

There are five steps for analyzing any predictable haplogroup:

- 1) Determine what level of haplotree is the best haplogroup to analyze.
- 2) Collect the data from many sources.
- 3) Determine the signature of the haplogroup.
- 4) Create a YSNP prediction model (statistical formula or empirical table lookup).
- 5) Create a chart using SAPP (or manually).

This process is highly iterative in nature, so you may need to repeat the cycle many times as you learn more about the haplogroup.

Step 1) – Determine the haplogroup to be analyzed.

I selected L554 as this YSNP branch has been around since the “Walk the Y” test. However, with all the new YSNP branches that have been added, the haplogroup may need to move up or down the haplotree since this branch was discovered over ten years ago.

FTDNA tree:

L554 – 10 branch equivalents
BY43439 (father) – 1 branch equivalent
BY14481 (son 1) – 14 branch equivalents
BY65903 (son 1) – 34 branch equivalents
ZZ6_1 (grandfather) – no branch equivalents
FGC3268 (great grandfather) – no branch equivalents

YFULL tree (not much help):

It does have:

FGC3268 – no branch equivalents – TMRCA = 3900 YBP
FGC3222 – son of the uncle of L554 – TMRCA = 1750 YPB
FGC61771 – uncle of L554 – TMRCA = 1900 YBP
BY23575 – great grandson of FGC3268 (great grandfather of L554)

BigTree

ZZ6_1 (grandfather of L554) – TMRCA = 3550 YBP
20151562 (father of L554) – could not find – 2490 YBP
L554 – 11 branch equivalents – 890 YBP
(includes BY14481, son of L554)

Surname distribution

Wilcox – 9

Whitney & Farris (probable Wilcox NPE lines)

Humphrey – 5

Murphy (probable Humphrey NPE line)

Pike – 4

Unknown (4) – probable Pikes or Pike NPE lines)

Other – 4 (possible future surname clusters)

TMRCA = 1200 YBP (does not meet 1500 to 2500 YBP requirement)

Both BigTree and YFULL really lack many relevant testers for L554. This means that TMRCA estimates will not be very accurate.

BigTree – L554 – TMRCA = 890 YBP

Surnames – L554 – 1200 YBP

These TMRCA's are really too young of a haplogroup for prediction and charting. After collecting the data for L554, the YSTR signature for L554 was found to be only six markers at Y67. This is the minimum size for a signature and would lower the accuracy of prediction. The YSNP prediction model is only 78 % accurate. Also, the sample size of confirmed testers is only 17 testers.

Criteria for good haplogroup to analyze:

- 1) TMRCA – too young for YSNP predictions – but can be done with lower accuracy.
- 2) Branch equivalents – 10 (average). However, the two known sons of L554 average 24 branch equivalents each (this is very high).
- 3) Signature size of six – minimum allowed – accuracy will suffer.
- 4) Sample size of confirmed testers – 17 – very low – needs to be 50.
- 5) Confirmed BigY testers (BigTree) – only three – very low.

Possible improvements:

- 1) Submit CSV files to Information Warehouse for analysis by BigTree.
- 2) Collect and update data in second pass.
- 3) Target testing of boundary condition testers.
- 4) Get testers for Pike surname cluster to add “Earliest Known Ancestor.”
- 5) Expand analysis of father of L554 (BY43439) – one found which does not match L554 very well.

Conclusions:

Even though YSNP prediction is only 78 % (well below the normal 98 % accuracy), charting reveals three large surname clusters: Wilcox, Humphrey and Pike. The signatures associated with these surname clusters are extremely strong: Wilcox (18 markers), Humphrey (13 markers) and Pike (13 markers).

There are only four false negatives which is not that many. But since the sample size is only 17 testers, this lowers prediction accuracy down 76 % accuracy. The boundary condition testers (those that barely pass prediction or those that barely fail prediction should be tested for targeting YSNP testing).

Even though prediction lacks accuracy, this should improve somewhat as the sample size grows. But SAPP charting works very well and quickly identifies three surname clusters as well as the most likely NPE lines of these surname clusters.

Recommendations:

1) Submit VCF files to Information Warehouse to get better TMRCA estimates and understand branch equivalents better.

2) Targeted Big Y testing of those that barely pass prediction:

Hall/B1327 and Morgan/153987.

3) Targeted Big Y testing of those that barely fail prediction:

Bonhan/191819, Brown/848178, Waters/121695 and Williams/201457.

4) Target Big Y testing of NPE lines of surname clusters:

Whitney/132150, Farris/H1249 and Murphy/B103734.

5) Build L554 cluster by upgrading Y37 testers to Y67 or Y111.

6) Collect new testers and upgrades to existing testers. With new testers, update YSNP prediction model and charting.

7) Proactively recruit testers to join L554 projects, post this analysis for project members to review and update the analysis as new information is added.