

## Big Y results for the Coffey family

By Timothy Peterman, Sep. 13, 2015

Many members of the Coffey family have participated in y-DNA testing over the last decade or so and have results for y-12, y-25, y-37, y-67, or y-111. These tests measure short tandem repeats (y-str) data for specific locations on the y-chromosome, where a very short segment, usually 4 nucleotides in length, repeats itself over and over. Occasionally copying errors happen and a repeat is deleted or a repeat is added. Since y-DNA doesn't recombine & most of it codes for nothing, these errors have no impact on the biological viability of the y-chromosome and build up over time. They have allowed genetic genealogists to predict with reasonable accuracy how closely two men are related. If two men match on say 65/67 of these markers, the TIP calculator shows that there is a 95% probability of a common ancestor within the last 11 generations, and a 99% probability of a match within the last 15 generations. The counts vary at random, sometimes diverging & sometimes converging. Sometimes people look, based on these counts, like they are a lot more closely related than they really are; or they look a lot more distant than they really are. Some men who match at 63/67 have been demonstrated to have a MRCA as far back as 1,600 years ago. Others who match at this distance have been shown to have MRCAs as recent as just a few hundred years ago.

There is another form of y-DNA testing that has been around for some time and that is SNP (or snip) testing. This stands for "Single Nucleotide Polymorphism". Anyone who has studied genetics knows that there are four types of nucleotides, with names shortened to the first letter: A, C, G, or T. If one of these changes between generations, it is called a mutation. A SNP is another name for a mutation. These define haplogroups. At Family Tree DNA, there are predicted haplogroups and tested haplogroups, which appear on the haplotree on your personal page as follows:

Your Predicted Haplogroup is **R-M512**

Your Confirmed Haplogroup is **R-L20**

A predicted haplogroup is based on a person's y-str results. A tested haplogroup means the relevant SNP has been tested and found in that person. A few years ago, haplogroups were given names like R1b, which means the SNP labelled M343 has been found. There is a complete haplotree of male humanity that has been found based on SNP testing. To see the tree that International Society of Genetic Genealogists concur is most accurate at present, go to [www.isogg.org](http://www.isogg.org). This tree goes all the way back to the common ancestor of all men & identifies different branches based upon SNP testing. The tree has ascertained the order in which SNPs have appeared. The Coffey y-DNA and 99.999% of all men are in a haplogroup called A0-T, defined by the SNP L1085+. Next in the line of descent is A1 (P305). Next down is A1b (P108). Next down is BT (M91), etc. This early ancestry occurred in the Middle Paleolithic. Eventually, as we follow the descent down, we get to the R haplogroup (M207), estimated to be 31,000 years old.

Over time, the number of SNPs that have been identified and published has grown tremendously. R1b1a2 (M269) is the ancestor of most western European R1b. The Coffey family is nested under R1b1a2a1a2 (P312). A little over a year ago, Family Tree DNA dropped the nomenclature R1b1a2.... P312+ is simply called R P312. One has to consult a haplotree to see where R P312 is situated. It is just beneath L11, which is just beneath L51, which is just beneath L23, which is just beneath M269. Each of

the names for a SNP are indexed to an exact location on the y-chromosome, where the SNP occurred. Because these are actual mutations, the odds of these reversing themselves are really low, so if someone is P312+, it means that he descended from a man who had a C mutate to an A at y-DNA location 22,157,311.

The terminal SNP is the most refined subclade for a patriline that has been identified. For the Coffey family, that has been L176 or Z198, which is nested under Z195, which is nested under DF27, which is nested under P312. The only y-str matches worth considering are those who are also L176+ or Z198+. Terminal SNPs are like guideposts. If a person is negative for L176 or Z198, the MRCA of that person with the Coffeys lived thousands of years ago and isn't related in a meaningful way, regardless of what his surname is or how many short tandem repeat counts he shares. The terminal SNP enables us to separate the chaff from the wheat. A negative result for someone who should be positive might also suggest a testing error. The Big Y report gives good data for ascertaining whether a testing error has occurred.

As time has passed, y-DNA testing has focused increasingly more on the discovery of ever more refined terminal SNPs, and a lot of progress was made before the advent of Big Y testing. Wouldn't it be wonderful if we could identify a terminal SNP that occurred as recently as say 500 years ago? The only men with the Coffey surname to consider for the Edward Coffey family would be those that share such a terminal SNP. Let's say we identify a SNP that occurred 1,000 years ago. This might include more distantly related Coffeys, as well as the Keough family. A terminal SNP from say 400 to 800 years ago would probably reveal roughly when the MRCA of the Coffey and Keough families lived. A SNP from 1,500 years ago would no doubt include even more Irish families and would involve many surnames. Contrast this to DF27, which is estimated to be 4,600 years old. Z195 and Z198 each does a wonderful job of refining the DF27 tree to those matching families we are interested in, but even these are only a few generations younger than DF27. So how do we get down to terminal SNPs as recent as 1,500 years, or 1,000 years, or 500 years ago?

Enter Big Y testing. Anyone who has done any y-str testing will see the Big Y advertised on his personal page. The y-chromosome is about 55 million base pairs in length. Each SNP occurs at one specific location in this series of 55 million nucleotides. The Big Y tests for about 36,000 known SNPs on the y-chromosome. It looks at the published location and determines whether the mutation has occurred: Yes or No or ?. Sometimes, no read can be made & the Big Y returns a ? value. In addition, the Big Y looks at about 10 million base pairs on the y-chromosome and searches for "novel variants"; that is SNPs that have never been discovered before and are thus, unpublished. The Big Y usually finds anywhere from 80 to 120 novel variants in each participant. The Big Y uses what is called Next Generation Sequencing, which uses a new technology to analyze the y-chromosome. It is faster, more comprehensive, and more accurate, and requires specialized testing equipment, which is probably why Big Y tests cost more than most of the others.

An example of a SNP report from the Big Y follows on the next page. Derived means that the participant is positive for the SNP. Unknown means there wasn't a good read for the SNP. Medium indicates that the SNP was read, but was low quality. "On Tree" indicates whether the known SNP appears on the Family Tree DNA haplotree. The Coffey data includes a report like this on 36,384 SNPs. The vast majority can be found in other haplogroups. These are always shown as No in the Derived column. The count of 36,384 ANPs also includes 114 novel variants.

Type	Position	SNPName	Derived	OnTree	Reference	Genotype	Confidence
3	Known SNP	PF5469	Yes (+)	No	C	C	High
4	Known SNP	PF5468	?	Yes	G	?	Unknown
5	Known SNP	PF5467	?	No	G	?	Unknown
5	Known SNP	PF5466	?	Yes	T	?	Unknown
7	Known SNP	PF5465	?	Yes	C	?	Unknown
3	Known SNP	PF5464	?	No	A	?	Unknown
9	Known SNP	L344	?	No	T	?	Unknown
9	Known SNP	PF5462	?	No	T	?	Unknown
1	Known SNP	PF5461	?	No	T	?	Unknown
2	Known SNP	PF5460	?	No	G	?	Unknown
3	Known SNP	PF546	?	No	T	?	Unknown
4	Known SNP	PF5459	Yes (+)	No	C	C	High
5	Known SNP	PF5456	No (-)	Yes	G	G	High
5	Known SNP	PF5455	No (-)	No	G	G	High
7	Known SNP	PF5454	?	No	A	?	Unknown
3	Known SNP	CTS6061	No (-)	Yes	C	C	High
9	Known SNP	PF5452	No (-)	No	G	G	High
9	Known SNP	PF5451	No (-)	No	T	T	High
1	Known SNP	PF545	?	No	T	?	Unknown
2	Known SNP	L398	No (-)	Yes	T	T	High
3	Known SNP	PF5448	?	No	T	?	Unknown
4	Known SNP	PF5447	No (-)	No	C	C	Medium
5	Known SNP	L397	No (-)	Yes	G	G	High

I have done Big Y tests for several lines that I descend from: Petermann, Robinson, Ellis, etc. I noticed earlier this year that, although the Coffey project has over 100 participants, none of them had participated in the Big Y project. I asked one of my mother's second cousins, a descendant of Benjamin Coe Coffey (1845-1927), if I could order a Big Y test for the DNA he had contributed as part of the Family Finder project. He readily agreed.

His results came in a few weeks later (late August, 2015). He hadn't done y-DNA testing before. The Big Y placed his terminal SNP at Z198, which is consistent with the rest of the Coffey family. The Big Y found 114 novel variants in this Coffey participant! Of these, 84 were shared with other Big Y testers, most of whom are DF27+. Many of these occurred earlier and can be found among all R1b men, or maybe all BT men. The exciting news is that 30 of these were not shared with any other participant – they are truly unique to his line of descent.

The novel variants reported by the Big Y are estimated to occur about every 135 to 150 years. With 30 of these, it means the MRCA for the Coffey line and all other current participants is between 4,050 and 4,500 years ago, consistent with the age estimate for DF27 & Z198. This string of 30 unshared

novel variants includes the terminal SNPs in our family from 1,500 years ago; as well as the SNP for 1,000 years ago & the one for 500 years ago. However, we can't tell which is which.

I uploaded the Coffey participant's results to [www.ytree.net](http://www.ytree.net) and learned within a day that his terminal SNP is actually Z292. This separates our branch from the rest of the Z198+ crowd. Nice! With the passage of time, as more Z198+ men participate in Big Y testing, we will learn which of the 30 unshared novel variants branched off first. Without doing any more Coffey testing, we will eventually see an ever more refined terminal SNP.

Edward Coffey was born perhaps 340 to 350 years ago. Each line of descent from Edward has been incurring novel variants since his marriage to Ann Powell. There were probably as many as one or two novel variants separating the current participant from his g-g-g grandfather, Eli Coffey (1775-1833), and perhaps another novel variant that happened between Edward Coffey and Eli Coffey. By counting the number of unshared novel variants between different lines of descent from Edward Coffey, we could confirm how closely related these lines of descent really are. This could probably be done just as easily with high resolution y-str testing. If any anomalies were to appear between DNA testing and paper trail documentation, we would then have to work on resolving these anomalies. Big Y results would confirm the findings of the y-str analysis.

The most important objective for me was to get one descendant of Edward (and preferably one rather closely related to me) to do the Big Y test. I would recommend that others do this if and when the price gets reduced considerably below what it is now. Since Family Tree DNA can usually run a Big Y test off of DNA left over from a Family Finder test, I encourage all of us to continue to get DNA samples, especially from the older family members and run either a y-12 or Family Finder on them at the current low prices. If there is a deeply discounted sale on the Big Y, we might consider upgrading some more of these results.

I strongly recommend that Big Y testing be done in conjunction with a y-DNA haplogroup project and that any P312+ results be added to [www.ytree.net](http://www.ytree.net). Just watch this over the next year or so. The Coffey family's place in the tree will probably become even more refined.