

Criteria for Y SNP prediction

- 1) Y SNP should be between 1500 to 2500 YBP. The oldest may go up a very small amount if other criteria are well met but can go more recent as well.
- 2) You need to have at least 10 to 20 branch equivalents in the highest two levels. But going over 30 branch equivalents across three levels becomes problematic. Having less than 10 branch equivalents is also less accurate as well.
- 3) You need at least 10 to 20 positive testers in your sample size and around 25 to 50 testers (both confirmed and predicted) to achieve very higher accuracy with the prediction model. You need at least 50 to 100 testers (confirmed and predicted) to be able to chart with reasonable accuracy.
- 4) If there are obvious major branches based on YSTR signature, this is a yellow flag that you may need to move down to a lower level where these YSTR signatures exist.

TMRCA Analysis

TMRCA YBP	BigTree	YFULL
BY114	2120	2800
L563	310	1550
BY23739	670	NA

BigTree implies BY114 is the best for a predictable haplogroup (green flag) but YFULL does not (yellow flag)

Branch Equivalent Analysis

		FTDNA Tree
Father	BY114	10
Son1	L563	15
Son2	BY23739	15
Gson1	FT75336	18
Gson2	Y31307	10
Averages across 3 levels		44 – yellow flag being too high

Analysis of Sample Size

	Branches	Positive	Predicted	Total
BY114	7	16	45	88
BY23739	2	6	16	22
L563	4	8	13	21

Even for Y SNP prediction, this appears that BY114 would be the best. Both sons are a little bit too low but are probably acceptable as well.

For charting, BY114 would be acceptable for charting accuracy but the two sons are significantly lower than the criteria.

Analysis of YSTR signature

Even though BY114 had a very acceptable signature of eight markers, there was some obvious convergence with other haplogroups (yellow flag). However, there were two major signatures revealed for L563 and BY23739 (red flag). After splitting up the analysis for two predictable haplogroups, both predictable haplogroups have 100 % accuracy. However, there are lot of untested boundary condition testers that could change the accuracy over time.

CONCLUSIONS

The decision to push the Y SNP prediction from L563 (in my L21 SNP predictor tool) up to its father was worth a try – but there were just too many alarm bells going off that should have been heeded. But when the analysis revealed two distinct signatures for the two sons of BY114, it became obvious that two prediction models would be required. But this did produce two good prediction models instead of just one – but these are very small scope in size.

The two confirmed testers that were BY114 were predicted to be BY23739. This is hopefully done via testing individual Y SNPs vs. Big Y. The model for BY23739 has very little boundary condition testing to date. So the constants for this prediction model will definitely change over time. It is unclear if the accuracy will remain as well.

Since the sample size was too low for charting, I did not produce any charts with SAPP.