

YDNA Analysis Tools

<http://www.rcasey.net/present>

By Robert Casey

February, 2019

Types of YDNA tools

- YSTR & YSNP reports (database of sorts)
- YSTR match screens (best FTDNA tool)
- YDNA haplotrees (FTDNA, BigTree & YFULL)
- Haplogroup EXCEL YDNA spreadsheets
- Big Y screens (Admin NGS/WGS EXCEL)
- Y500 spreadsheet summary (no FTDNA report)
- YSEQ YSNP testing spreadsheet (relevant YSNPs)
- YSNP prediction tools (NevGen & EXCEL macros)
- Charting tools (SAPP & manual)

YSTR reports

- This is primarily a database of YSTRs
- It also includes the currently “terminal” YSNP
- Admins group testers together but many are just not YSNP tested enough or some groupings can be several thousand years old
- Does not show known relevant testers from other projects or all relevant YSNP results
- Missing 25 to 50 % of testers that are private

Y SNP reports

- This is primarily a database
- Has no grouping function
- Missing significant Y SNP data from Big Y (bloated with too much meaningless data)
- Does not include significant Y SNP testing from YSEQ and Full Genomes
- Missing Y SNP testing from non-project members (generally 25 to 50 % of testers)
- Missing 25 to 50 % more from private testers

YSTR Match Screens

- Can be a very power tool for analysis
- Used to recruit new project members
- If your surname cluster is genetically isolated, can be a very useful and accurate tool
- For the vast majority of testers that are not genetically isolated, 50 to 90 % error rates at even 67 markers vs. advertised 5 % errors
- With some genetic filtering – becomes useful

Genetically Isolated Cluster – Part 1

67 MARKERS - 21		
Genetic Distance †	Earliest Known Ancestor	Y-DNA Haplogroup
0	Jesse E. Casey, b. 1797, Georgia, died 1863 Ark	R-M269
1		R-M269
1	Pleasant Casey, b. 1810, Roane Co., TN	R-M269
1	Pleasant Casey, b. 1810, Roane Co., TN	R-FGC5639
1	William Harvey, b.c. 1740, d. after 1780	R-M269
2	Hilliard John Casey, b. 1813 (SC) and d. 1870 (MS)	R-FGC5639
2	John Meredith, Virginia b. 1806 d. 1884	R-M269
2	Henson Casey, b. 1834, Roane Co, TN	R-M269
2	John Casey b.SC d 1849 IL	R-M269
2	Moses Casey, r. SC	R-M269

Genetically Isolated Cluster – Part 2

2	Abner Casey, b. 11 July 1786, Spartanburg, SC	R-L226
3	Frank Casey	R-M269
3	Love Casey, b. 1823 and d. 1890	R-M269
3	Randolph Casey, b. 1738	R-M269
3	John Casey, b. 1782, d. 1856	R-M269
3	James Hill Casey, b. 1813, Spartanburg, SC	R-L226
4		R-M269
5	Levi Casey	R-M269
6	Dempsey Casey Sr, b. 1762 and d. 1814	R-M269
7	William Kersey, b. 1722 and d. 1764	R-FGC5647
7	Thomas Carey, b. 1717	R-M269

Genetically isolated cluster

- Works due to a very strong eight mutations from the L226 signature (making it genetically isolated)
- The observed NPE rate is very reasonable
Casey (15), Kersey (1), Meredith (1), Hanvey (1)
and Carey (1) – 79 % Casey testers
- NPE rates below 50 % are reasonable (over 30 to 40 generations, there will be a lot of NPEs)
- Haplogroups are consistent – L226, FGC5639 and FGC5647 (do not suggest older than 1000 AD)

Erin Garcia – Murphy or Mahony ?

Genetic Distance ↑	Earliest Known Ancestor	Y-DNA Haplogroup
3	Michael Mahoney; b. ca. 1800; of Kilbehenny	R-M269
3	James Mahony m 1832	R-DC199
3	Daniel Mahony b.c. 1835 Carrick	R-M269
4	Michael Mahoney; b. ca. 1800; of Kilbehenny	R-M269
4	William O Mahony, Ballyfaskin	R-L226
4	John O'Mahony b. 1807, Ireland	R-L226
5	O'Mahony, Mitchelstown	R-L226
5		R-L226
5	John Donoghue b.abt.1800, Knockane,Toomevara,CoTip	R-L226
5	John Hannon. b. abt 1810 and d. abt 1885	R-BY4100
5	Michael Mahony b 1823	R-L226

Erin Garcia – Part 2

Surnames/Haplogroups not reasonable

Genetic Distance ↑	Earliest Known Ancestor	Y-DNA Haplogroup
6	Thomas Jefferson Noland B:1826 TN D:1887 Conway,AR	R-L226
6	Timothy O'Neill b.1803 / Kilfintinan Clare	R-DC30
7		R-M269
7	John Neale, b.c.1687 Ireland, d. 1733 CharlesCo MD	R-L226
7	John Coy, 1710 - 1776	R-L226
7	Henry Anderson bn 18 jly 1716 d 1790 Kilkenny IRE	R-DC15
7		R-M269
7	William MacCraw, d. c. 1752, Virginia thru son 2	R-M269
7	Charles Hannigan (c1803-1882) Rathdowney, Ireland	R-M269

Worst Case Scenario Donohoe/11144

67 MARKERS - 677 - MATCHES



Worst Case Scenario

- Donohoe/11144 has only two 67 marker matches of his surname at genetic distance of three
- But there are 94 testers that do not match his surname at a genetic distance of three
- This results in 97 % NPE rate even at a genetic distance of three at 67 markers
- But at 111 markers, the NPE rate drops to 40 % for a genetic distance of seven but jumps to a NPE rate of 93 % at GD = 10 (FTDNA matches)
- But all matches are positive for L226 branches (which is 1500 years old)

Y500/Y700 YSTRs

- There is no report with Y500 values
- Requires haplogroup admins to track and post Y500 results via EXCEL spreadsheet
- Due to smaller sample sizes, Y500 are not currently very useful for charting
- Y500/Y111 ratios vary a lot – so does help understand degree of relatedness better
- Full of too many very slow mutating markers

Conclusion YSTR Matches

- The accuracy of FTDNA's match lists has been proven to have very low accuracy over many haplogroups (this is the rule not the exception)
- Testers should always look at reasonable NPE rates as well as looking at YSNP branches that are over 1,000 years old that match (common sense)
- With genetic distance filters lower than FTDNA recommendations and with higher Y111 resolution, you can still find reasonable matches

Action Items for YDNA reports

- Projects are the lifeblood of analysis, not only should you join your surname and haplogroup projects – but encourage your matches as well
- Privacy settings just remove valid data from your database – you should enable “opt in to sharing” and encourage your matches to enable as well
- No matter how small the scope of your interests are – track your matches in EXCEL
- Use the match YSTR match list pop-up menu to contact your matches directly - network

Haplogroup Trees - YSNPs

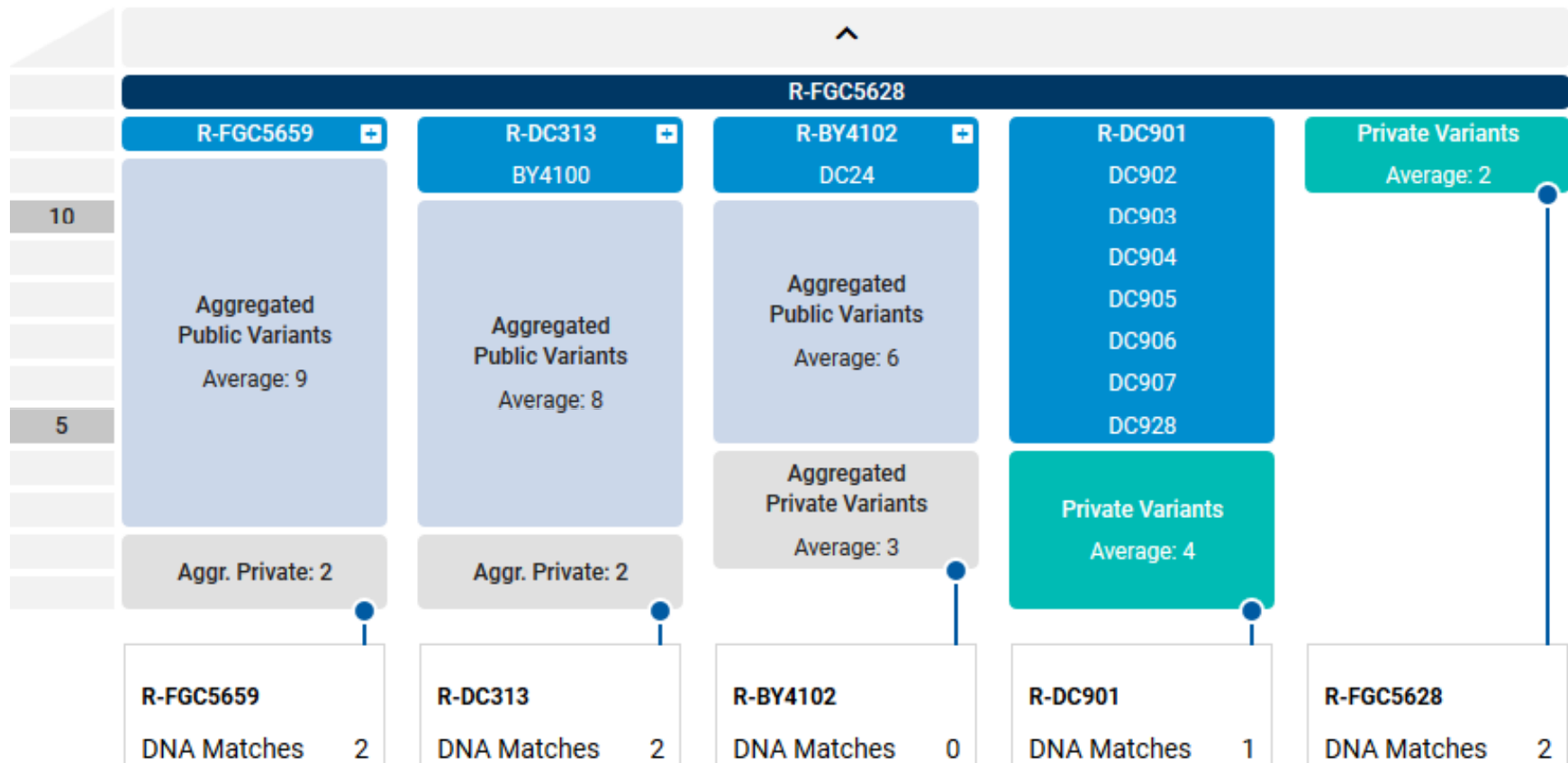
- FTDNA now has three haplotrees: (on your MyFTDNA – click on YDNA haplogroup button)
- If you are tested with Big Y500 – click on Big Y Block Tree for a lot of detailed summaries
- Public version – Haplotree summary shows how many are confirmed to be positive (check how many missing)
- Big Tree and YFULL are enhanced third party versions of the FTDNA Big Y block tree (but smaller numbers)
- Some haplogroup projects have their own their haplotrees (can have key differences)

FTDNA Haplotree

Tested Positive	Tested Negative	Presumed Positive
Downstream	Presumed Negative	Test in Progress

BY4102 More...	R-BY4102
DC40	R-DC40
DC936	R-DC936
DC18 More...	R-DC18
A6097 More...	R-A6097
DC50 More...	R-DC50
DC55 More...	R-DC55
DC245	R-DC245
DC246 More...	R-DC246
DC901 More...	R-DC901
DC28	R-DC28
FGC12295 More...	R-FGC12295
FGC12290 More...	R-FGC12290
FGC12296 More...	R-FGC12296
FGC12293 More...	R-FGC12293
DC201	R-DC201
DC360	R-DC360
DC205	R-DC205
DC435 More...	R-DC435
DC437 More...	R-DC437
DC199	R-DC199
DC464 More...	R-DC464

FTDNA Big Y Block Tree



Action Items

- Get familiar with your part of the haplotree
- Add YSNP information from YSNP reports & Big Y screens to your EXCEL summaries
- From YSTR reports – determine your closest matches that have been Big Y500 tested
- Track not only relevant “current” terminal YSNP but also track known negative branches that are downstream (negatives are key for charting)
- Let haplogroup admins do the heavy lifting but do learn as much as possible about Big Y results

Your personal database - EXCEL

- YSTR and YSNP reports are very crude forms of your database (groupings are marginal)
- In order to get better analysis and data collection – you (or your admin) need to track your data via EXCEL spreadsheets
- When you find valid matches – get them to join your projects (also track them yourself)
- Your matches should be encouraged to enable “opt in to sharing” to maximize access to data

Basics to include your EXCEL

- Of course track as many relevant testers as possible (both YSTR and YSNP data)
- Surnames are very important to track (extract surname from male Earliest Known Ancestor or Donor name)
- Kit numbers are key (do not create your own IDs as this defeats the purpose of sharing)
- Keep track of which project that you found the data (otherwise you may not catch upgrades)
- Country of Origin can be useful – or misleading
- Tracking source of YSNP testing is useful – NGS (Big Y, FGC, YSEQ, Dante) Individual YSNPs (FTDNA & YSEQ), SNP packs (FTDNA & YSEQ), Nat Geo

Basics to include your EXCEL

A	B	D	E	F	I	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB
f878164	Kane	R1b-M269	zzCountry	R-L226	BLR	13	24	14	11	11	14	12	12	11	13	13	16	17	8	9	11
f880027	Montgomer	R1b-M269	zzCountry	R-L226	BLR	13	24	14	11	11	14	12	12	12	13	13	16	17	8	9	11
f1N41478	Connor	R1b-L21>Z253>L226>DC889	India	Conner	Indiv	13	24	15	10	11	14	12	12	12	13	13	16	17	8	9	11
fB382283	Fineman	R1b-M269	zzCountry	R-L226	BLR	13	24	14	12	11	15	12	12	11	13	13	17	17	8	9	11
f537546	Ahearn	R1b-M269	zzUnknown	R-L226	BLR	13	24	14	11	11	14	12	12	11	13	13	17	17	8	9	11
fN201977	Alexander	R1b-M269	zzUSA	R-L226	BLR	13	24	14	11	11	14	12	12	12	13	13	16	19	8	9	12
fB199399	Anderson	R1b-L21>Z253>L226>DC15	Ireland	R-L226	NGS	13	24	14	11	11	14	12	12	12	14	13	16	17	8	9	11
f16274	Anderson	R1b-L21>Z253>L226>DC15	Ireland	R-L226	NGS	13	24	14	11	11	14	12	12	11	14	13	16	17	8	9	11
f750701	Balise	R1b-M269	zzUSA	R-L226	BLR	13	24	14	10	11	14	12	12	11	13	13	17	17	8	9	11
f207264	Barding	R1b-M269	zzCountry	R-L226	BLR	13	24	14	11	11	14	12	12	12	13	13	16	17	8	9	11
f767419	Barrett	R1b-L21>Z253>L226>DC29	zzUnknown	R-L226	L226	13	24	14	11	11	14	12	12	11	13	13	15	16	8	9	11
f92065	Barry	R1b-L21>Z253>L226>FGC5660	Ireland	R-L226	L226	13	24	14	11	11	14	12	12		13	13	16	17	8	9	11
f107338	Beedy	R1b-M269	zzUnknown	Beede	BLR	13	24	14	11	11	14	12	12	12	13	13	16	17	8	9	11
f15455	Benton	R1b-M269	zzUSA	R-L226	BLR	12	24	14	11	11	14	12	12	11	13	13	16	16	8	9	11
f217593	Bergesen	R1b-L21>Z253>L226>DC40	zzUnknown	R-L226	NGS	13	24	14	10	12	14	12	12	12	13	13	16	17	8	9	11
f248507	Berry	R1b-M269	Ireland	Berry	BLR	13	24	14	11	11	14	12	12	12	13	13	16	17	8	9	11
f5822	Blair	R1b-L21>Z253>L226>FGC5660>	zzUnknown	R-L226	L226	13	24	14	12	11	14	12	12	11	13	13	16	17	8	9	11
f303923	Bowers	R1b-M269	zzUnknown	Adoptees	BLR	13	24	14	10	11	14	12	12	13	13	13	17	18	8	9	11
f833307	Boyd	R1b-L21>Z253>L226>DC722	zzBlank	R-L226	NGS	13	24	14	11	11	14	12	12	12	13	13	16	17	8	9	11
f335951	Boyd	R1b-L21>Z253>L226	United Kingdom	R-L226	Z253	13	24	14	11	11	14	12	12	12	13	13	16	17	8	9	11
f192466	Bray	R1b-L21>Z253>L226>FGC5660	Ireland	R-L226	Indiv	13	24	14	11	11	14	12	13	11	13	13	16	17	8	9	11
f180168	Bray	R1b-L21>Z253>L226>FGC12295	Ireland	R-L226	L226	13	25	14	11	11	14	12	12	11	13	13	16	18	8	9	11
f21421	Bresnan	R1b-L21>Z253>L226>FGC5628	Ireland	R-L226	Z253	13	23	14	11	11	14	12	12	12	13	13	16	17	8	9	11
fN22028	Bresnan	R1b-M269	Ireland	Bresnan	BLR	13	23	14	11	11	14	12	12	12	13	13	16	17	8	9	11
f24629	Bresnan	R1b-M269	Ireland	Bresnan	BLR	13	23	14	11	11	14	12	12	12	13	13	16	17	8	9	11

Add Signatures to your EXCEL

- Determine the signature of the genetic cluster that you are tracking
- Track the both signature match and genetic distance from the signature in EXCEL
- These can be easily done with EXCEL macros borrowed from others and modified to your haplogroup or surname cluster
- Once you determine your macros, you just copy and paste the macros for each new tester found
- Keep track of false hits – you do not want to continue to “re-discover” false hits repeatedly

Add Signatures & GD to your EXCEL

- Signature Match (most are seven to ten markers)
=+COUNTIF(\$U17,"<=11")+COUNTIF(\$Z17,"<=8")+COUNTIF(\$AA17,"<=9")+COUNTIF(\$AG17,"<=29")+COUNTIF(\$AH17,"<=13")
- Genetic Distance (for one YSTR panel)
=+COUNTIF(\$BU17,"=12")+COUNTIF(\$BV17,"=11")+COUNTIF(\$BW17,"=13")+COUNTIF(\$BX17,"=11")+COUNTIF(\$BY17,"=11")+COUNTIF(\$BZ17,"=12")+COUNTIF(\$CA17,"=12")
- Total Genetic Distance (all 65 markers – do not include CDY)
=65-(+CH17+CI17+CJ17+CK17+CL17+CM17)
- Enter these just once – then copy & paste for new testers

Finding L226 Casey signatures

f190465	Casey	R1b-L21>Z253>L226	13	24	13	16	17	8	25	15	19	29	13	13	15	17	11	16	12	21	23	12	14	8	11	22	20
f58301	Casey	R1b-L21>Z253>L226	12	24	13	16	16	8	25	15	19	30	13	14	15	17	12	16	12	23	23	12	14	8	12	23	20
f107623	Casey	R1b-L21>Z253>L226	12	24	13	16	16	8	25	15	19	30	13	14	15	17	13	16	12	23	23	12	14	8	12	23	20
f45068	Casey	R1b-L21>Z253>L226	12	24	13	16	16	8	25	16	19	30	13	14	15	17	12	16	12	23	23	12	14	8	12	23	20
fB3520	Casey	R1b-L21>Z253>L226>DC50	13	24	13	18	17	8	25	14	19	27	13	13	15	17	11	16	12	23	23	12	15	8	12	22	20
f77349	Casey	R1b-L21>Z253>L226>FGC5639	12	24	13	16	16	8	25	15	19	30	13	14	15	17	13	16	12	23	23	12	14	8	12	23	20
f51924	Casey	R1b-L21>Z253>L226>FGC5639	12	24	13	16	16	8	25	15	19	30	13	14	15	17	13	16	12	23	23	12	14	8	12	23	20
f93773	Casey	R1b-L21>Z253>L226>FGC5639	12	24	13	16	16	8	25	15	19	30	13	14	15	17	12	16	12	23	23	12	14	8	12	23	20
f539983	Casey	R1b-L21>Z253>L226>FGC5639	12	24	13	16	16	8	25	15	19	30	13	14	15	17	13	16	12	23	23	12	14	8	12	23	20
f56031	Casey	R1b-L21>Z253>L226>FGC5660>DC331	13	24	13	16	17	8	25	15	19	28	13	13	15	17	11	16	12	21	23	12	14	8	12	22	20
f131349	Casey	R1b-L21>Z253>L226>FGC5660>DC331	13	24	13	16	17	8	26	15	19	28	13	13	14	17	11	16	12	19	23	12	14	8	12	22	20
fN146628	Casey	R1b-L21>Z253>L226>FGC5660>DC377	13	24	13	16	17	8	25	15	19	29	13	13	15	15	11	16	12	21	23	12	14	8	12	22	20
f373282	Casey	R1b-L21>Z253>L226>FGC5660>DC377	13	24	13	16	17	8	25	15	19	27	13	13	15	16	11	16	12	21	23	12	14	8	12	22	20
f412309	Casey	R1b-L21>Z253>L226>FGC5660>DC69	13	24	13	16	18	8	25	15	19	30	13	13	15	17	11	16	12	21	23	12	14	8	12	22	20
f277927	Casey	R1b-L21>Z253>L226>FGC5660>DC69	13	24	13	16	16	8	25	15	19	29	13	13	15	17	11	16	12	21	23	12	14	8	11	22	20
f53484	Casey	R1b-M269	13	24	13	16	17	8	25	15	19	28	13	13	14	17	11	16	12	19	23	12	14	8	12	22	20
f245778	Casey	R1b-M269	13	24	13	16	17	8	25	15	19	28	13	13	14	17	11	16	12	19	23	12	14	8	12	22	20
fN42374	Casey	R1b-M269	13	24	13	16	17	8	25	15	19	29	13	13	15	17	11	16	12	21	23	12	14	8	12	22	20
f34073	Casey	R1b-M269	13	24	13	16	17	8	25	15	19	29	13	13	15	17	11	16	12	21	23	12	14	8	13	22	20
f56130	Casey	R1b-M269	12	24	13	16	16	8	25	15	19	30	13	14	15	17	12	16	12	23	23	12	14	8	12	23	20
f56874	Casey	R1b-M269	12	24	13	16	16	8	25	15	19	30	13	14	15	17	13	16	12	23	23	12	14	8	12	23	20
f119586	Casey	R1b-M269	12	24	13	16	16	8	25	15	19	30	13	14	15	17	12	16	12	23	23	12	14	8	12	23	20
f54166	Casey	R1b-M269	12	24	13	16	16	8	25	15	19	30	13	14	15	17	13	16	12	23	23	12	14	8	12	23	20
f831266	Casey	R1b-M269	12	24	13	16	16	8	25	15	19	30	13	14	15	17	13	16	12	23	23	12	14	8	12	23	20
f172505	Casey	R1b-M269	12	24	13	16	16	8	24	15	19	30	13	14	15	17	12	16	12	23	23	12	14	8	12	23	20
f379785	Casey	R1b-M269	12	24	13	16	16	8	25	15	19	30	13	14	15	17	12	16	12	23	23	12	14	8	12	23	20
f99381	Casey	R1b-M269	12	24	13	16	16	8	25	15	19	30	13	14	15	17	11	16	12	23	23	12	14	8	12	23	20
f844322	Casey	R1b-M269	12	24	13	16	16	8	25	15	19	30	13	14	15	17	12	16	12	23	23	12	14	8	12	23	20
fB89528	Casey	R1b-M269	12	24	13	16	16	8	25	15	19	30	13	13	13	15	11	16	12	23	23	12	14	8	12	23	20

Advanced Tool – YSNP prediction

- Try NEVGEN predictor tool
<http://www.nevgen.org/>
- Try my L21 SNP predictor tool
http://www.rcasey.net/DNA/R-L21_SNP_Predictor_Intro.html
- Create your own empirical predictor model
- Add your statistical model macro using AcaStat to generate the constants (\$20)

John Marostica Prediction

Graphical view of probability and fitness

	Haplogroup	Probability	Fitness	Fitness 2
1	R1b L21>DF13> ZZ10>Z253> Z2534>L226	100	73.15	1.09
2	R1b L21>DF13> Z39589>Z251>> S11556	0	46.55	0.83
3	R1b L21>> DF21>>DF5> ZZ32> L1403	0	48.07	0.73
4	R1b U152>L2> Z49>Z142> Z51	0	45.72	0.82
5	R1b DF27>ZZ12> ZZ19>Z31644	0	45.84	0.84
6	R1b L21>DF13> Z39589>Z251> Z16943	0	44.32	0.79
7	R1b L21>> ZZ10>Z253> Z2186>L1066	0	44.05	0.78
8	R1b U152>L2> Z258>L20	0	43.63	0.79
9	R1b DF27>ZZ12> FGC20747	0	43.64	0.77
10	R1b U152>Z36	0	44.01	0.81
11	R1b U152>L2> FGC22501	0	43.53	0.78
12	R1b L21>DF13> DF21>S5488> BY518	0	43.83	0.71
13	R1b U152> PF6658> BY3642	0	43.39	0.76
14	R1b U106>Z381> Z156>DF96> S11515> L1	0	43.21	0.72

Randy Whited Prediction (mother's line)

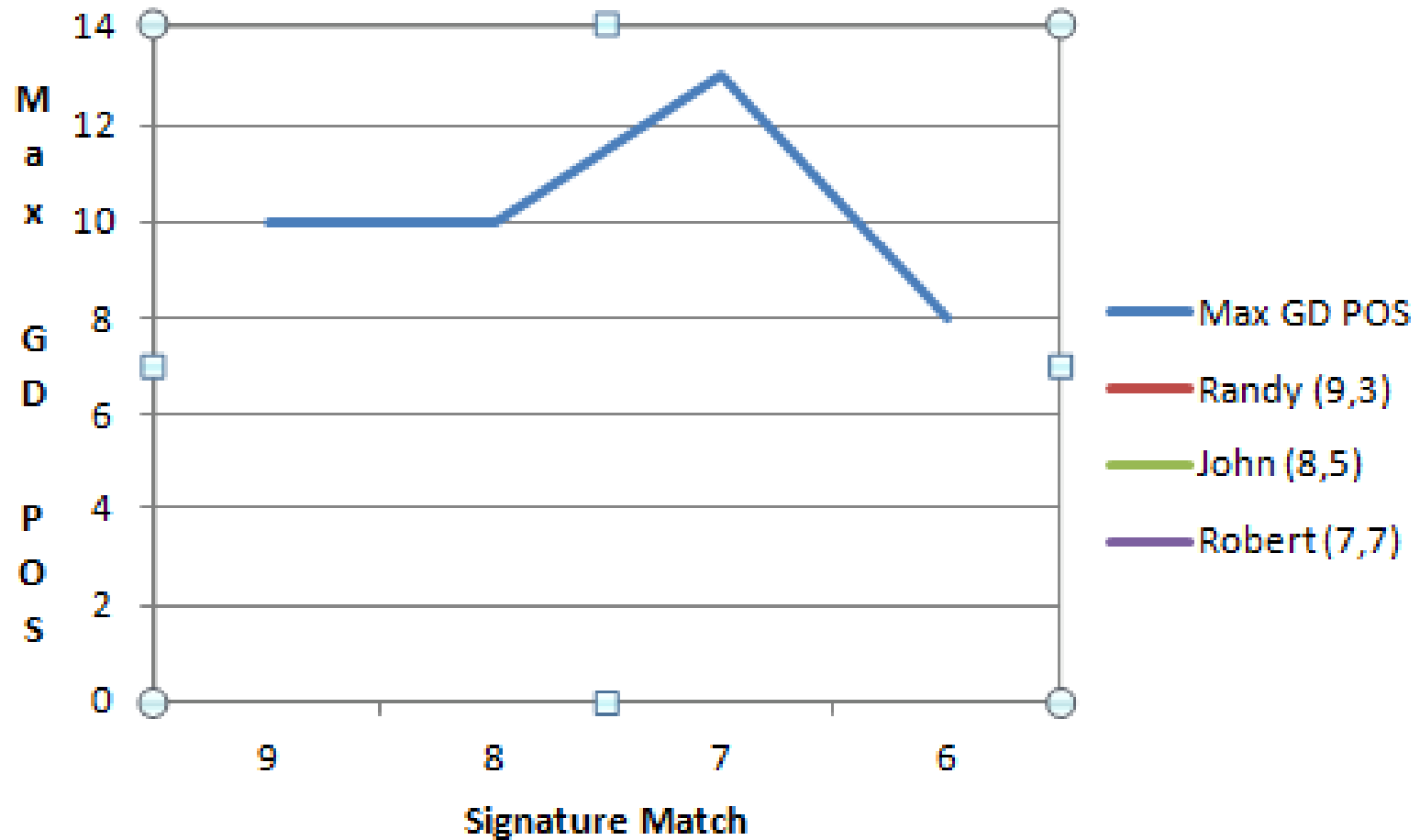
Possible R-L21 Y-SNP Testing Candidates

All columns can be sorted by clicking on column header

SNP	Probability of testing positive	Fingerprint Match	Fingerprint Breadth	Fingerprint Match Percentage	Genetic Distance
L226	95%	9	9	100%	4
L144 & L195 (K)	0%	6	12	50%	13
L159.2-C	0%	7	15	46.7%	15
DF41_M	0%	4	9	44.4%	13
L583	0%	4	11	36.4%	16
L744 & L745 & L746	0%	3	9	33.3%	16
L894 & L895	0%	4	12	33.3%	16
L908 & L909	0%	5	15	33.3%	18
L580	0%	5	16	31.3%	20
L69.4 (S)	0%	4	13	30.8%	18
L130	0%	4	13	30.8%	21
DF41_D	0%	3	10	30%	17
L564	0%	3	10	30%	16
L96	0%	3	11	27.3%	16
L554	0%	3	11	27.3%	19
L1335_P	0%	4	15	26.7%	19

Empirical Prediction (L226)

(Randy, John & Robert)



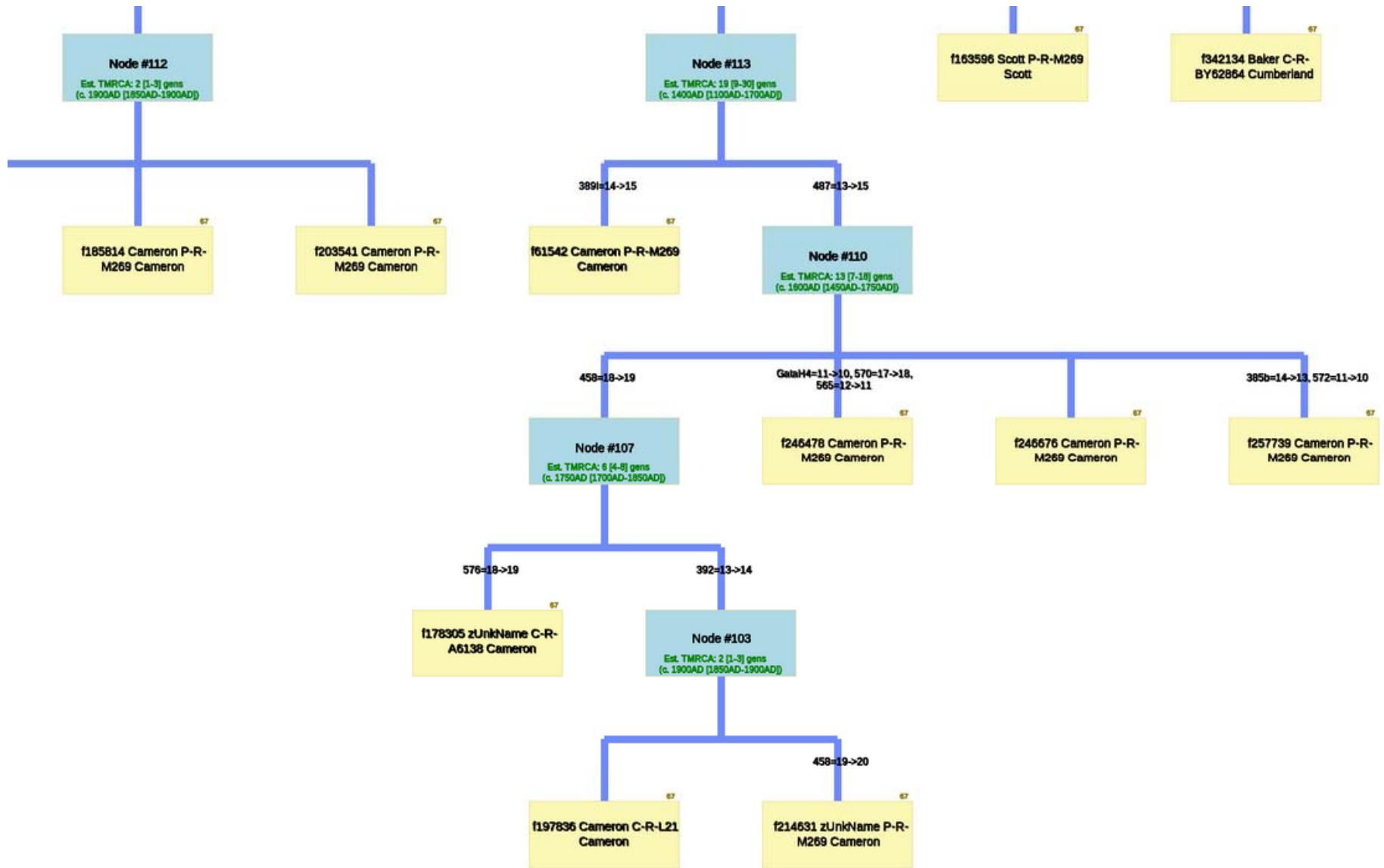
Statistical Model (L226)

- $$=(2.71828^{(-109.84+27.701*K17-5.838*L17)})/(1+(2.71828^{(-109.84+27.701*K17-5.838*L17)}))$$
- Formula never changes – only three constants change (you need positive & negative results)
- K is Signature match and L is genetic distance
- -109.84, +27.701 & -5.838 are the three coefficients that AcaSat (\$20) generates

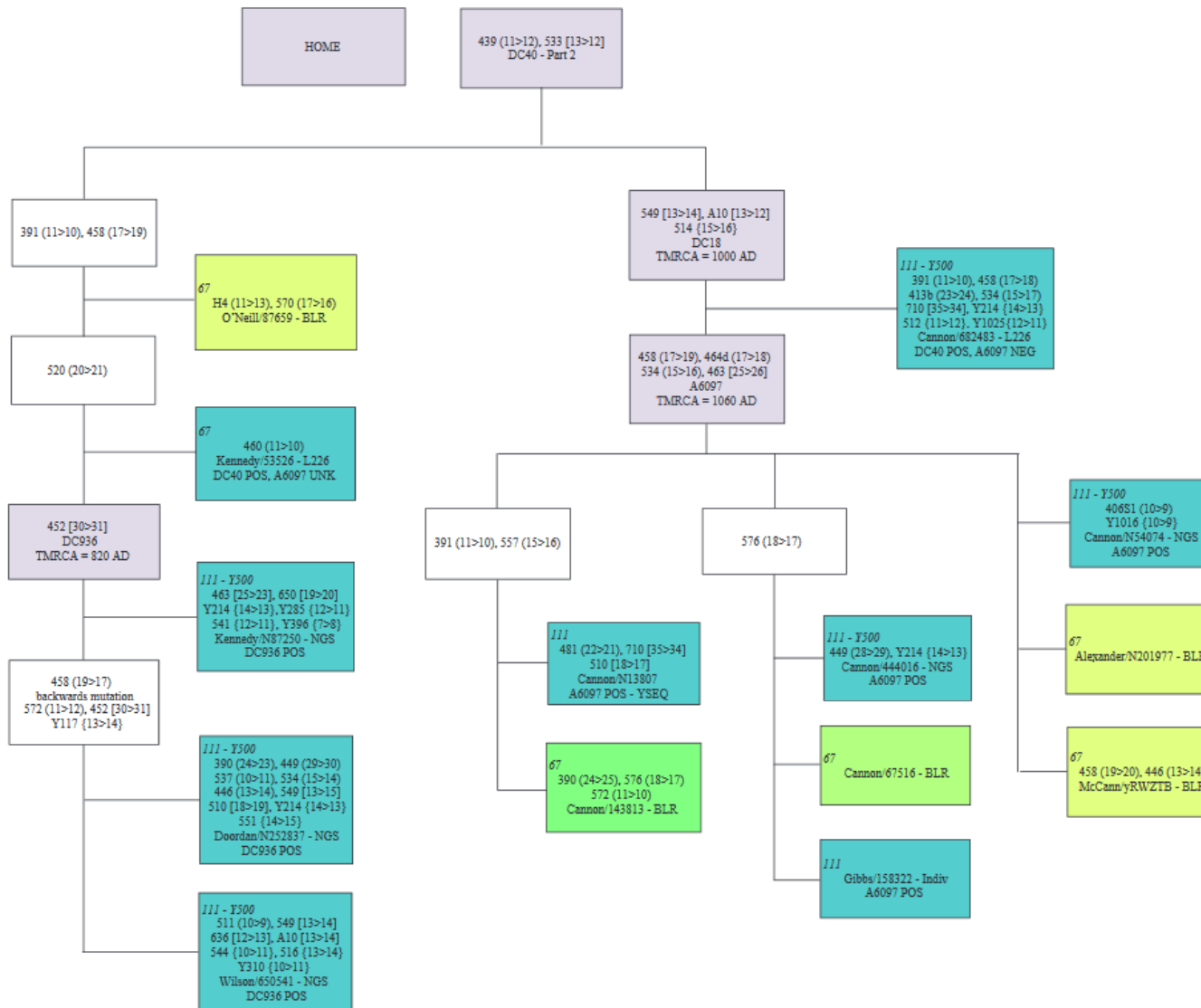
Charting with SAPP

- /STRDATA
fB199399 13 24 14 11 11 14 12 12 12 14 13
- /SNPDATA
f21421 (FGC5628+ FGC5659- DC40-)
- /INFO
f107338 Beedy P-DC728 Ireland
- /IGNORE
CDYa CDYb
- /SNPTREE
Z19670 BY31281
BY31281 A7298
A7298 A7300
A7300 A6138
- /OPTIONS
NODBTREE

Graphic SAPP Output – Z19670



Manual Chart with Adobe InDesign – L226



YouTube Videos

- Search “Genetic Genealogy Ireland”
- Search “Genetic Genealogy Robert Casey”
- Search “Genetic Genealogy First/Last Name”
Maurice Gleeson, John Cleary, James Irvine,
Dave Vance, Debbie Kennett, Paddy Waldron,
CeCe Moore, Rob Warthen, Blaine Bettinger,
Roberta Estes, Dennis O’Brien, Dennis Wright,
Margaret Jordan and Barbara Rae Venter