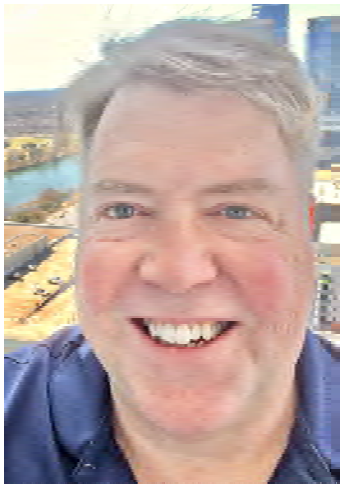
Intermediate and Advanced Y-DNA Topics

Beyond STRs

Brad Larkin

Prepared for the
Genealogical Forum of Oregon
Advanced DNA Special Interest Group
22 October 2022



Overview

- Deep dive into Y-DNA NextGen SNP genetic genealogy tests like Big Y-700 and YSeq WGS
  - From Biology
  - to Data
  - to the Phylogenetic Tree
  - to Geographic Interpretation
- Topics Include:
  - Explanation of raw read alignment to the genome which produces your SNP results
  - Understanding your terminal SNP and why it changes
  - How to use Novel Variants / Unnamed Variants to extend the phylogenetic tree for your family
  - Tips on getting the most info and avoiding bugs in current tools using real world examples

Topics

- Part I – SNP Detection
  - Biology & Sequencing
  - Alignment of Reads
  - Types of Files
  - BAM Viewing
- Part II – Trees and Interpreting SNPs
  - Nomenclature of SNPs in Y-DNA Genetic Genealogy
  - Phylogenetic Tree
  - Y-DNA Tree Backbone
  - Coincident mutations on same branch
  - Terminal SNP
  - Novel Variants / de Novo Mutations
  - Geography & Ethnic Attributions
  - Connecting Paper & Genetic Genealogy
- Part III - Tips

Objective

- Have a deeper understanding of how NextGen Y-DNA results are produced.
- Learn about the existence of visual tools for reviewing Y-DNA SNP results.
- Understand the concepts of Terminal SNP and Phylogenetic tree as they are developed from biology to data.

Topic Sections

Introduction

Part I SNP Detection

Part II Trees and Interpreting SNPs

Part III Tips

Introduction

- Intended Audience
- Speaker Background

Intended Audience

- Those using NextGen DNA Y-DNA tests of millions of base pairs.
  - Y-DNA Project Administrators
  - Y-DNA aficionados and independent researchers
    - Those trying to find new mutations / branches
  - Genetic genealogists & test takers
    - Who want to get deeper understanding
    - Greater visualization of their DNA results
  - DNA Software & Tool builders

Next-Gen Genetic Genealogy Labs

- Here we use Next-Gen Y-DNA sequencing products available from these commercial laboratories:
  - **Family Tree DNA** (Y)
  - **YSeq** (Y or whole genome)
  - **Full Genomes Corp** (Y or whole genome)
  - **23mofang** (Y)
  - **Nebula Genomics** (whole genome)

Presentation Not Intended For

- Adoption Solving
- Criminal or Missing Persons Investigations
- DNA tests involving less than 2 million base pairs:
  - Ancestry
  - 23andMe
  - Living DNA
  - MyHeritage

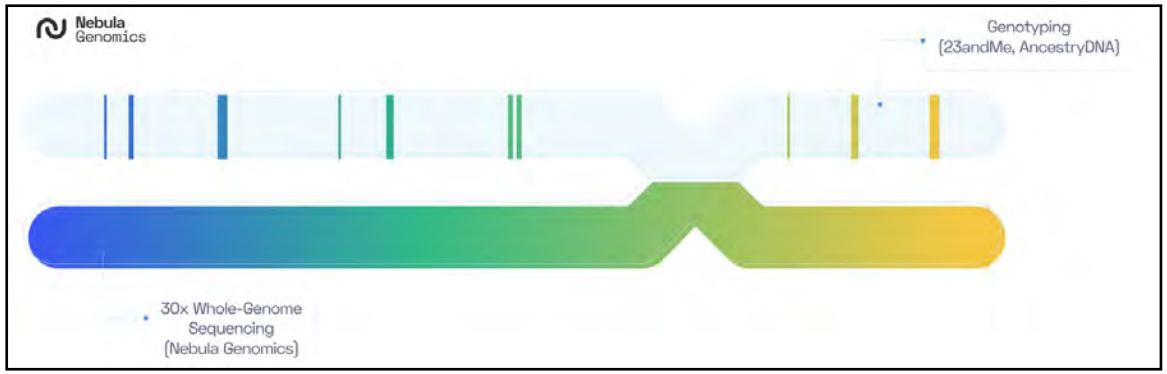


Image illustrating 30x coverage versus microarray technology adapted from video by Nebula Genomics https://nebula.org/whole-genome-sequencing-dna-test/

Topic Sections

Introduction

Part I SNP Detection

Part II Trees and Interpreting SNPs

Part III Tips

Part I – SNP Detection

- Biology
- Human Reference Sequence
- NextGen Sequencing
- Alignment of Reads (aka Mapping Reads)
- Types of Files
- BAM Viewing

Tree Branches are not usually Consecutive Generations

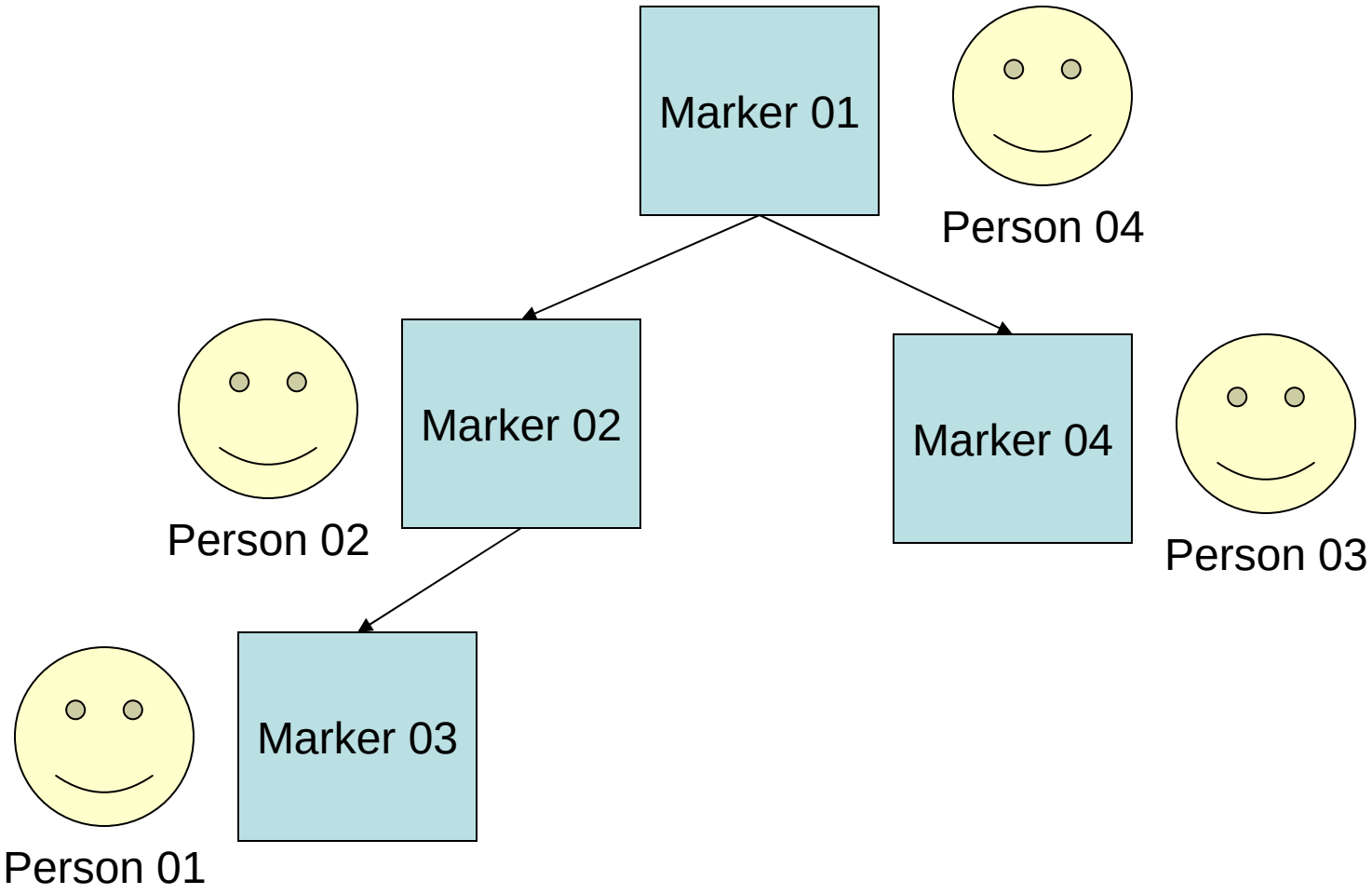
- Bear in mind the tree is not every mutation observed.
  - Just an assembly of those mutations which we have found which clearly distinguish groups of human beings.
- There are hundreds of thousands of mutations observed, but only a fraction of those have been found useful for tree branching.

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.

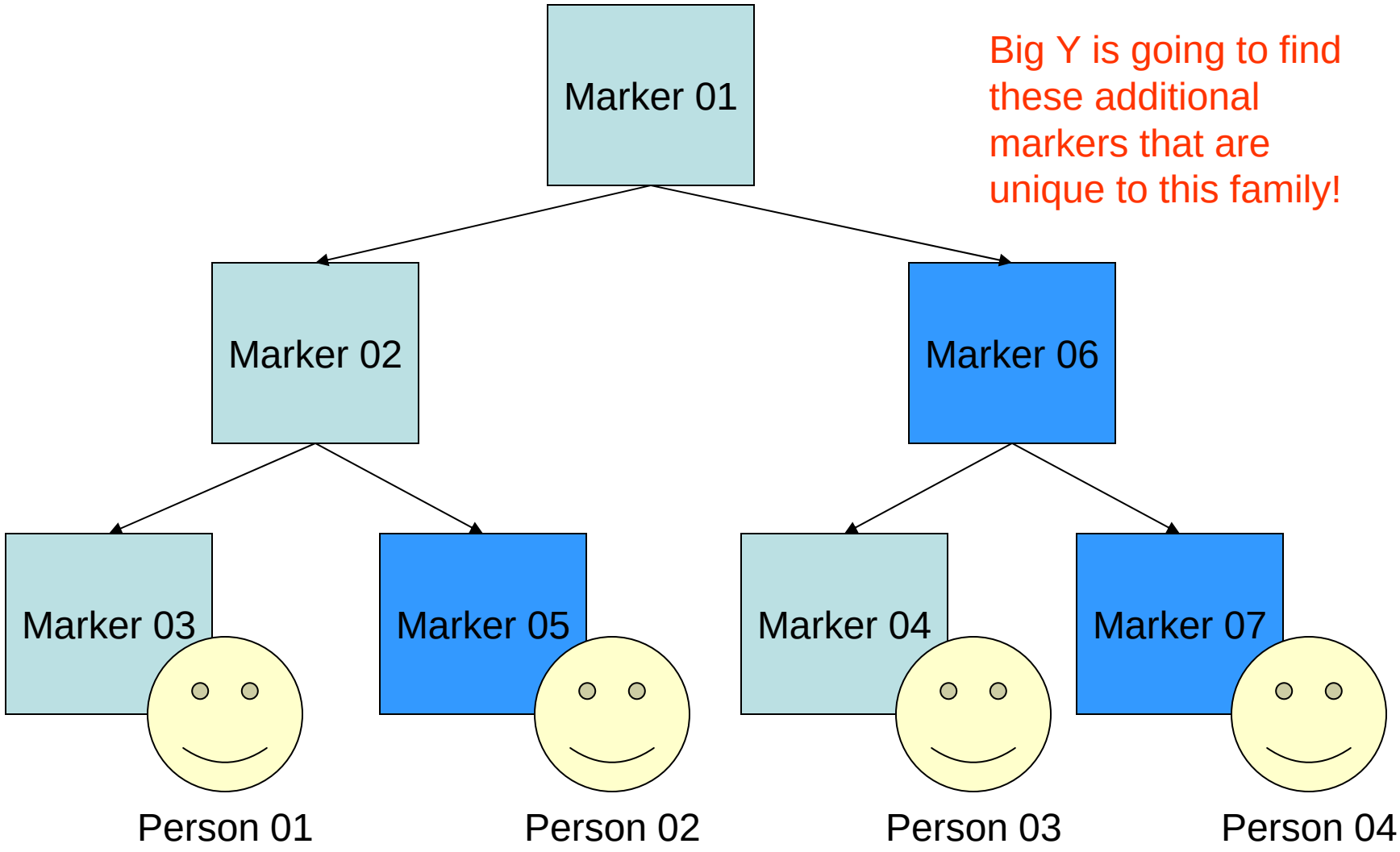
Marker Phylogeny Solution



Terminal SNP

- The term *Terminal SNP* refers to the deepest SNP marker on the Phylogenetic Tree that an individual has tested positively for.
  - As of 10/19/2022 Genetic Homeland Y-DNA phylogenetic tree has 31,920 distinct Terminal SNPs

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

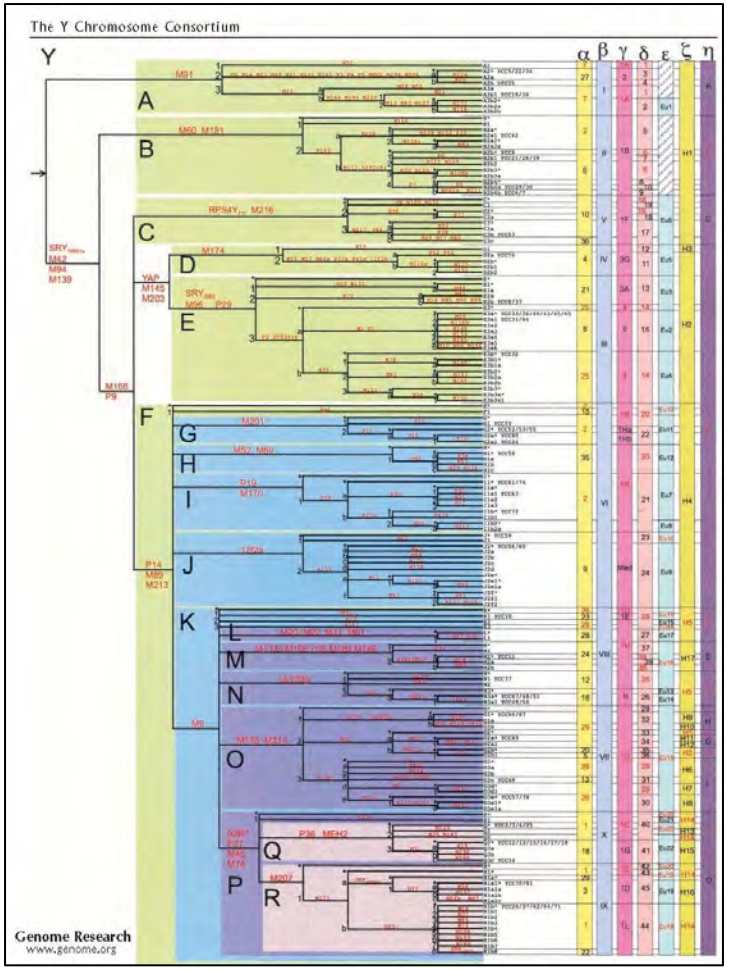
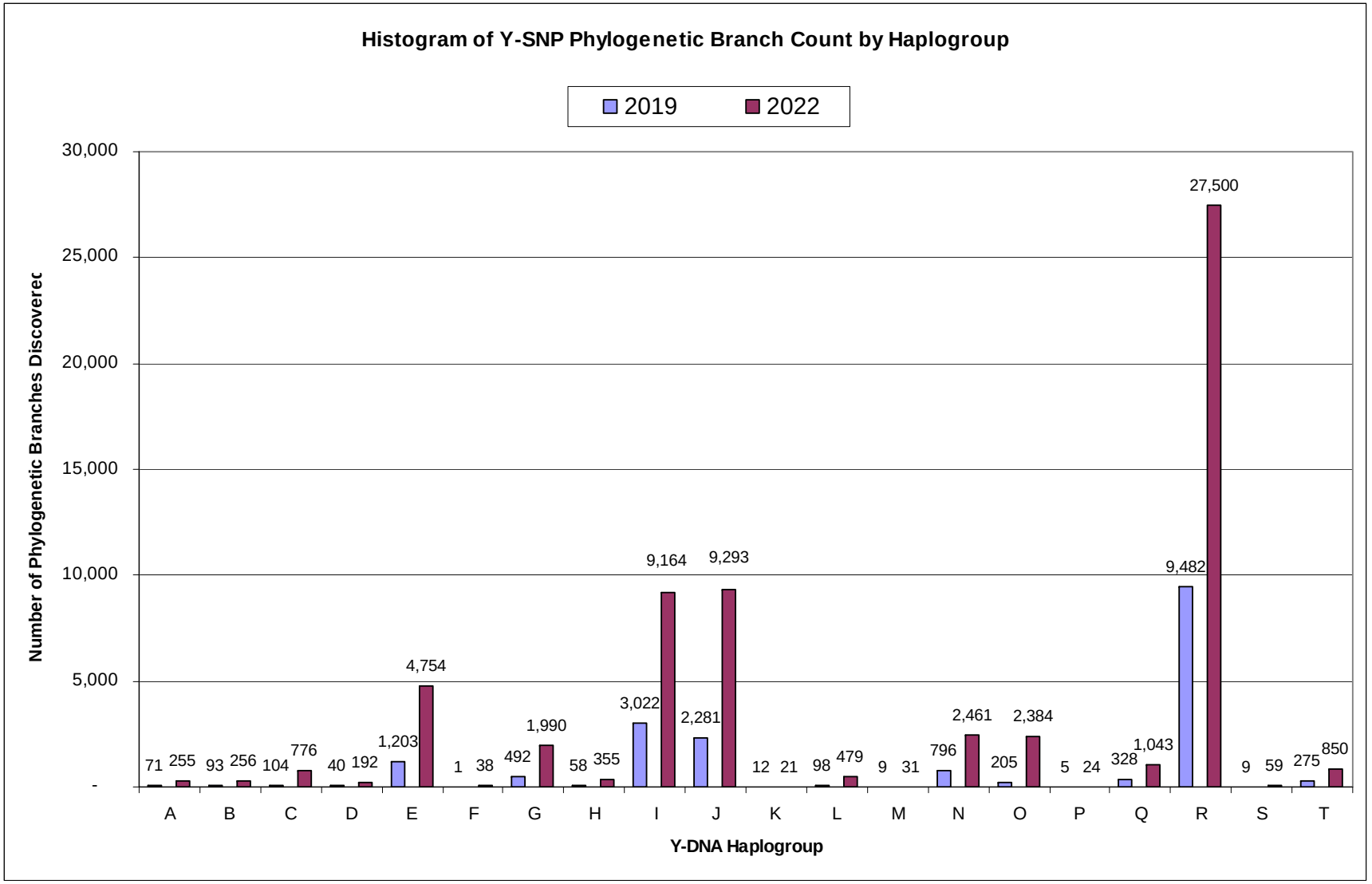


Image taken from Figure 1 from Michael F Hammer et al (2002) as The Y Chromosome Consortium in "A Nomenclature System for the Tree of Human Y-Chromosomal Binary Haplogroups" in Genome Research 12:339-348 www.genome.org

Y Phylogenetic Tree 2022

- Genetic Homeland Y-DNA phylogenetic tree as of 10/19/2022
  - 70,425 branches on entire Y-DNA tree
  - 2,352,323 SNP marker names (labels)
    - e.g. '*R-M269*'
    - 2,029,101 mutation events
  - Growing at a rate of roughly 750 new branches and 15,000 new SNPs per month!



Figures from Genetic Homeland DNA Ancestral Pedigree Tree as comparing 23 Feb 2019 with 5 Apr 2022.

Current Haplogroup Figures

- Count of phylogenetic tree branches and SNP markers by major ISOGG haplogroups in Genetic Homeland database.
  - https://www.genetichomeland.com/welcome/haplogroupcount.asp

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

Count of phylogenetic tree branches and SNP markers by major ISOGG haplogroups in Genetic Homeland database.

Updated at: *Friday, October 21, 2022 2:16:23 PM*

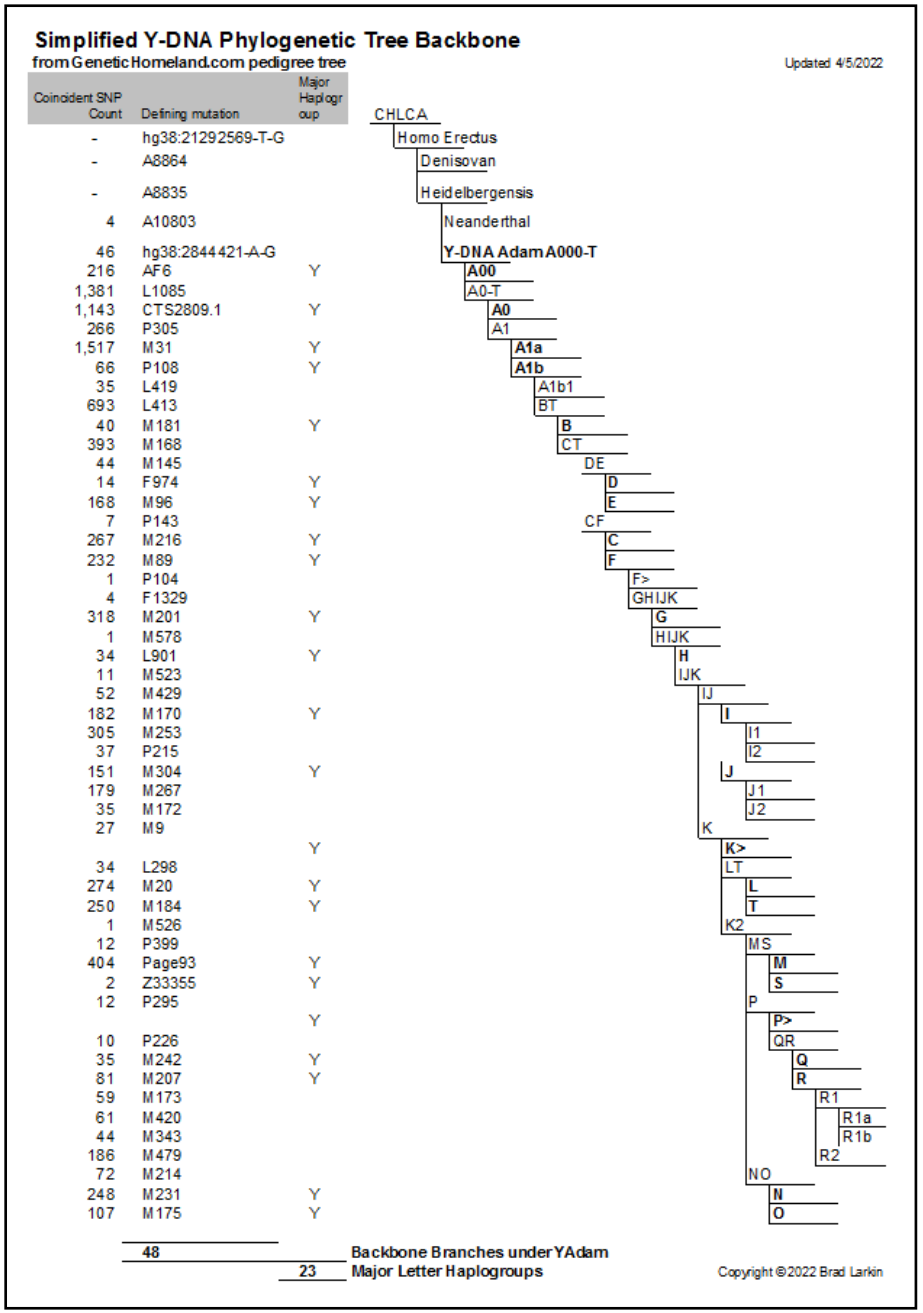
https://www.genetichomeland.com/welcome/haplogroupcount.asp

Chromosome	Haplogroup	Branch Count	SNP Count
Y	A	272	24,063
	A0	19	6,292
	A00	11	4,666
	A1a	41	2,279
	A1b	193	9,425
Y	B	283	16,653
Y	C	972	50,996
	C1a	68	2,441
	C1b	189	5,253
	C2a	307	4,915
	C2b	402	5,135
Y	D	213	12,779
Y	E	5,426	88,817
	E1a	160	3,388
	E1b	5,189	52,790
	E2a	11	1,130
	E2b	40	1,129
Y	F	41	2,482
Y	G	2,198	39,438
	G1a	105	1,737
	G1b	6	273
	G2a	864	14,140
	G2b	72	1,083

Chromosome	Haplogroup	Branch Count	SNP Count
Y	H	388	17,537
Y	I	10,185	99,623
Y	J	10,491	126,188
	J1a	6,470	32,974
	J1b	34	1,219
	J2a	3,179	35,203
	J2b	804	7,971
Y	K	21	1,418
Y	L	530	10,228
Y	M	57	1,514
Y	N	2,660	44,815
Y	O	3,700	405,165
	O1a	578	4,947
	O1b	1,581	10,488
	O2a	1,507	14,827
	O2b	20	564
Y	P	24	1,942
Y	Q	1,130	34,409
Y	R	31,049	302,710
	R1a	4,994	54,691
	R1b	25,765	234,720
	R2	288	8,129
Y	S	82	4,620
Y	T	961	15,459
All Root Haplogroups		70,683	1,300,856

Y-DNA Tree Backbone

- 46 backbone branches under YAdam
- 23 Major Letter Haplogroups



Phylogenetic Tree for BY21680 in 2022

- Phylogenetic tree from HomoErectus to BY21680 using Genetic Homeland Ancestral DNA Marker Pedigree tool
- In 2008 it was 43 branches.
- As of 10/19/2022 it is 56 branches.
  - That is 13 levels deeper in 14 years.

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	HeidelbergensIs	A0033	Presumed Y Ancestor of Humans and Neander thals. ISOGG calls this haplogroup A000-T.
3	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. Believed coincident with PR2921. ISOGG calls this haplogroup A00-T.
4	L1085	A0-T hg38:2922685-T-C	Defining mutation for near-root haplogroup A0-T (aka A0'T'23'4). That is to say father of ALL modern Y-DNA lineages except A00. hg38 Ref does not match ancestral allele value.
5	P305	hg38:2842113-A-G A1 rs72625368	Defining mutation for ISOGG haplogroup A1 branch. hg38 Ref does not match ancestral allele value.
6	P108	A1b rs761539052	Defining mutation for ancient haplogroup branch A1b at ISOGG. hg38 Ref does not match ancestral. Also enumerated as hg38:13314368-C-T.
7	L413	PF1409 V31 BT rs192935307	Defining mutation for haplogroup branch BT at ISOGG. hg38 Ref does not match ancestral allele value.
8	M168	PF1416 CT-M168 CT rs2032595	Defining mutation for haplogroup branch CT. hg38 Ref does not match ancestral allele value.
9	P143	PF2587 CF-P143 CF rs4141886	Defining mutation for haplogroup branch CF. hg38 Ref does not match ancestral allele value. Example is ancient sample I8142 from Granada, Spain 2,100 bce.
10	M89	PF2746 F rs2032652	Defining mutation for ancient haplogroup F. hg38 Ref does not match ancestral allele value.
11	F1329	M3658 PF2622 YSC0001299 V2308 GHIJK-F1329 GHIJK F-F1329 rs9786482	Defining mutation for haplogroup branch above HIJK (sic). hg38 Ref does not match ancestral allele value. V2308 has alleles reversed.
12	M578	F929 S6397 PF3494 S6378 HIJK-PF3494 HIJK rs73614810	Mutation at haplogroup branch HIJK. hg38 Ref does not match ancestral allele value.
13	M523	L15 PF3492 S117 Z4413 IJK-L15 IJK rs9786139	Defining mutation for haplogroup branch IJK. hg38 Ref does not match ancestral allele value.
14	M9	PF5506 K rs3900	Defining mutation for haplogroup K. Predecessor of most non-African haplogroups. Arose about 45,000 bce. hg38 has incorrect Ref versus ancestral allele.
15	M526	PF5979 rs2033003	Found in ancient haplogroup K2 on ISOGG, YFull, and FTDNA trees. hg38 Ref does not match ancestral allele value. Example is ancient sample Ust-Ishim from Russic 43,000 bce.
16	M1221	PF5911 YSC0186 P331 YSC0000186 MF44733 rs9785994	Defines haplogroup K2b on ISOGG, YFull, and FTDNA trees. Ancestor of haplogroups M, P, Q, R and S. hg38 Ref does not match ancestral allele value.
17	PF5850.1	PF5850 rs74550416	On academic study. Defining mutation for haplogroup P root on YFull and FTDNA trees. Example is sample JHM6 (accession ERS2589709) from Johoi tribesman of Malaysia. hg38 Ref does not match ancestral allele value.
18	P295	PF5866 S8 rs895530	Defining mutation for haplogroup P root (K2b2) on ISOGG tree. Originated about 42,000 bce. hg38 Ref allele does not match ancestral.
19	M1254	PF6062 rs66540167	Found in ancient haplogroup P above P337 (aka K2b2) on YFull tree. hg38 Ref does not match ancestral allele value. Example is ancient DNA sample from Andoman Islands.
20	P337	F1857 Page83 PAGES00083 PF5901 PAGE083 rs13305774	As intermediate branch on YFull tree. hg38 Ref does not match ancestral allele value. Example is ancient sample Yana1 from Russia 29,600 bce.
21	P284	rs4116821	On YFull and FTDNA trees again. Example is ancient sample Yana2 from Russia 29,600 bce.
22	P226	PF5879 rs17250992	Found in ancient haplogroup P1a1 (cka QR) - ancestor of Q and R. hg38 Ref does not match ancestral allele value.
23	M207	Page37 PAGES00037 PF6038 UTY2 rs2032658	Designates major haplogroup R, ancestor to R1b, R1a, and R2 (aka K2b2a2). Arose about 30,000 bce. hg38 Ref does not match ancestral allele value. Example is ancient sample MA1 from Irkutsk, Russic 22,000 bce.
24	M173	P241 Page29 PAGES00029 PF6126 rs2032624	Designates major haplogroup branch R1*. Arose about 26,000 bce. hg38 Ref does not match ancestral allele value.
25	M343	PF6242 R1b rs9786184	Designates mutation for haplogroup R1b root. Arose about 22,000 bce. hg38 Ref does not match ancestral allele value. Example is ancient sample I4315 from Uzbekistan 1450 bce.
26	L754	PF6269 YSC0000022 rs7893073	Defines branch R1b1 on ISOGG, YFull, and FTDNA trees. hg38 Ref does not match ancestral allele value. Example is ancient sample I4315 from Uzbekistan.
27	L389	PF6531 rs1358368	Defines branch R1b1a on FTDNA and YFull trees. hg38 Ref does not match ancestral allele value.
28	P297	PF6398 MF48762 R1b1a1a rs9785702	As an early branch that probably originated in central Asia about 13,600 bce. hg38 Ref does not match ancestral allele value. Formerly labeled R1b1a1a haplogroup in old literature.
29	M269	PF6517 R1b1a1a2 MF53029 rs9786153	Defining mutation for Western Atlantic Modal Haplotype (WAMH) of R1b in Europe. Originated about 11,000 bce. hg38 Ref does not match ancestral allele value. Britain's DNA labeled this branch: Anatolian. Formerly labeled R1b1a1a2 haplogroup in older literature.
30	L23	PF6534 S141 R1b1a1a2a rs9785971	Arose about 6,000 bce. hg38 Ref does not match ancestral allele value. Sometimes labeled as R1b1a1a2a haplogroup in literature using older nomenclature.
31	L51	M412 PF6536 S167 rs9786140	In Europe, almost entirely west of the Danube river. hg38 Ref does not match ancestral allele value. Example is ancient sample PCW070 from Poland 2,300 bce.
32	P310	PF6546 S129 rs9786283	On ISOGG and FTDNA trees. Believed coincident with L52. hg38 Ref does not match ancestral allele value. Examples include ancient samples PCW361 and PCW362.
33	PF6536	rs55975037	Under L52 on YFull tree. hg38 Ref does not match ancestral allele value. Example is ancient DNA sample PCW040 from Poland 2,350 bce.
34	L151	PF6542 rs2082033	Under M269 on ISOGG, YFull, and FTDNA trees. hg38 Ref does not match ancestral allele value.
35	P312	PF6547 S116 MF52579 rs34275300	Largest branch under haplogroup R1b. Arose about 5,000 bce. hg38 Ref does not match ancestral allele value. Example is ancient sample esp005 from Spain 1,300 bce. Britain's DNA labeled this branch: Beaker Folk and formerly Bell Beaker.
36	Z290	S461 rs146019383	Largest branch under under R1b M269
37	L21	M529 S145 rs11799226	Largest European group under R1b P312. Highly correlated with geography of ancient Celts. Britain's DNA labeled this branch: Pretani.
38	S552	FGC3218 Y2598 rs150868296	On FTDNA tree as an intermediate branch under L21 and ancestral to DF13 et al.
39	DF13	CTS241 S521 rs55989227	Major branch of L21 in haplogroup R1b. Originated about 2,500 bce.
40	Z39589.1	Z39589 Z39589_2	Major branch in haplogroup R1b under L21 and DF13 with many descendant branches from Ireland. Deletion of 10 base pairs. Note that hg38 position description not linear translation from hg19 liftover map. Also enumerated as hg19:44439911-TGCAGCTTCACTCCTGAGG-T.
41	DF49	S474 AM01922 rs769171919	Large branch under L21.
42	Z2980	S6154 rs1006335787	Under L21 and DF49.
43	Z2976	S476 S644 S614 / AM01919 rs770179167	Under L21 and DF49.
44	DF23.1	DF23 S193 rs773517566	Under Z39589 on ISOGG, YFull, and FTDNA trees.
45	Z2961	rs771631896	Under DF23 on ISOGG, YFull, and FTDNA trees. Most concentrated in Ireland.
46	Z2956	AM01923 rs1006912931	On academic study.
47	Z2965	S6155 rs1282364675	Under L21. Originated about 2200 bce.
48	M222	Page84 PAGES00084 USP9Y-3636 rs35720707 rs20321	Sometimes called Northwest Irish, concentrated in Ireland and western Scotland. Associated with Niall of the Nine Hostages and Ui Neill clans. Britain's DNA labeled this branch: Ancient Irish.
49	Y2605	FGC4124 rs148684963?	On YFull tree as an intermediate branch under M222.
50	S659	DF106 FGC4100 Y2841 rs747839864	Under M222 on ISOGG, YFull, and FTDNA trees. Example is ancient sample VK95 from Iceland 1200 ce.
51	DF104	S661 Y2842 FGC4099 rs752415261	Under M222 on ISOGG, YFull, and FTDNA trees.
52	DF105	S659 Y2843 FGC7927 rs75574899	Under M222 and DF104 on FTDNA tree. Believed coincident with DF109 / S660.
53	A18726		Under M222, DF109, and DF105 on YFull and FTDNA trees.
54	BY198	A738	Under M222 and A18726 on YFull and FTDNA trees. Discovered by Thomas Krahn & Iain Kennedy.
55	BY20835	A15864	Under M222 and BY198 on FTDNA tree. Believed coincident with A15865.
56	BY20834	S27575 rs3097069	Under M222, BY198, and BY20835 on FTDNA tree.
57	BY21680	R-BY21680 A15870	Found in Larkin Type 01 of Ireland.

View Map of descendants of BY21680

Giant Tree of Mankind

- "Tapestry of individual lineages"

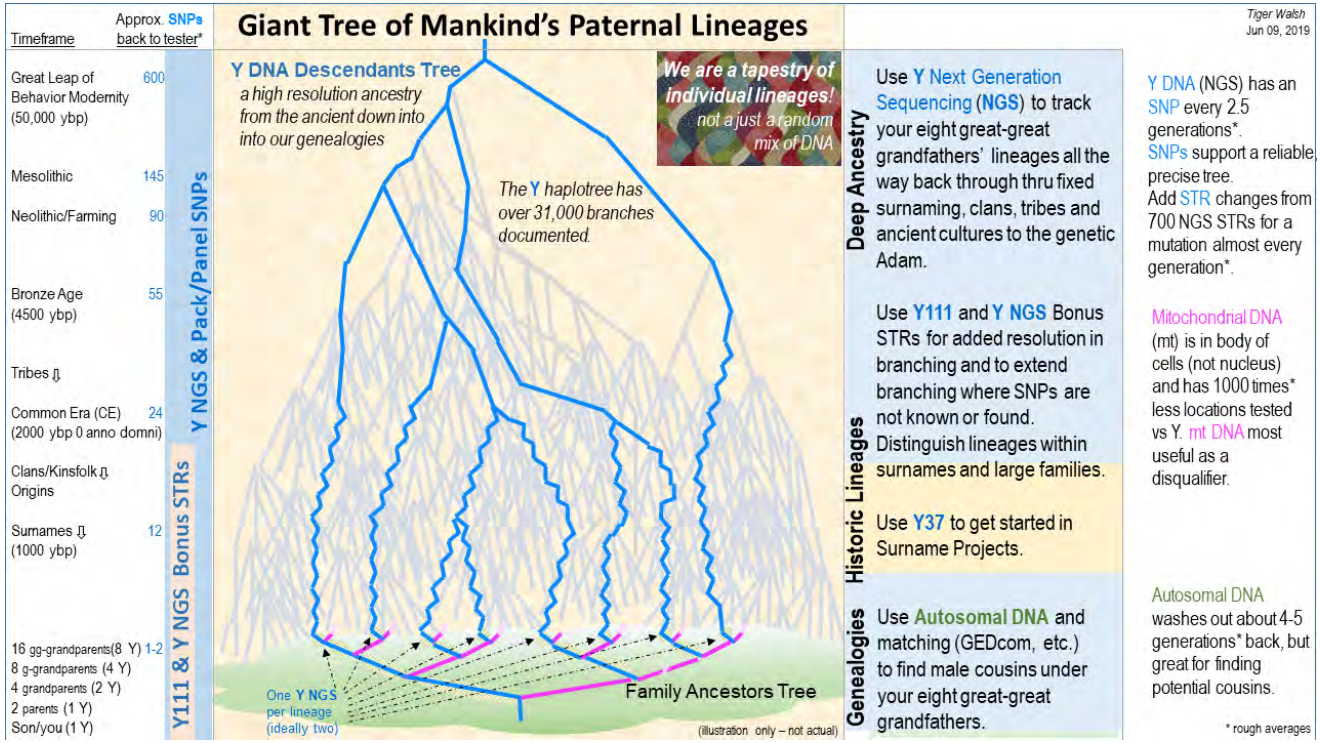


Image & quotation from Tiger Walsh 2019, R1b All Subclades DNA Project at FTDNA, https://www.familytreedna.com/groups/r-1b/about/background

Discovery Time vs Ancestral Time

- Branch Level 40 ~ about 2200 years of ancestral time.
- M222 marker was only discovered in 2006
  - (16 years ago)
- A738 / BY198 discovered 2014
  - then only 2 examples
  - today there are 77
- BY21680 correctly located in tree in 2018

Phylogenetic Ancestral Tree for marker [BY21680]

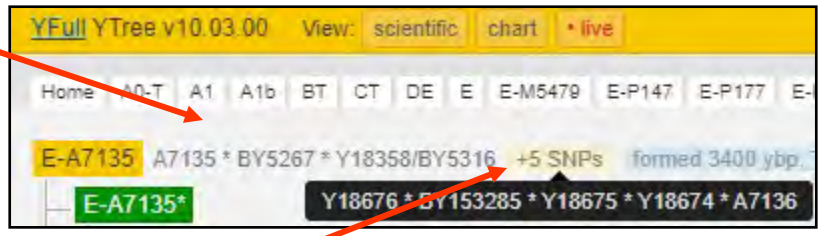
Tree Level	Marker / Branch Name	Alternative Names	Notes
48	M222	Page84 PAGES00084 USP9Y+3636 rs35720707 rs20321	
49	Y2605	FGC4124 rs1486849632	
50	S658	DF106 FGC4100 Y2841 rs747839864	
51	DF104	S661 Y2842 FGC4099 rs752415261	
52	DF105	S659 Y2843 FGC7927 rs755714899	
53	A18726		
54	BY198	A738	
55	BY20835	A15864	
56	BY20834	S27575 rs3097069	
57	BY21680	R-BY21680 A15870	

Coincident Mutations

- Groups of mutations which always *coincide* together!
  - SNP mutations that are always found or not-found in all samples so far tested.
  - Useful in phylogeny because when we find one of the markers, we can usually make a branch assignment of that individual sample.

Coincidents on YFull Tree

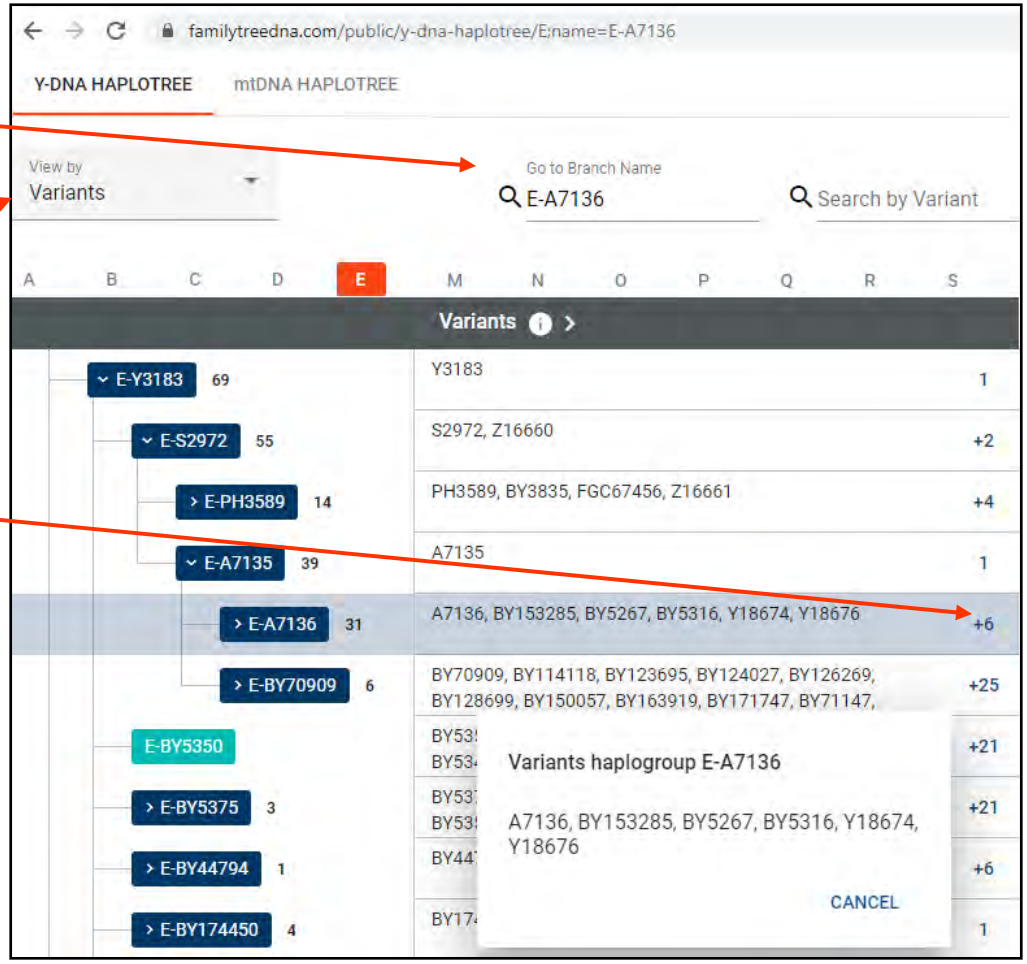
- Up to three displayed on branch page.
  - BY5267 and Y18358/BY5316
  - When more than three identified, mouseover the "+5 SNPs" item.
    - Y18676, BY153285 are shown in this example



Note also that phylogenetically: YFull lists A7136 as coincident with A7135

Coincidents on FTDNA Y-DNA Haplotree

- Navigate to your SNP branch node.
  - e.g. E-A7136
- Change *View by* to Variants
- Two rows display immediately as well as a numeric count of the total number of markers on this branch..
  - Click on the numeral to see the full list in a pop-up box.



Note also that FTDNA recognizes A7136 phylogenetically as a descendant branch of A7135.

What Causes Coincident Y Mutations?

- Biological
  - One ancestral individual got two or more SNPs from his or her parent.
  - Accumulation of mutations across many generations in our ancestral lineage combined with population bottlenecks.
    - Most common reason for coincident mutations in our deep ancestral branches – tens of thousands of years ago.
  - Palindromes, large DNA segment shifts or repeats
- Limited Sampling
  - Have not tested enough of those descendants to see the branching.
    - Most common reason for coincident mutations in our most recent branches.

Coincident Mutations

- Coincident mutations can also be very helpful in cases where a ‘back-mutation’ of one base pair to the ancestral condition occurs.
- From a tree-building perspective, coincident mutations on are like *star dust from which the next branches will be born*.

Terminal SNP

- In this pedigree, at this moment, it is BY21680

42	Z2980	S6154 rs1006335787	Under L21 and DF49.
43	Z2976	S476 S644 S6147 AM01919 rs770179167	Under L21 and DF49.
44	DF23.1	DF23 S193 rs773517566	Under Z39589 on ISOGG, YFull, and FTDNA trees.
45	Z2961	rs771631896	Under DF23 on ISOGG, YFull, and FTDNA trees. Example is YF001534 from Ireland.
46	Z2956	AM01923 rs1006912931	Under Z2961 on FTDNA tree. Believed coincident with S645.
47	Z2965	S6155 rs1282364675	Under L21, DF23, S645, and Z2956 on YFull and FTDNA trees. Originated about 2150 bce. Example is YF071698.
48	M222	Page84 PAGES00084 USP9Y+3636 rs35720707 rs20321	Sometimes called Northwest Irish, concentrated in Ireland and western Scotland. Associated with Niall of the Nine Hostages and Ui Neill clans. Britain's DNA labeled this branch: Ancient Irish.
49	Y2605	FGC4124 rs1486849632	On YFull tree as an intermediate branch under M222.
50	S658	DF106 FGC4100 Y2841 rs747839864	Under M222 on ISOGG, YFull, and FTDNA trees. Example is ancient sample VK95 from Iceland 1200 ce.
51	DF104	S661 Y2842 FGC4099 rs752415261	Under M222 on ISOGG, YFull, and FTDNA trees.
52	DF105	S659 Y2843 FGC7927 rs755714899	Under M222 and DF104 on FTDNA tree. Believed coincident with DF109 / S660.
53	A18726.1	A18726	Under M222, DF109, and DF105 on YFull and FTDNA trees.
54	BY198	A738	Under M222 and A18726 on YFull and FTDNA trees. Discovered by Thomas Krahn & Iain Kennedy.
55	BY20835	A15864	Under M222 and BY198 on FTDNA tree. Believed coincident with A15865.
56	BY20834	S27575 rs3097069	Under M222, BY198, and BY20835 on FTDNA tree.
57	BY21680	R-BY21680 A15870	Found in Larkin Type 01 of Ireland.
		View Map of descendants of BY21680	

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.

What is your Terminal SNP

- I think of it as the most unique SNP marker that is visible at a moment in time.
  - Your DNA is not changing. But the parts that we can see is changing.
    - What we 'see' changes when
      - You get an upgrade or laboratory expansion of to test more base pairs on the Y-chromosome
      - More persons provide samples to test at all the same positions that you have tested
      - The phylogenetic tree expands around you when laboratories and researchers derive the sequence of mutations amongst related individuals

Big Y SNP Matching

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

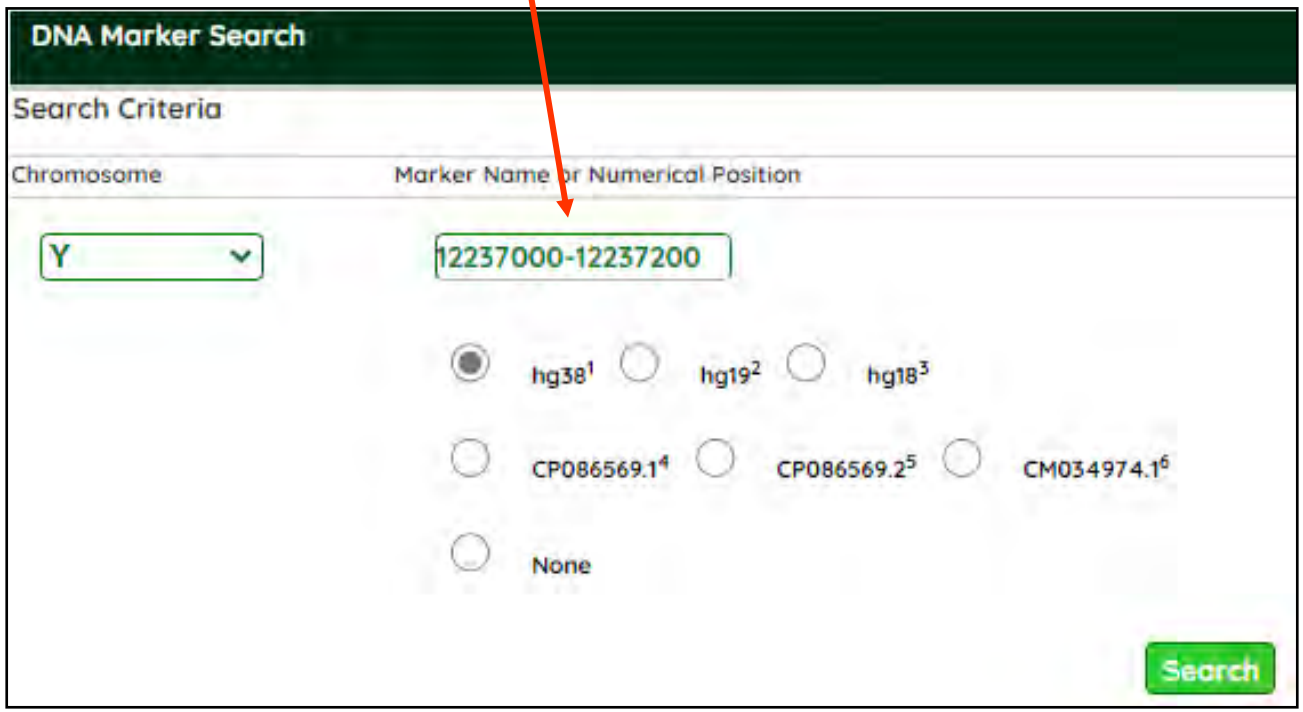
[1]As of Jan 30, 2019 FTDNA classifies Big Y matches based on "30 or fewer differences in SNPs with you, and their haplogroup is downstream from your haplogroup or downstream from your four closest parent haplogroups" https://www.familytreedna.com/learn/y-dna-testing/big-y/big-y/

Novel Variants

- NOVEL VARIANTS ARE YOUR UNIQUE ANCESTRAL 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

Search for SNPs by Range of Positions

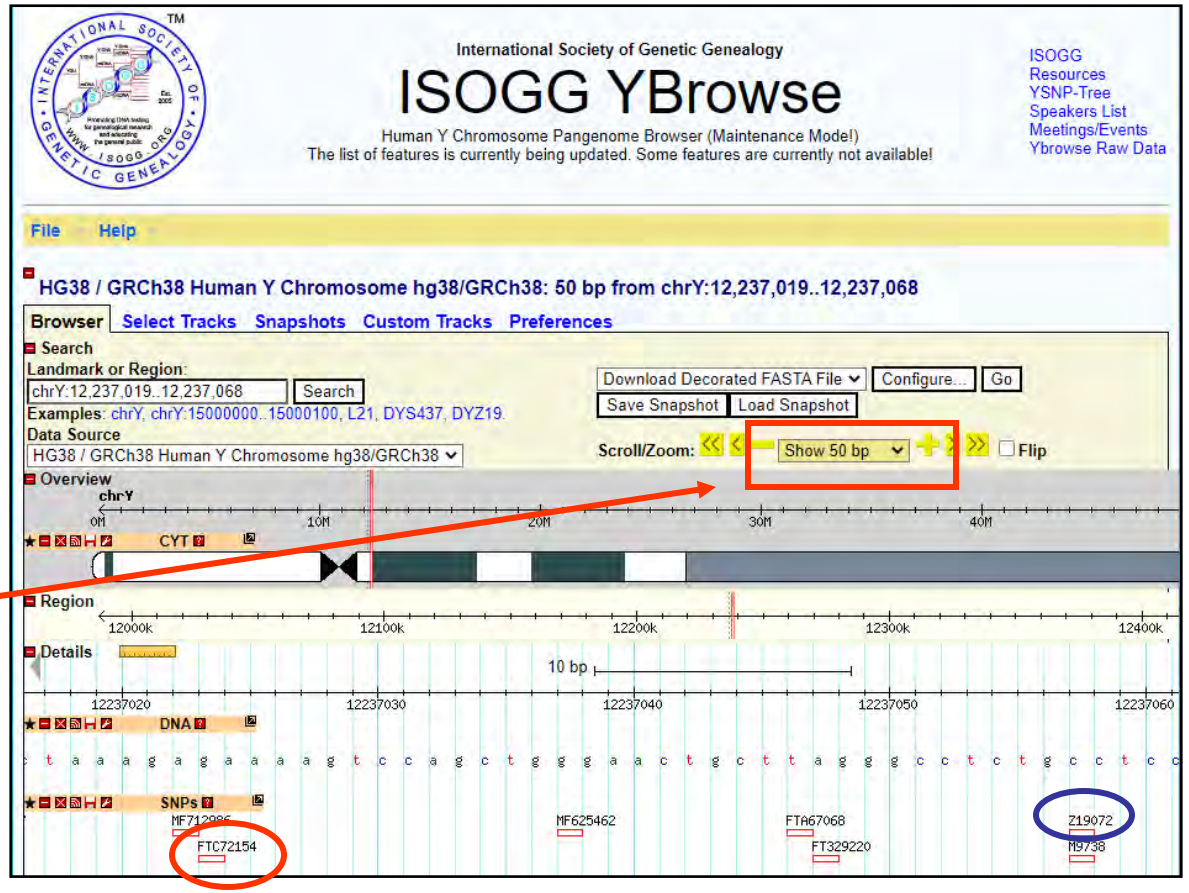
- Use a tool which can be queried for a list of SNPs within a range of chromosome positions. You might be able to identify SNP Names proximate to your Private Variant.



YBrowse Range Example

- Viewing a range of Y chromosome positions using YBrowse tool.

  Zoom out to ~ 50 bp



Genetic Homeland Range Example

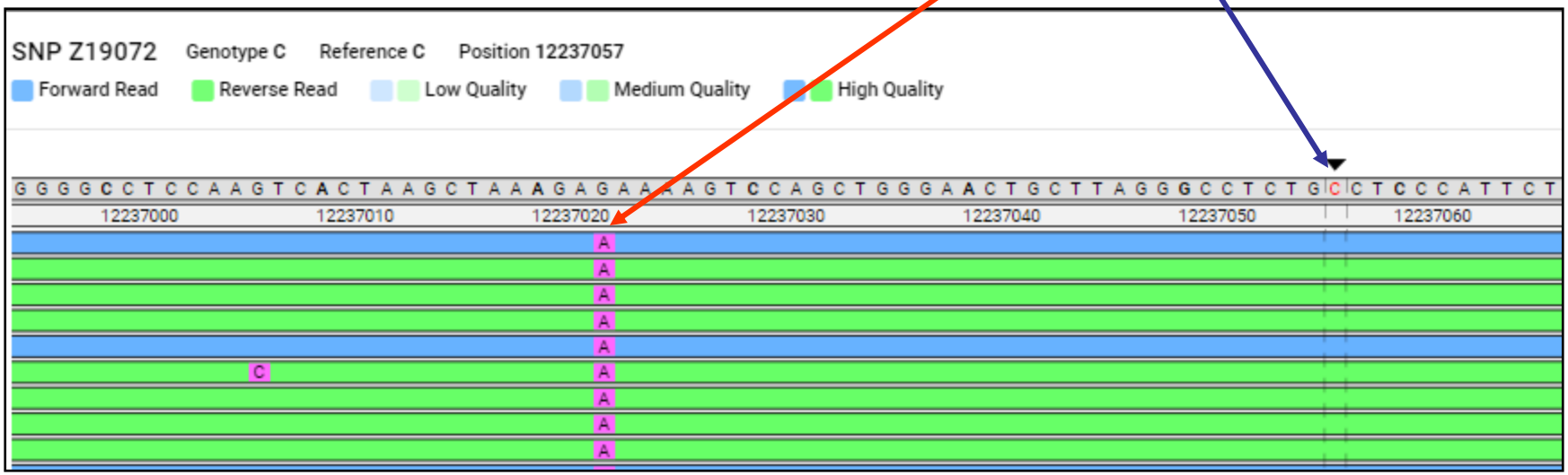
- By trial & error, found that Z19072 would work on FTDNA Chromosome Browser

DNA Marker Index data for Marker: 12237000-12237200 on Chromosome: Y

Marker Name(s)	Notes	Identification	Ancestral	Derived	Chromosome	Position (hg38)
O-MF712985 MF712985	Major Haplogroup: O.	23mofang 2021	C	T	Y	12237000
FT311737		FTDNA 2020	A	G	Y	12237014
O-MF712986 MF712986 rs1471811847	Major Haplogroup: O. NIH DBSNP b153 GRCh38p12 ClinVar? Position hg18: 12867728	23mofang 2021	A	G	Y	12237022
F-FTC72154 FTC72154	Found in haplogroup R1b under M222 and S588. Coincident with FTC72153. View Pedigree Using Coincident Marker [FTC72153] on Tree: 3259057	FTDNA 2022	G	A	Y	12237023
O-MF625462 MF625462	Major Haplogroup: O.	23mofang 2021	G	A	Y	12237037
FTA67068		FTDNA 2021	T	C	Y	12237046
E-FT329220 FT329220 rs771177491	Found in haplogroup E on academic study. Coincident with Z36911. View Pedigree NIH DBSNP b153 GRCh38p12 ClinVar? Position hg18: 12867753	FTDNA 2020	A	G	Y	12237047
H-Z19072 Z19072	Found in haplogroup H2a on YFull and FTDNA trees. Example is ancient sample JP14 from Ireland 3,550 bce.	Ray Banks 2015	C	A	Y	12237057

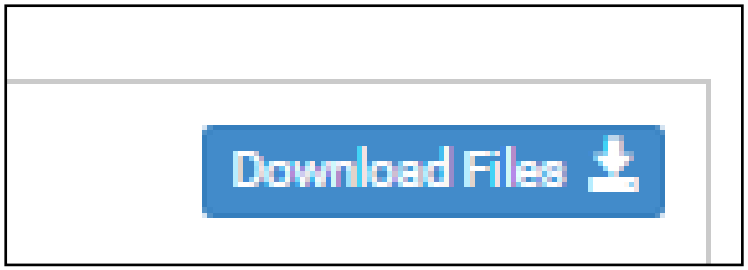
Scroll From Named Marker to Position of Interest

- Once you've found a recognized, named SNP, try scrolling the FTDNA Chromosome Browser to see if there were reads at this position.
  - Position of Z19072 = 12237057
  - Position of Novel Variant: 12237023



Download Raw Data BAM File

- If these workarounds and limitations did not whet your appetite, best thing is to pay the money to get your raw BAM file from FTDNA for a fee.



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	-

Geography & Ethnic Attributions

- Proof that I am Antartican!

- *THE MARKER*
  - that tells me I have been right all along ….
  - matches what my grandmother told me ….
  - shows that I am royalty ….

Be Cautious and Open Minded

- Geographical & Ethnic Attribution
  - Humans always mixing and traveling.
  - For every marker where we conclude marker XYZ originated in geography ABC, there are exceptions.
  - But where we see a lot of geographical continuity in a lineage which is well-sampled, it does suggest a geographical and cultural heritage.
    - But usually at much higher resolution than the researcher has tested.

Viking Study

Margaryan (2020) study of viking graves.

- n = 297 samples with Y-DNA haplogroups
- I1 was most common 32%.
- R1b also common, 29%.
- Also 8 haplogroups found at low frequency today

=> Viking was an occupation as much as an ethnicity.

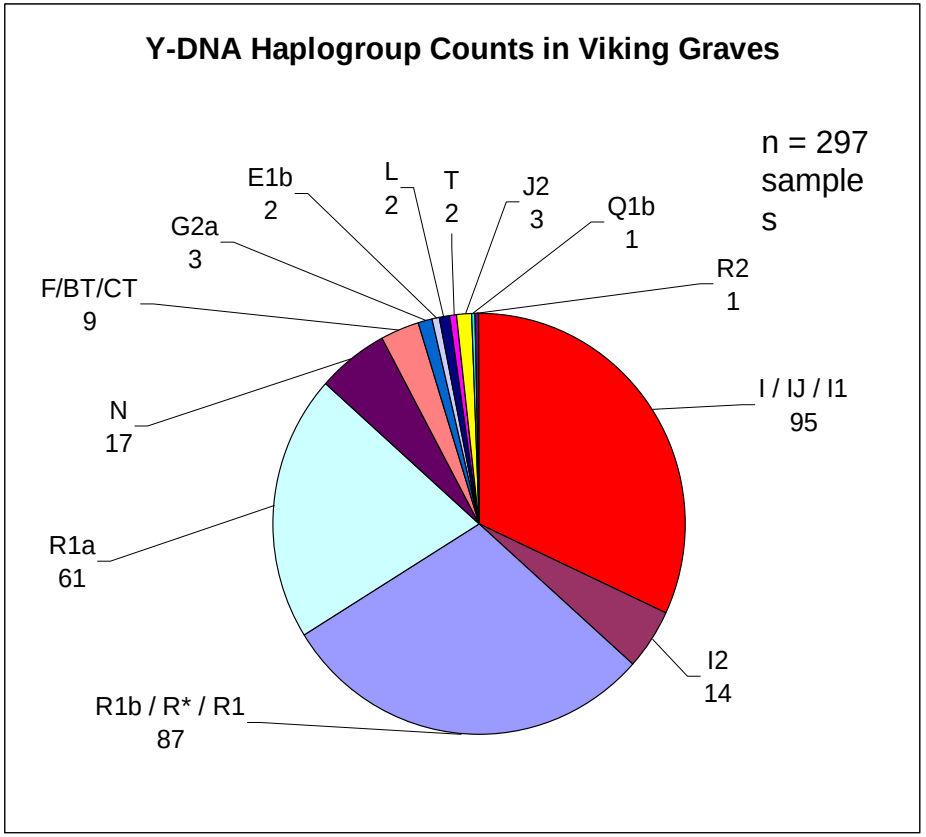
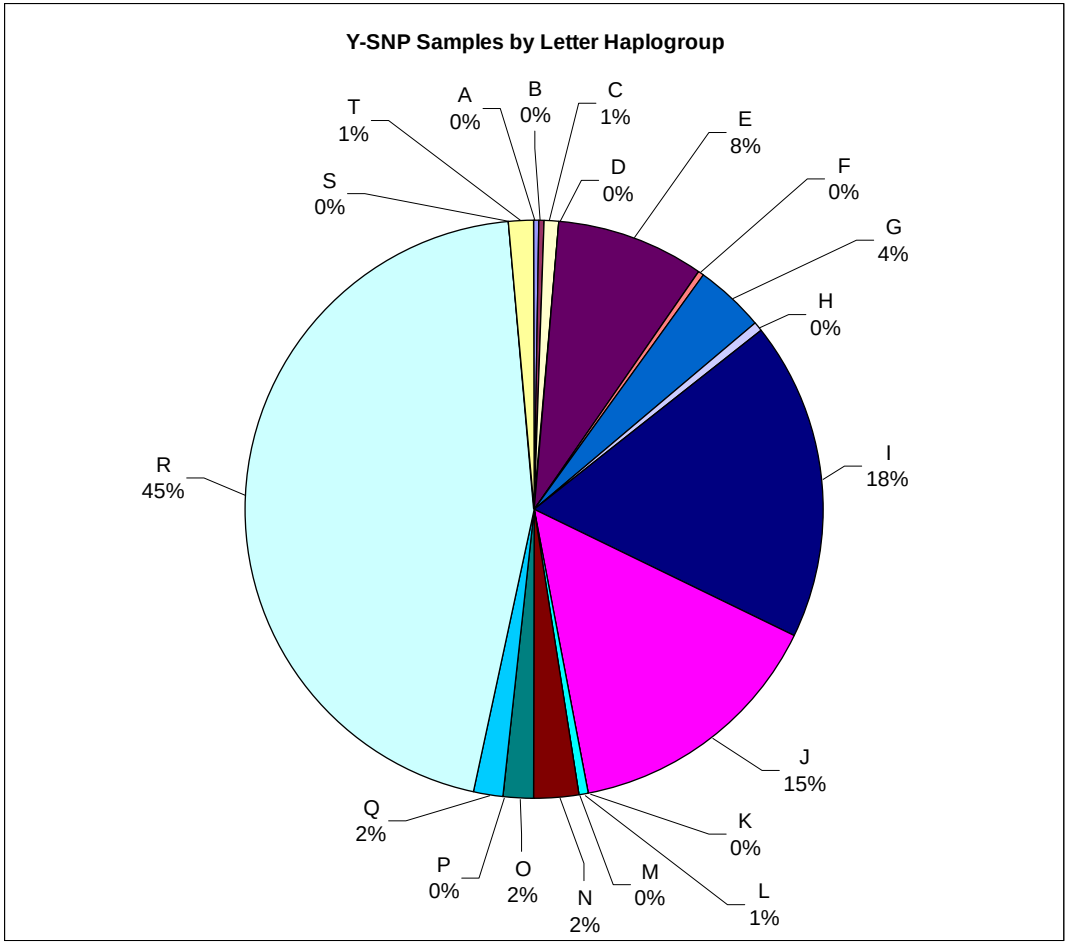


Image by author using data from Margaryan et al, Population genomes of the Viking world, Nature vol 585, 2020

Low Resolution Fallacies

- STR and early SNP findings are low resolution Y-DNA resolution.
- Ethnic attributions to these markers like R-M269, M223, L21, etc occurred in Neolithic and bronze ages.
  - Are meaningful only at the continental level.
  - Genetics attributed to historical events and populations is vastly over-simplified.
- Avoid *making a mountain of conclusion from a mole-hill of SNP data*.

Y-SNP Samples by Haplogroup



Y-DNA World Map - Pre-Colonial

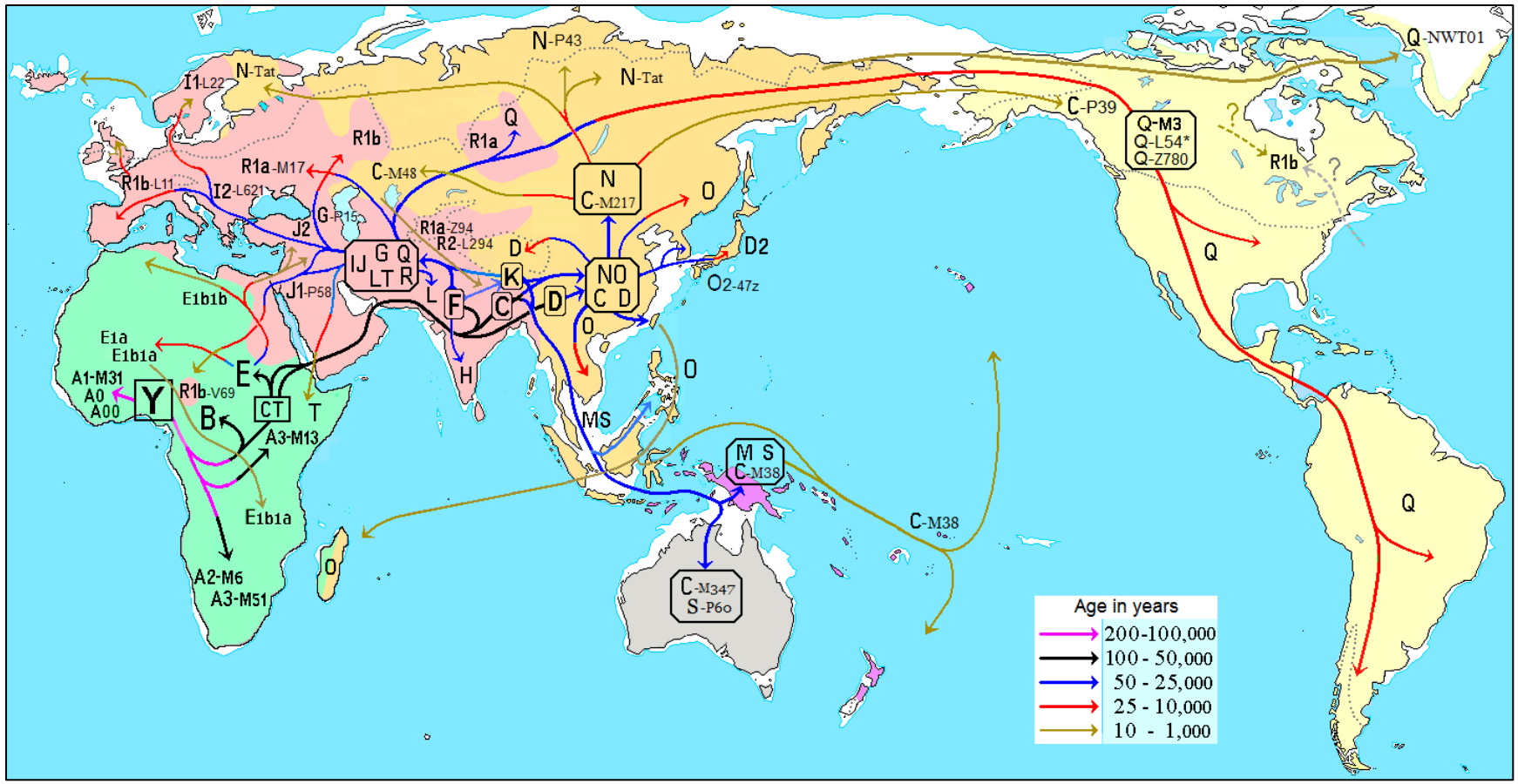


Image by Chakazul (2013) on Wikipedia, World Map of Y-DNA Haplogroups

Y-DNA World Map - Today

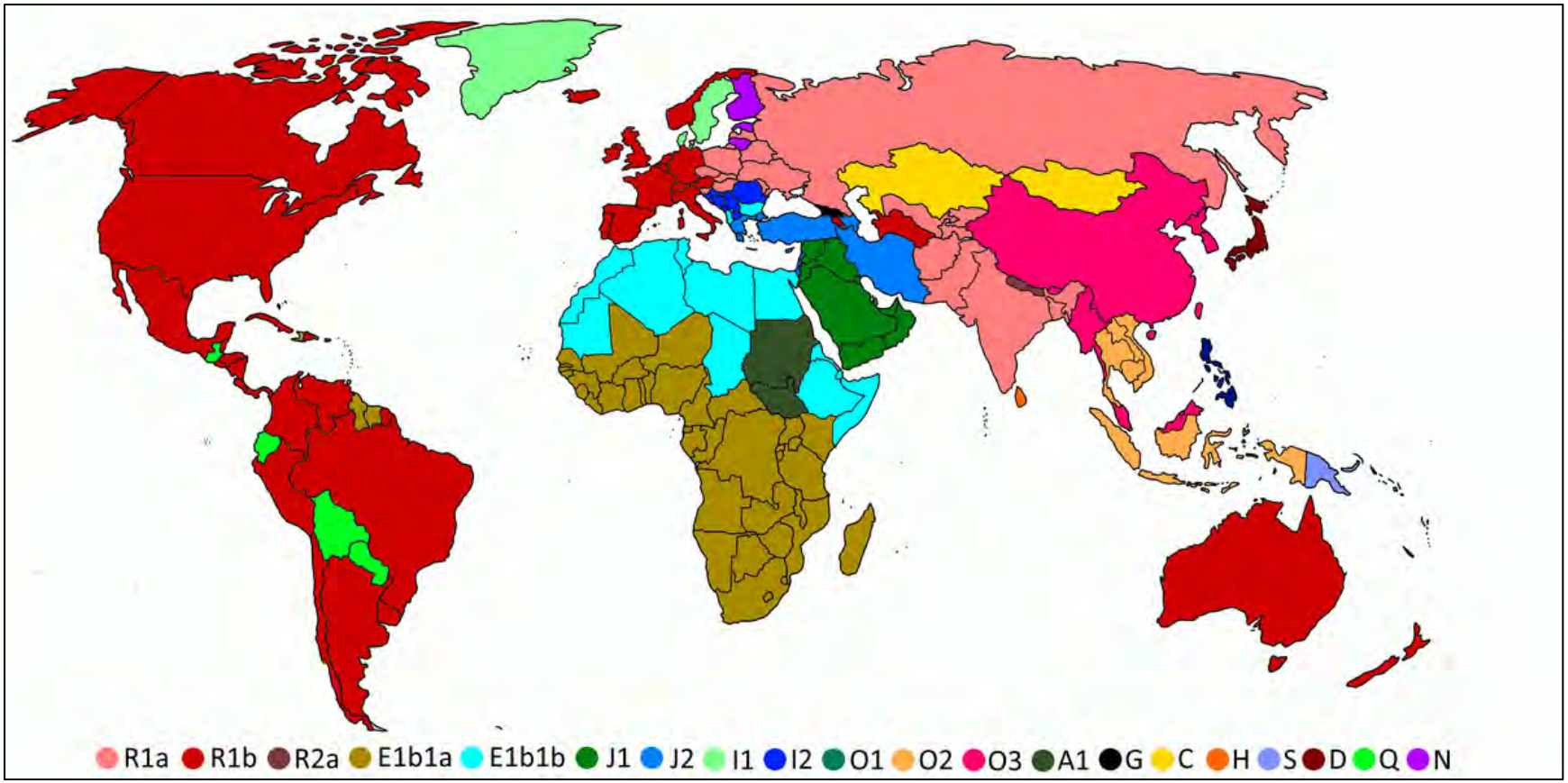
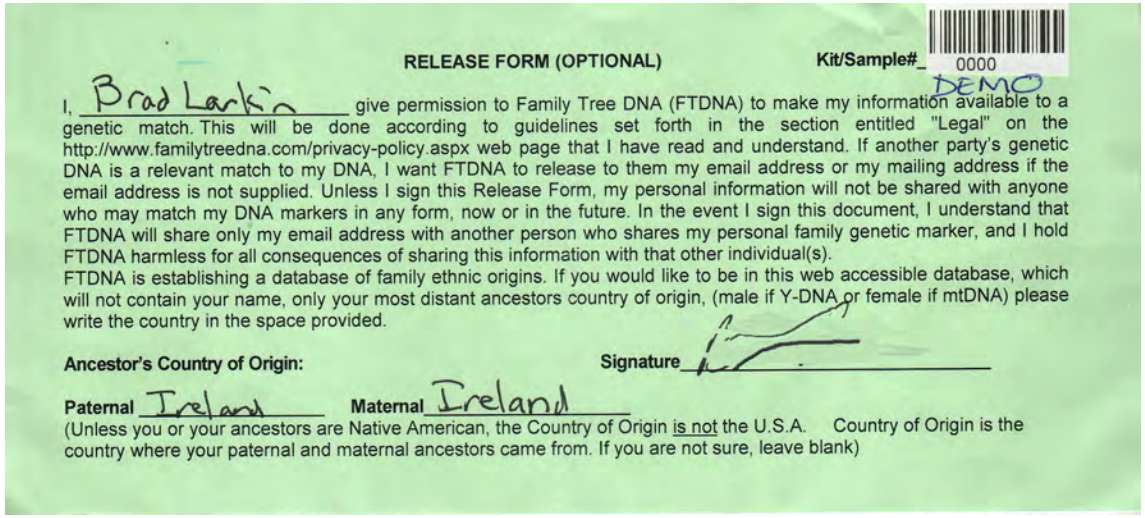


Image by Costa (2017) on genetics – Haplogroups message board on TheAPricity.com

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

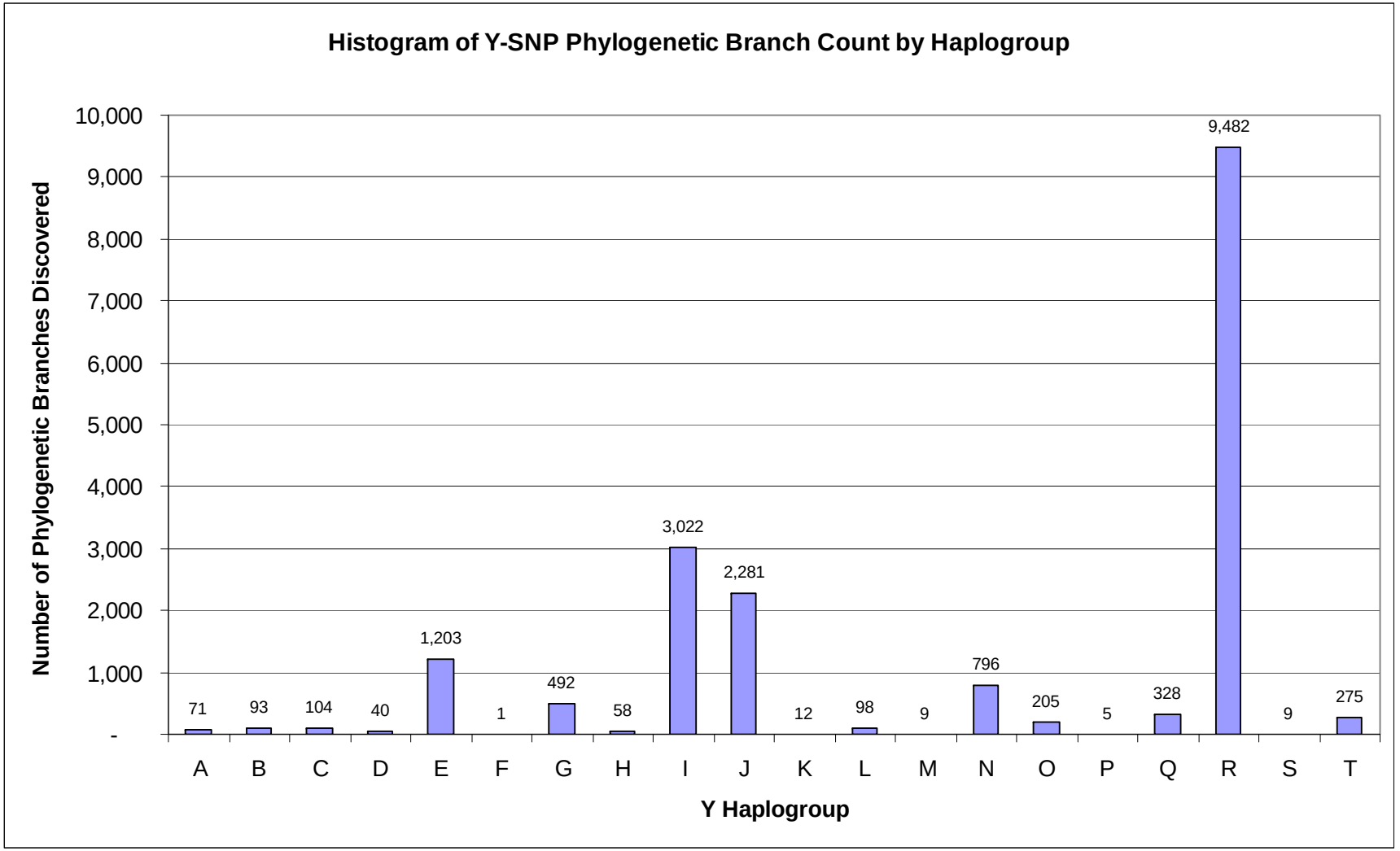
I, Brad Larkin give permission to Family Tree DNA (FTDNA) to make my information available to a genetic match. This will be done according to guidelines set forth in the section entitled "Legal" on the http://www.familytreedna.com/privacy-policy.aspx web page that I have read and understand. If another 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.

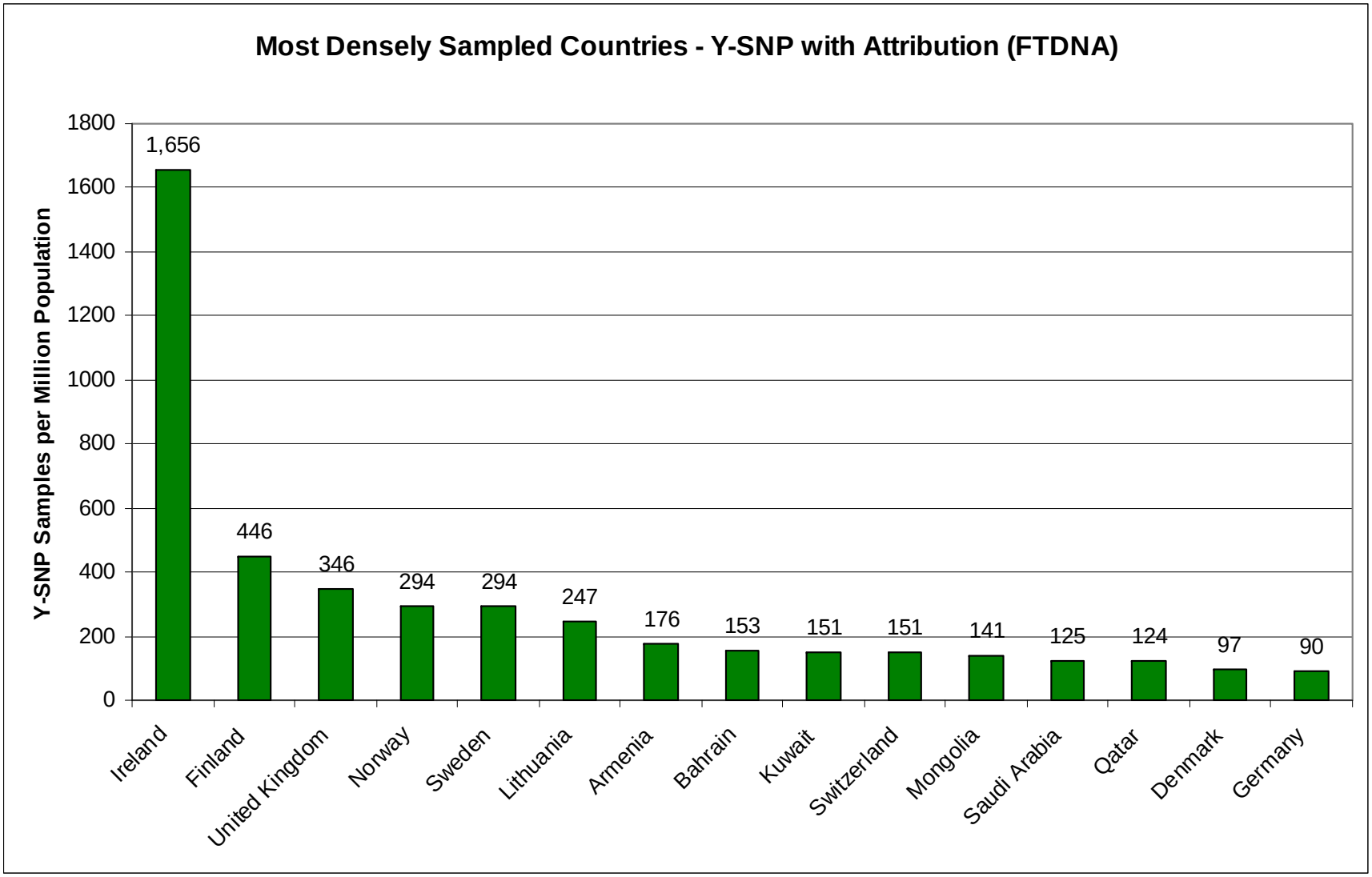
Ancestor's Country of Origin: Signature

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)



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



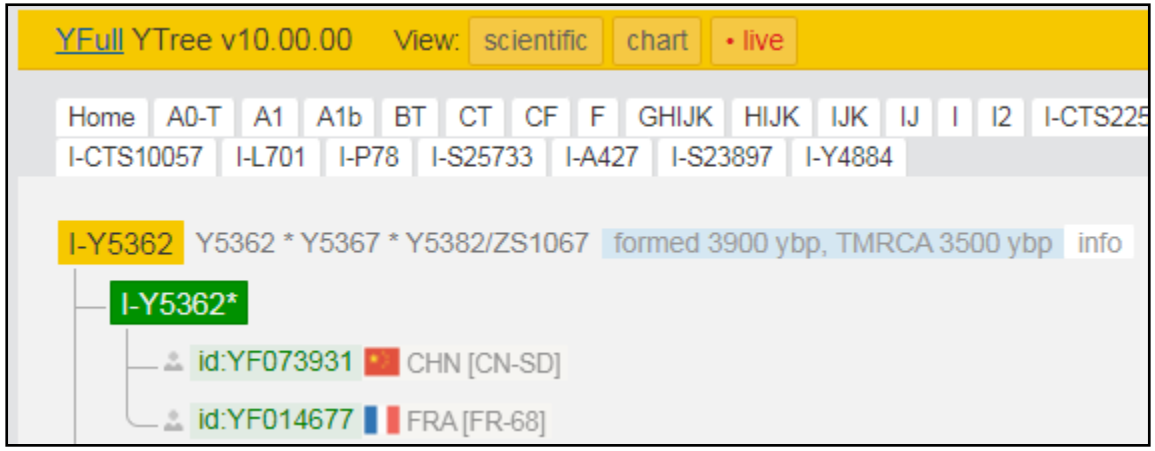
Population data from United Nations 2017 World Population Prospects on Wikipedia
https://en.wikipedia.org/w/index.php?oldid=879939654

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

Exercise – Estimating Ethnicity using Phylogenetic Tree

- Look at YFull tree for I-Y5362.
  - Has two examples
  - One from China.
  - One from France.



- If you did not have a preconceived notion and this was your terminal Y-DNA SNP, are you most likely to be Chinese or French?

Estimating Ethnicity: Phylogenetic Context

- Look at ancestral, sibling, and descendant branches if available.
  - Ancestral: I-Y4884 and I-S23897[1]
    - Samples with national attribution are all European.

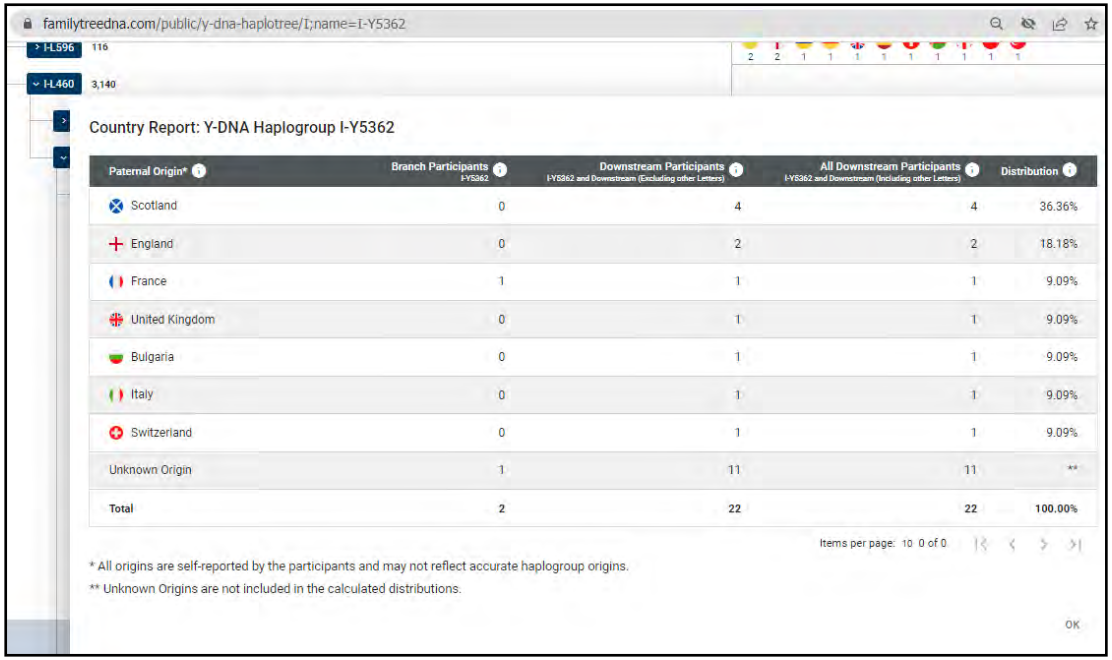
[1]YFull Haplogroup YTree v10.02.00 (06 April 2022)

Estimating Ethnicity: Descendants

- Descendant branches:
  - Again, all European, no East Asian examples.
- Conclusion
  - I-Y5362 likely indicates a European origin with a recent introduction to Asia.

Estimating Ethnicity: Multiple Tools

- FTDNA Tree has *Country Report*[1] which gives a nice tally of national attribution for marker and descendants.



[1]See FTDNA Public Y-DNA Haplotree Overview for instructions and more info.

When – Mutation Age

- I am much more interested in getting the sequence of mutations
  - Getting the pedigree markers in the right order.
    - This we can do by collecting facts (sampling)
  - Most chronological age estimating is based on modeling and assumptions.
    - Worthwhile, but not my particular interest at this point in the journey.

Review

- The Phylogenetic Tree
  - Tells us who our paternal ancestors were.
  - Let's us compare our result to another person based on terminal SNPs.
- NextGen Y-DNA 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

Connecting Paper & Genetic Genealogy

- Usually requires numerous DNA Project participants who have all had NextGen tests.
- Deep paper pedigree for at least some of the DNA test takers.

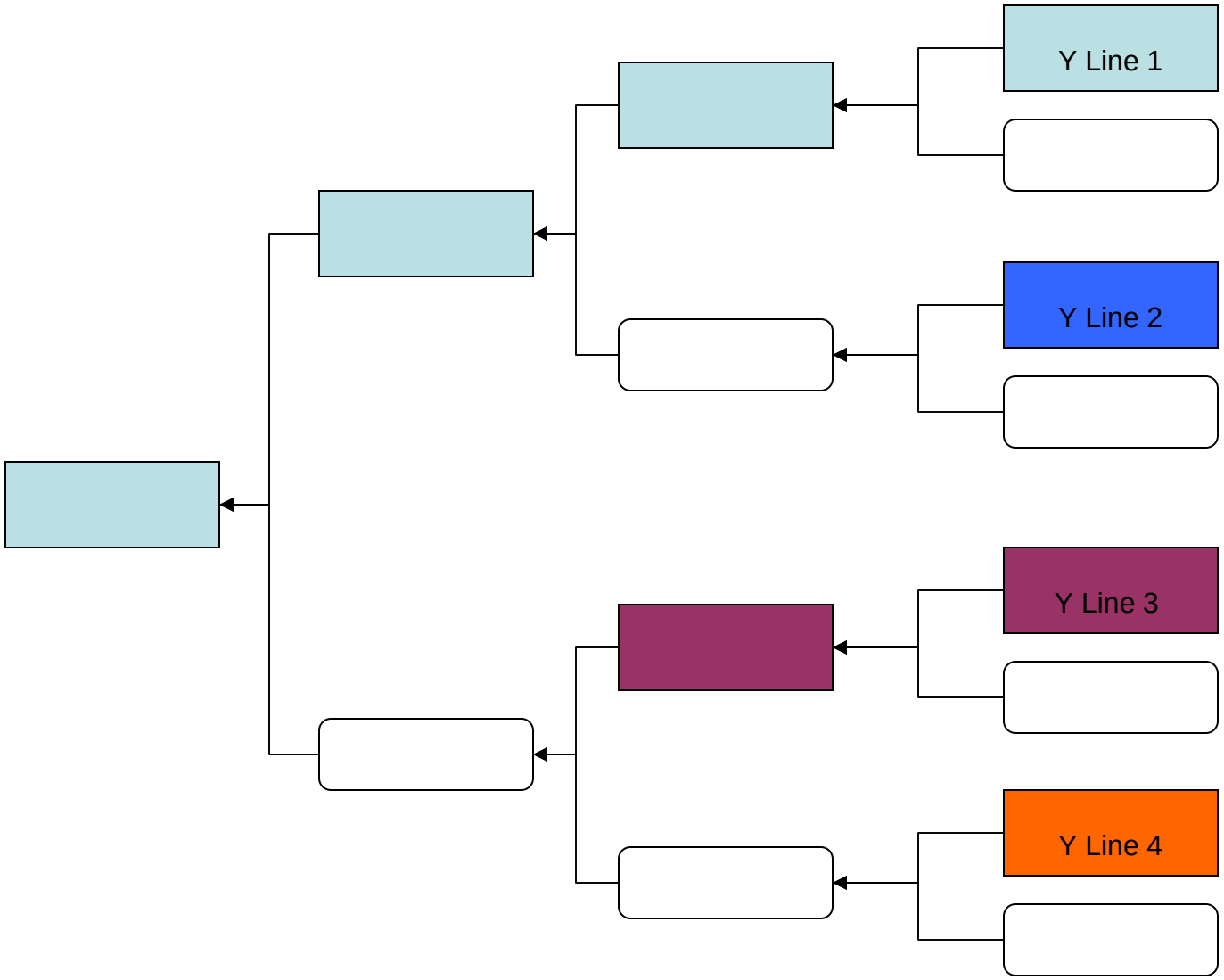
Flipping the Surname Identification Paradigm

- 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.

Most Frequently Asked Question

- *Give me contact info / help tracing my genealogy with Y-DNA matches*.
  - 22% of inquiries
  - Privacy regulations prevent revealing information not already provided by the testing laboratory.
  - Best answer:
    1. Get most complete Y chromosome test for yourself
    2. Recruit samples from others from geography / surnames most likely to be informative to your interest
       - e.g. Ancestral Parish Sampling presentation

Big Y all your Paternal Lineages



Low Resolution Fallacies

- STR and early SNP findings are low resolution Y-DNA results.
- Ethnic attributions to these markers like R-M269, M223, L21, etc occurred in Neolithic and Bronze ages.
  - Are meaningful only at the continental level.
  - Genetics attributed to historical events and populations is vastly over-simplified.

Discovery Time vs Ancestral Time

- Branch Level 40 ~ about 2200 years of ancestral time.
- M222 marker was only discovered in 2006
  - (16 years ago)
- A738 / BY198 discovered 2014
  - then only 2 examples
  - There are 82 as of 10/25/2022
- BY21680 correctly located in tree in 2018

Phylogenetic Ancestral Tree for marker [BY21680]

Tree Level	Marker / Branch Name	Alternative Names	Notes
48	M222	Page84 PAGES00084 USP9Y+3636 rs35720707 rs20321	
49	Y2605	FGC4124 rs1486849632	
50	S658	DF106 FGC4100 Y2841 rs747839864	
51	DF104	S661 Y2842 FGC4099 rs752415261	
52	DF105	S659 Y2843 FGC7927 rs755714899	
53	A18726		
54	BY198	A738	
55	BY20835	A15864	
56	BY20834	S27575 rs3097069	
57	BY21680	R-BY21680 A15870	

Royal Descent

- Celebrities & Royals not fond of revealing detailed results, but some are known.[1]
- Myths in family genealogy not uncommon
- Low resolution markers useful for disproving a relationship, but cannot prove one.
  - Low resolution markers like *R-M269* are shared by hundreds of millions of people.
- High resolution SNP marker matches needed.

[1] Y-DNA of the British Monarchy by Bradley T Larkin 2013

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.

Topic Sections

Introduction

Part I SNP Detection

Part II Trees and Interpreting SNPs

Part III Tips

Part III - Tips

- Unreliable Mapping
- InDel versus sliding point mutation
- Faux Mutations relating to hg38 Reference Sequence
- Palindromic regions
- X-Y crossover / recombination
- Comprehensive cross-reference of SNPs
- Relative Strengths of Various Published Phylogenetic Trees

InDel versus sliding point mutation

- Watch out for a deletion that has also been labeled separately as a point mutation.
  - *phylogenetic equivalent mutations*
    - *Only 1 piece of biology happened, a single deletion.*
  - Different labs software (and same lab using different technology) might enumerate this biology as:
    - T -> deletion at posn 3
    - T->C mutation at posn 3
      - If we think of the result as 'sliding' the succeeding sequence to the left.
    - C->G mutation at posn 4

Position	1	2	3	4	5	6
Ref	A	G	T	C	G	G
Biology	A	G	del	C	G	G
Result	A	G	C	G	G	?

Faux Mutations relating to hg38 Reference Sequence

- Happens where mutation occurred ancestrally in one of samples used to compose the hg38 reference.
  - Typically a position which had a mutation in haplogroup R1b
    - e.g. L20
- All samples outside of this lineage (R1b) should actually be expected to have the derived allele in hg38.
  - e.g. MF181966 hg38:6154452-A-del
    - Is really a faux mutation for what was actually an INSERTION that occurred anciently in haplogroup R1.
- Most common for newer laboratories that are relying entirely on human reference sequence.
  - And even more if they are using an older reference sequence.

Other Markers at this Position

- MF181966, described as a deletion at hg38:6154452
- But see ACT2970 – an insertion in haplogroup R1b.
- So the *human reference sequence* is carrying an [A insertion] at this position.
  - Which we now recognize happened in haplogroup R1 ancestor to the person whose DNA was used to construct the human genome reference.

DNA Marker Index data for Marker: 6154452-6154452 on Chromosome: Y

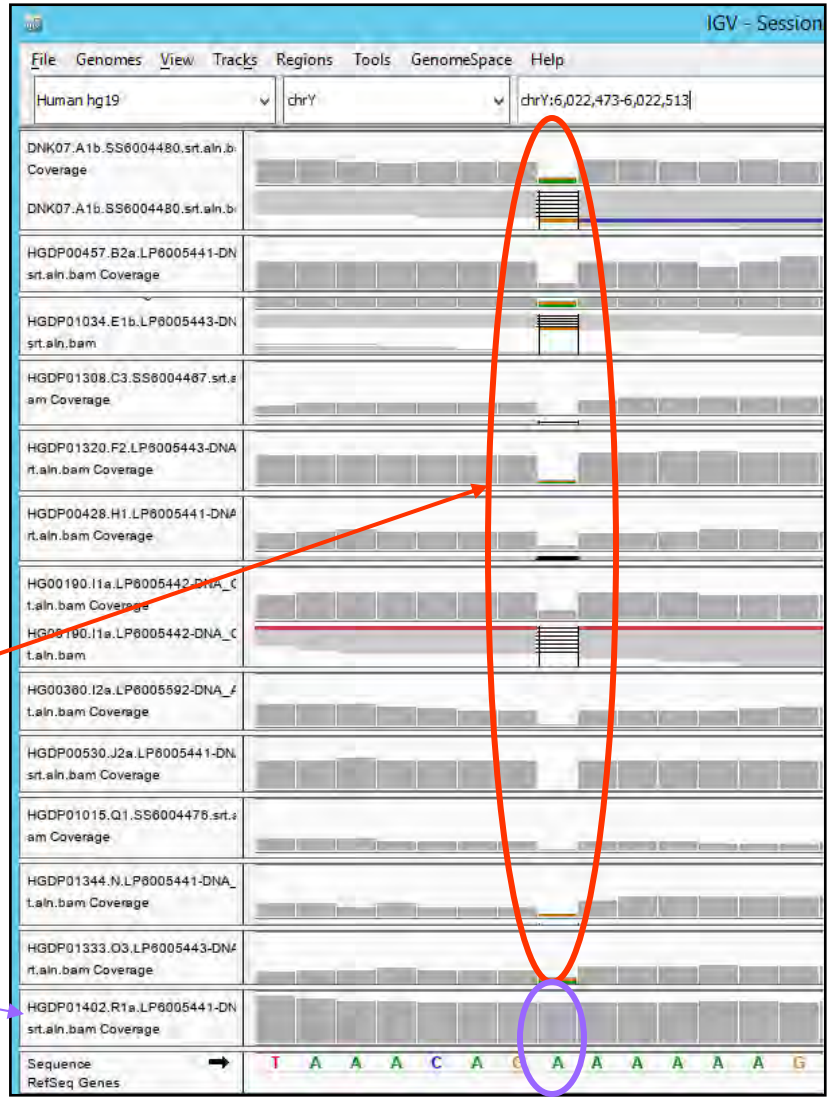
Marker Name(s)	Notes	Identification	Ancestral[7]	Derived[8]
R-ACT2970 ACT2970	Found in haplogroup R1, approximate to M173. Insertion of one base pair in a series of repeating A alleles, 5A->6A. hg38 Ref contains the extra A allele insertion and so all other haplogroups may report a deletion at this position. Could also be enumerated on FTDNA tests as deletion at hg38:6154457 (search from FT8154 backwards). Coincident with M173. View Pedigree	Ryan Lan-Hai Wei 2018	ins	A
FT184756	See also ACT2970 with different mutation at same position.	FTDNA 2019	A	C
O-MF181966 MF181966	Claimed to be in haplogroup O and other haplogroups, but probably a faux marker as hg38 Ref contains an extra A allele insertion. See also ACT2970.	23mofang 2020	A	del

==> When working outside of haplogroup R1b, use a position-searchable tool to look for other markers listed at this location:

- Genetic Homeland DNA Marker Index https://www.genetichomeland.com/dnamarkerindex
- ISOGG YBrowse http://ybrowse.org/gb2/gbrowse/chrY/?name=chrY%3A6154451..6154453

Confirming with BAM viewer

- Using a group of high coverage samples from SGDP project.
  - A set of samples covering many Y haplogroups.
- Looks like a deletion in almost every case.
- Except: The R1a sample has the A allele.



Palindromic Regions & Duplications

- Palindrome
  - Same forward as reverse[1]
    - *RADAR*
    - *Madam, I'm Adam*
- DNA palindromic sequences tend to chemically attach to themselves in a loop.
  - Loops can cause large insertions, deletions, and shifts in chromosome position as well as duplicate sequences.
  - Creating *palindromic arms*

[1] Thomas Krahn, *Matching Multicopy Y-STR Markers in Closely Related Individuals* presentation at FTDNA 3rd International Conference on Genetic Genealogy 2006

NextGen Sequencing & Palindromes

- Palindromic arms more problematic in the STR-era of genetic genealogy.
- Because NextGen sequencing ‘cuts up’ the strands into small pieces and matches to a reference sequence
  - The biological position on the palindromic arm is irrelevant.
- May become a bigger challenge with longer read length sequencing coming.

X-Y crossover / recombination

- *pseudoautosomal regions* PAR1 and PAR2 are portions of the Y chromosome which undergo recombination with the X chromosome
  - PAR1 spans hg38 Y position 10,001 to 2,781,479
  - PAR2 spans hg38 Y position 56,887,903 to 57,217,415
- Some researchers believe they should not be relied upon for patrilineal ancestry markers.
- Other labs are reporting & using mutations within PAR1 & PAR2 in their trees.

MSY Complexity

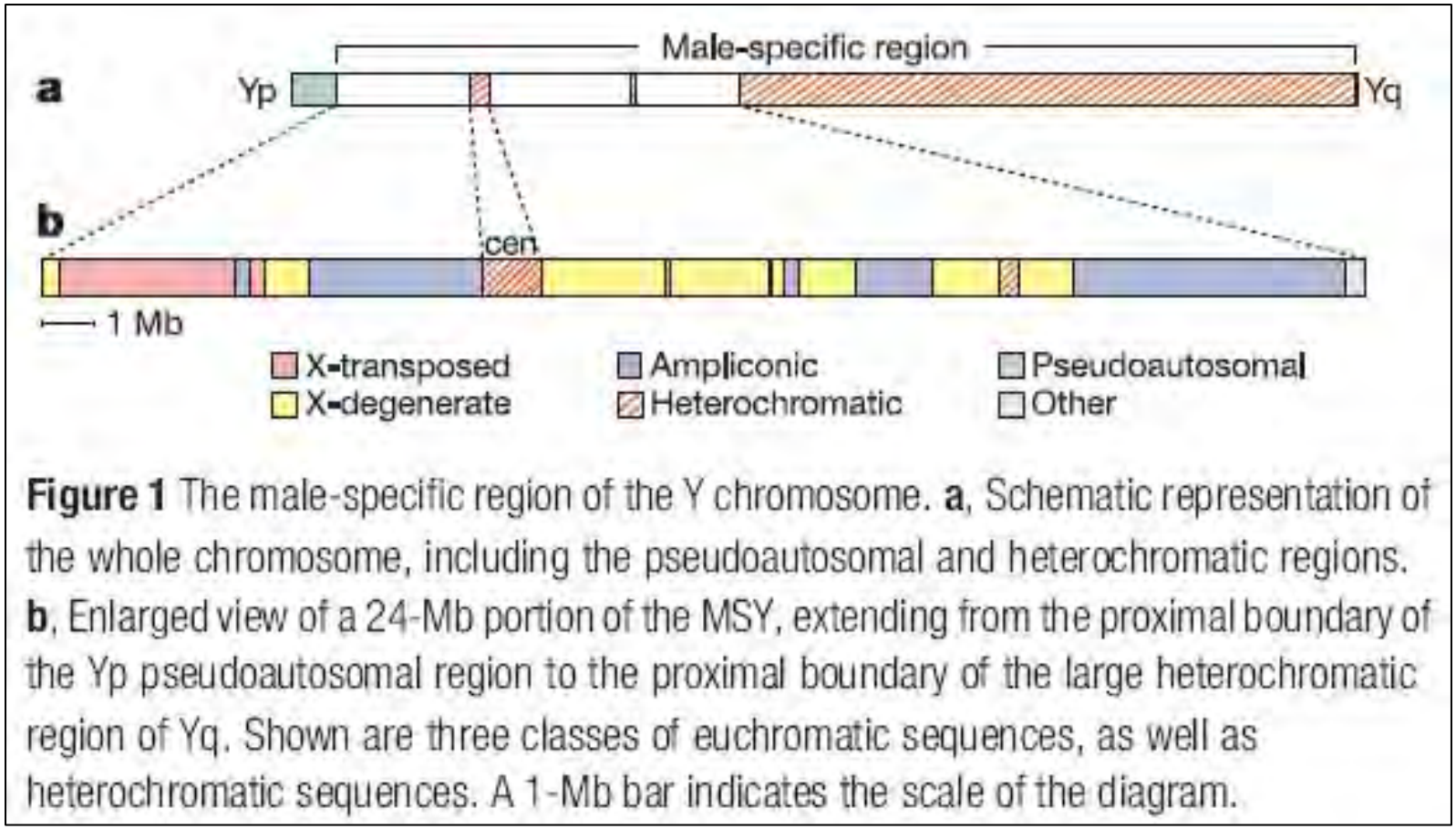


Figure 1 The male-specific region of the Y chromosome. **a**, Schematic representation of the whole chromosome, including the pseudoautosomal and heterochromatic regions. **b**, Enlarged view of a 24-Mb portion of the MSY, extending from the proximal boundary of the Yp pseudoautosomal region to the proximal 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)

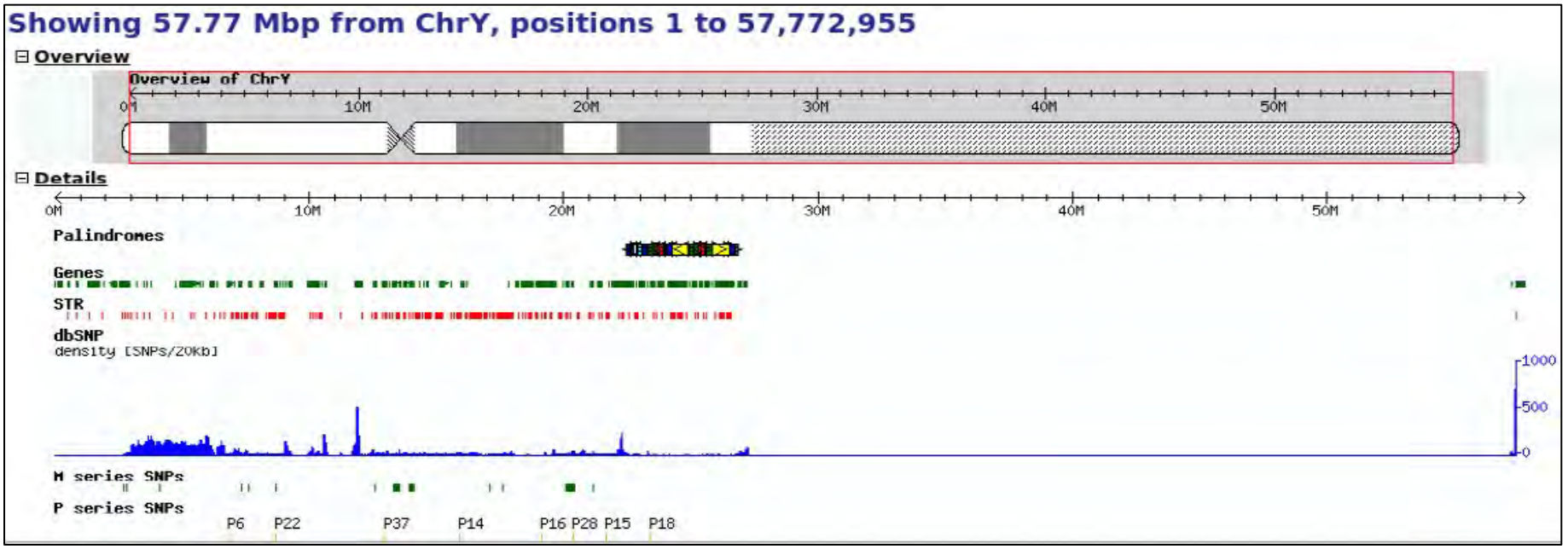


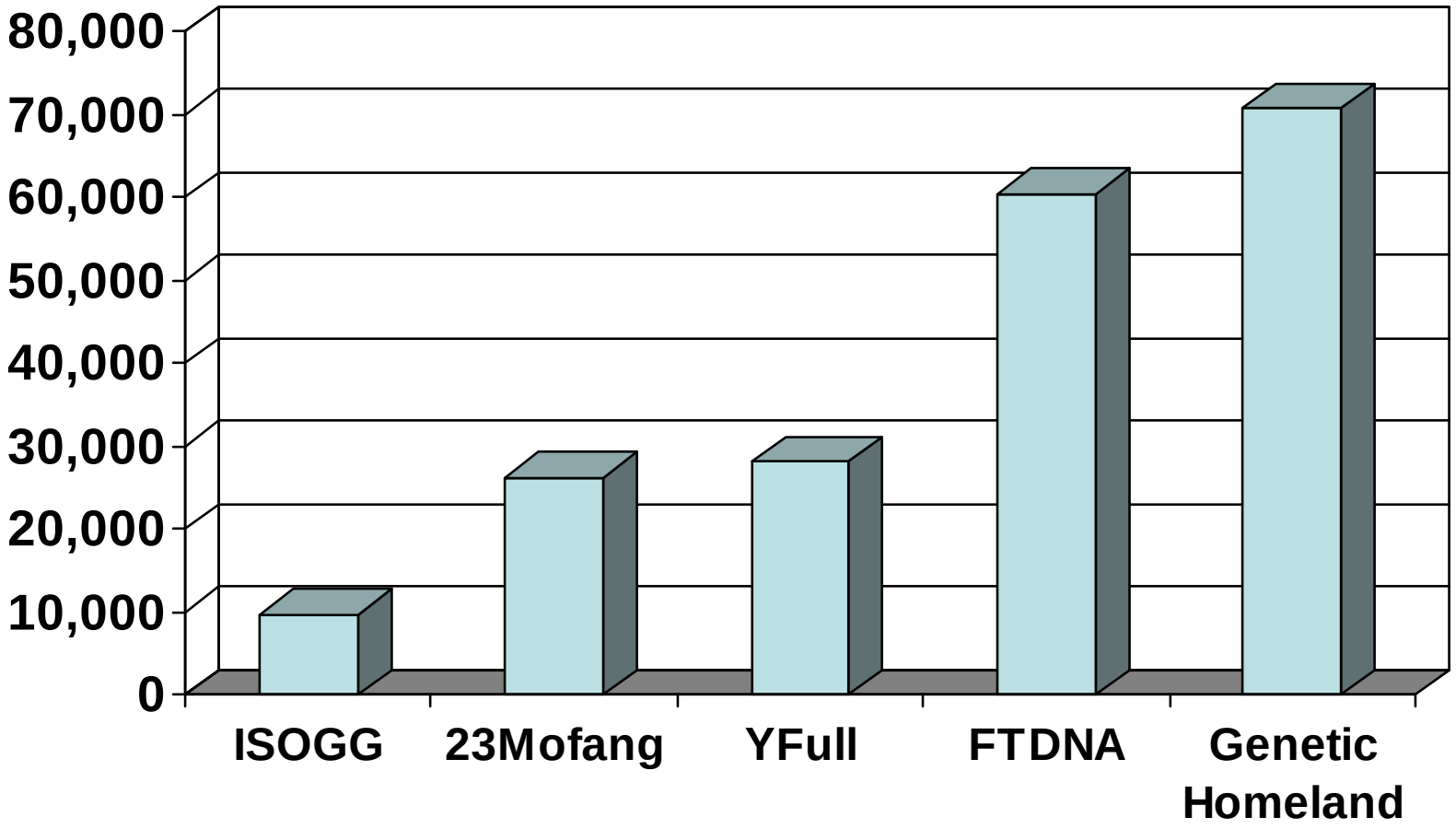
Image adapted from Thomas Krahn, *Walk On Y Project* presentation at FTDNA *4th International Conference on Genetic Genealogy* 2007.

Relative Strengths of Various Published Phylogenetic Trees

- FTDNA
  - Has the most Y samples with SNP testing and growing.
    - New SNPs every week.
  - Public Haplotree has best public linkage of surname, country, and Y-DNA tree branches.
    - About 60,000 branches
  - Customer base skewed towards U.S., Western Europe, and Middle East.
    - Best in R1b, I, J, E1b.
  - Low resolution, legacy data: many customers only have STR results with minimal SNPs
- YFull
  - Not a lab but rather a results analysis service.
  - Has oldest well-maintained public tree.
    - About 28,000 branches
    - Superior collection of R1a and Asia samples.
    - Incorporates a time origin estimate at each branch.
- ISOGG
  - Used by scientific / academic community
  - Tools and tree have not kept pace, last full update was 2018
    - About 9,700 branches
- YSeq
  - Has internal tree where product packages developed with independent researchers.
    - So it can be the best in niches. Uses YFull tree in many cases.
- 23mofang
  - Largest East Asian laboratory with many more samples in haplogroups O, C, Q.
  - Not easy to navigate as site is not published in English
    - About 26,000 branches
- GeneticHomeland.com
    - About 70,400 branches

Y Phylogenetic Tree Branch Count

Estimated as of 10/21/2022



Need for a comprehensive cross-reference of SNPs

- DNA Marker Index on GeneticHomeland.com
- dbSNP

GeneticHomeland.com DNA Features

- Cross references SNP label names across all labs along chromosome positions.
  - Largest SNP name cross reference.
    - Over 2.4 million labels as of 10/19/2022.
- Integrates lab trees as well as scientific research papers.
  - Has the most branches.
    - 70,425 Y-DNA tree branches as of 10/19/2022.
- Intended for individual pedigree and head-to-head comparison for divergence of two markers.
  - Not designed to be 'bushy' tree display.
  - Generally does not have specific samples for each branch.
- Mapping of geographies for all descendant branches for a given terminal SNP.
  - Includes Ancient Y-DNA dataset of over 4,800 records.
  - Can compare geographic distribution of up to three different SNPs at same time.

Recap

Part I SNP Detection

Part II Trees and Interpreting SNPs

Part III Tips

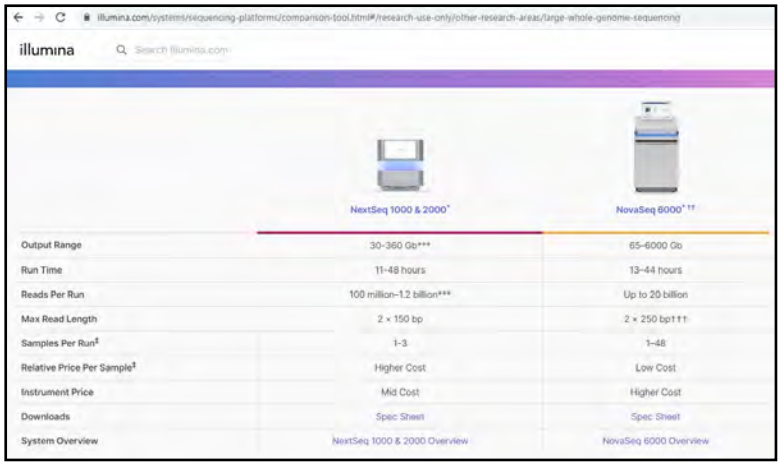
Topics

- Part I – SNP Detection
  - Biology & Sequencing
  - Alignment of Reads
  - Types of Files
  - BAM Viewing
- Part II – Trees and Interpreting SNPs
  - Nomenclature of SNPs in Y-DNA Genetic Genealogy
  - Phylogenetic Tree
  - Y-DNA Tree Backbone
  - Coincident mutations on same branch
  - Terminal SNP
  - Novel Variants / de Novo Mutations
  - Geography & Ethnic Attributions
  - Connecting Paper & Genetic Genealogy
- Part III - Tips

Sequencers & Read Lengths

- 150 base pair read length
  - Illumina *HiSeq*
  - Illumina *NextSeq*
- 250 base pairs
  - Illumina *NovoSeq*
- 300 base pairs
  - Illumina *MiSeq v3*
- 400 base pairs
  - Ion Torrent *PGM 314 Chip*
  - Roche *GS Junior 1 PTP*
- 4,500 base pairs
  - PacBio *RS II* using SMRT Cell
- 6,000 base pairs
  - Oxford Nanopore *GridION*

- Increasing the sequencing read length is probably the biggest change coming in next 5 years of Y-DNA genetic genealogy.



Illumina comparison of NextSeq 1000 & 2000 with NovoSeq 6000 sequencers

PacBio RS II

Sequencer data from https://genohub.com/ngs-instrument-guide/.

Future of Y Reference Sequences

- Haplogroup-specific, experimental reference sequences published in 2021 & 2022
  - GenBank entry CP086569.1
    - First released Telomere-To-Telomere (T2T)[1] sequence of a complete Y chromosome
      - 62,456,832 base pairs sequenced, published Nov 16, 2021
      - Corrected by Thomas Krahn ,YSeq, Jan 14 2022
    - Sample from 1000 Genomes Project participant *NA24385*[2]
      - Ashkenazi male, descendant of Y-DNA haplogroup J1-M267
    - GenBank entry CP086569.2
      - » Improvements over version 1
      - » 62,460,029 base pairs sequenced, published Apr 04, 2022
  - GenBank entry CM034974.1
    - Sample from 1000 Genomes Project participant *HG01243*[3]
      - Puerto Rican male, descendant of Y-DNA haplogroup R1b-DF27[4]
- Aligning longer read length sequencing onto haplogroup-specific reference sequences is probably biggest change coming in next 10 years of Y-DNA genetic genealogy.

[1] Telomere-to-Telomere (T2T) Consortium, https://sites.google.com/ucsc.edu/t2tworkinggroup/
[2] Nurk et al (2022) The complete sequence of a human genome. Science 2022 Apr;376(6588):44-53, doi:10.1126/science.abj6987
[3] Zimin et al (2021) A reference-quality, fully annotated genome from a Puerto Rican individual, doi10.1101/2021.06.10.447952
[4] YDNA-Warehouse, Telomere-to-Telomere Y Chromosome Experiments, https://ydna-warehouse.org/t2t-experiments

Audience Questions?

Intermediate and Advanced Y-DNA Topics

Brad Larkin

Genealogical Forum of Oregon
Advanced DNA Special Interest Group

Questions For You

1. InDel: Using either of these two examples of DNA marker information for Z16279, is this mutation?
   a. Insertion
   b. Deletion

Z16279 Details

Name:	Z16279
Type:	snp
Source:	indel
Position:	chrY:21813319..21813320 (+ strand)
Length:	2
allele_anc:	del
allele_der:	ins
comment:	downstream of DF21
count_derived:	0
count_tested:	0
isogg_haplogroup:	R1b1a2a1a2c1g (not listed)
load_id:	Z16279
mutation:	del to T
primer_f:	TBD
primer_r:	TBD
ref:	Alex Williamson (2014)
ycc_haplogroup:	R1b-DF21 (not listed)
yfull_node:	Not found on the YFull Ytree
primary_id:	305815
gbrowse_dbid:	chrY:database

DNA Marker Index data

Marker Name(s)	Notes	Identification	Ancestral	Derived	Chromosome	Position (hg38)	Position (hg19)
R-Z16279 Z16279 FGC32818	Found in haplogroup R1b under DF21. Also enumerated as hg38:21813319-GT-GTT.	Alex Williamson 2014	ins	T	Y	21813320	23959467

Questions For You

2. In this example with L192, what do the decimal suffixes .1 and .2 indicate?

DNA Marker Index data for Marker: L192 on Chromosome: Y

Marker Name(s)	Notes	Identification	Ancestral	Derived	Chromosome	Position (hg38)	Position (hg19)
R-L192 L192.1 L192 rs113781025	Found in haplogroup R1b-L513. See also L192.2 n haplogroup J2a. Phylogenetic Parent: BY11127 / FGC49373 View Pedigree View Map of descendants Phylogenetic Children: BY11150 NIH DBSNP b153 GRCh38p12 ClinVar? Position hg18: 6813342	Thomas Krahn, FTDNA	C	T	Y	6885301	6753342
L192 L192.2 rs113781025	Found in haplogroup J2a. See also L192.1 in haplogroup R1b. Phylogenetic Parent: SK1382 / Y14314 View Pedigree View Map of descendants Phylogenetic Children: FT203595 FGC30635 FT258230 NIH DBSNP b153 GRCh38p12 ClinVar? Position hg18: 6813342	Thomas Krahn, FTDNA	C	T	Y	6885301	6753342

Questions For You

3. In this example is Z39589, what do the two suffixes with **underscore** characters mean?

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

Marker Name(s)	Notes	Identification	Ancestral	Derived	Chromosome	Position (hg38)	Position (hg19)
R-Z39589 Z39589.1 Z39589 Z39589_2	Major branch in haplogroup R1b under L21 and DF13 with many descendant branches from Ireland. Deletion of 18 base pairs. Note that hg38 position description not linear translation from hg19 lifover map. Also enumerated as hg19:4 439911-TGCAGCTTCACTCCTGAGG-T. Phylogenetic Parent: DF13 / CTS241 / S521	Alex Williamson 2016	GCAGCT TCACTC CTGAGG	del	Y	4571907	4 39912
R-Z39589_1 Z39589_1	Major Haplogroup: R. Deletion of 18 base pairs. Z39589 hg38 position enumerated at hg38:4571871 based on liftover mapping. Later versions of FTDNA Big Y-700 test use this hg38 position. But instead See also Z39589_2 for correct hg38 position.	Alex Williamson 2016	GCAGCT TCACTC CTGAGG	del	Y	4571871	4 39912

Questions For You

4. In these phylogenetic tree examples, which of these markers is shown as **coincident** with M222?

   a. Z2965

   b. Z2963

   c. A7362

Questions For You

5. Using the illustration, which of these is a synonym for marker M222?
   a. Z2965
   b. FTC311
   c. PAGES0084

DNA Marker Index data for Marker: M222 on Chromosome: Y

Marker Name(s)	Notes	Identification	Ancestral	Derived	Chromosome	Position (hg38)	Position (hg19)
R-M222 M222 Page84 PAGES00084 USP9Y+3636 rs35720707 rs20321	Major Haplogroup: R. Sometimes called Northwest Irish, concentrated in Ireland and western Scotland. Associated with Niall of the Nine Hostages and Ui Neill clans. Britain's DNA labeled this branch: Ancient Irish. Phylogenetic Parent: Z2965 / S6155 View Pedigree View Map of descendants Phylogenetic Children: FTC311 Y2605 Z2959 NIH DBSNP b153 GRCh38p12 ClinVar? Position hg18: 13411808	Sun et al 1999	G	A	Y	12790481	14902414

Questions For You

6. Which of the following could be reasons that the Y-DNA Terminal SNP reported by your lab could change:
   a. Trick question – your Terminal SNP can never change once your sample has been completely processed and resulted at the laboratory.
   b. You sent two different kits from the same participant to the same laboratory and results could vary by kit.
   c. You upgraded from STR to Big Y-700 test.
   d. New samples are collected from other descendants of your earliest known ancestor, creating new branches.