

Y-DNA Evidence for an Ashkenazi Lineage's Iberian Origin

by Rachel Unkefer, J.B. Royal, and Wim Penninx

Most genealogists researching their Ashkenazi families encounter brick walls in the paper trail within the past 200 years. A new research strategy using Y-DNA Next Generation Sequencing (NGS) allows us to go back further in time and, in some cases, discover a history different from the expected one.¹ Using results from FamilyTreeDNA's "Big Y" test, we are uncovering just this kind of story, about a lineage of Ashkenazi Jews that appears to be descended from a man who lived in the Iberian Peninsula (Spain and Portugal) a thousand or more years ago. This article presents our preliminary findings.

Introduction

Jewish genealogy typically focuses on the past two centuries, since this is the primary period for which documentation of individuals and their families is available. Jewish

*A new research strategy using Y-DNA
Next Generation Sequencing (NGS)
allows us to go back further in time.*

families who lived in Eastern or Central Europe during that period generally consider themselves to be Ashkenazi (based on the Hebrew word for "Germany"). Some have family stories of having been Sephardic (based on the Hebrew word for "Spain") before moving eastward, either prior to or after the Spanish Edict of Expulsion in 1492, but documentation for such migration is sparse.

For those who cannot find a paper trail for their Ashkenazi ancestors back to the once-thriving Jewish communities of Spain and Portugal, another type of evidence may bridge that gap. New Y-DNA² tools and techniques provide just such evidence in at least one Ashkenazi lineage.

Several years ago, at the beginning of Y-chromosome testing for genealogy, little information was available about the haplogroups into which the testers' results were placed.³ The National Geographic's Genographic Project had begun to map the broad strokes of human migration tens of thousands of years ago, but technology for discovering the finer branching of haplogroups in a more recent time frame was not yet invented.

Some Y-chromosome haplogroups common among Jewish men were identified as having Middle Eastern or North African origins (e.g. J1, J2, E1b1b1),⁴ but a small percentage of Jewish men found themselves in the R1b haplogroup, shared with the majority of men in Western Europe. Many R1b Jewish males wondered what this meant about their Jewish genetic heritage. If their Y-DNA appeared to be more similar to that of British men than Middle Eastern men, how did their families wind up as Ashke-

nazi Jews in, for example, Belarus in 1820?

A decade and a half later, knowledge about each haplogroup has exploded, due to Next Generation Sequencing (NGS). These NGS tests examine about 12 million positions on the Y chromosome, record the "letters" (the A, C, G, and T nucleotide designations) at each position, and compare them to the "Reference Genome," a model of a representative person's nucleotide designations at each location. This is not the genome of an actual person, but a composite model of a number of different people's genomes. Each position in the tester's NGS results that has a letter different from the Reference Genome represents a mutation, or SNP.⁵

Over the past few years, more than 300,000 SNPs have been discovered and catalogued, and now, by grouping thousands of individuals' test results according to which SNPs are shared by which men, we can create robust phylogenetic trees (charts like the one in Figure 1, which show the evolutionary branching of mutations). New discoveries are made every day, and the tree increases in accuracy and complexity with each discovery.

This article presents the non-technical preliminary results of a Y-DNA study that began with a cluster of men of Ashkenazi descent and one of Sephardic descent in an R1b subclade⁶ known as R-DF27, today most frequently found in the Iberian Peninsula. The vast majority of the Ashkenazi men had no knowledge or family stories of any connection with Spain or Portugal. Later, we discovered that there were others who shared a SNP with these Ashkenazi men, including a prominent Sephardic lineage, and possible descendants of converso refugees in the 16th-century New World.⁷

Our haplogroup subclade, named for its defining SNP, R-FGC20747,⁸ is estimated to have originated around 4,500 years ago, some 400-600 years after the arrival of the first R1b ancestor in Western Europe from the Russian steppe about 3,000 BCE,⁹ but it has multiple Jewish and possibly converso descendant branches. Some central questions focused on the Jewish branches within our study. Who were the earliest Jewish ancestors? When and where did they live?

Methodology

The study began with four men of Ashkenazi paternal ancestry who were identified by FamilyTreeDNA (FTDNA) as STR¹⁰ genetic matches with each other and had independently been tested using FTDNA's Big Y. All four had joined the large haplogroup subclade project (R-DF27) and submitted their raw data files to the Big Tree created by Alex Williamson.¹¹

The four men appeared in a sparsely populated part of the Big Tree under a branch named for the SNP designated

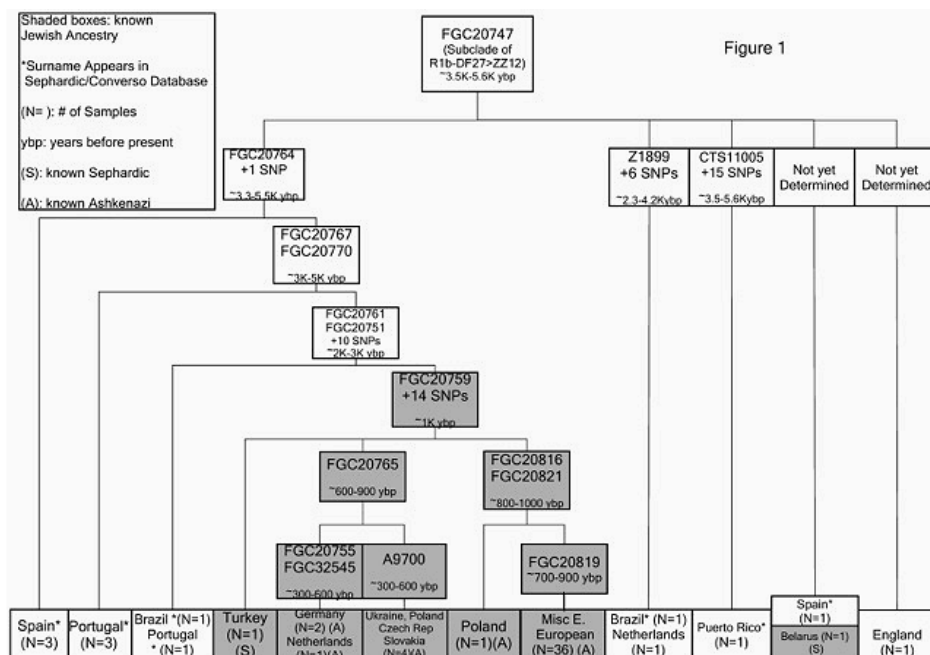


Figure 1. In the center cluster of shaded boxes, all but two of the 45 men have known paternal Jewish ancestry.

FGC20747, which they had in common with a few other men.¹² On another branch of this SNP, parallel to the Jewish cluster, was a Portuguese man with a surname that appeared to be Sephardic, based on a search of Jeffrey Malka's Consolidated Index of Sephardic Surnames.¹³ Although the Portuguese man did not have documentation of Jewish ancestry, he thought it was plausible. Nearly all the other surnames on the Big Tree under FGC20747 apparently were Iberian and potentially Sephardic as well. The authors of this article requested the creation of a subclade project at FTDNA and began to invite participants.

All members of our study came from the pool of individuals who had already independently taken Y-chromosome STR tests at FTDNA. We simply gathered those whose results indicated a possible close genetic connection with our group.

The starting point was the STR genetic match lists of the initial four Ashkenazi men and the matches of their matches. When combined, the match lists included approximately 100 individuals of known or presumed Ashkenazi descent, based on surname, geography, and closeness of genetic match. Those 100 individuals were invited to join the study; 45 did so. Among the group of matches to the Ashkenazi members of the group were several members of a family from Turkey and Bulgaria with known Sephardic ancestry; two members of this family joined the study.

In addition to those invited based on their STR matches with the Ashkenazi individuals in the study, all of the men shown on Alex Williamson's Big Tree under FGC20747 were invited to join, along with some of their STR matches. We, the authors, who are also the project administrators of R-FGC20747, worked with the administrators of the larger haplogroup projects of which FGC20747 is a subset (R-

P312 and R-DF27) to invite those project members whose new test results from Big Y and SNP Packs identified them as being within the FGC20747 part of the tree. Finally, we looked at various other FTDNA project public websites, e.g., "Spanish and Portuguese projects" to identify additional possible candidates.

The technical details of the testing we performed after assembling our group are outside the scope of this article, but in general, we created subgroups of men with similar STR values and administered Big Y tests to at least two individuals in each subgroup to find the shared SNPs that defined that subgroup. Once we established the terminal SNP¹⁴ or SNPs for each subgroup using the Big Y test, we tested the other individuals in that subgroup for only those few SNPs, rather than performing a Big Y test for each person. This is a methodology used by many other Y-DNA projects at FTDNA.

Results

Jewish Ancestry. In Figure 1, the cluster of shaded boxes at the center bottom contains 45 individuals, all but two of whom have known paternal Jewish ancestry. We may infer from the close genetic matches and their ancestral countries that the two exceptions also likely have paternal Jewish ancestry. The 45 men are shown descending from a block of 15 shared SNPs. At present, all the individuals who have done the Big Y test are positive for all of them. Rather than listing all 15, we have labeled this branch for one of the SNPs, FGC20759. Below the FGC20759 box are three main branches:

- Unlabeled. One individual whose paternal MDKA (Most Distant Known Ancestor) was from Turkey and known to be Sephardic

- FGC20765. Seven men whose paternal MDKA were from Czech Republic, Germany, the Netherlands, Poland, Slovakia, and Ukraine

- FGC20816/FGC20821. Thirty-seven individuals whose MDKA were from Belarus, Czech Republic, Germany, Hungary, Latvia, Lithuania, Poland, Romania and Ukraine

Below the third group (FGC20816/FGC20821), the box labeled “Misc E. European” can be broken down into several more recent SNP groups beneath FGC20819, but in order to simplify the overview chart, the details are not shown here. Several individuals included in this group who have not done the Big Y test likely belong on different branches below FGC20819, but they have not yet been tested for the SNPs that define those sub-branches. When all are tested, we will know their positions on the tree more specifically, and thus, which people are most closely related to each other.

Our data analysis is not yet complete, but at this time we estimate that the Most Recent Common Ancestor (MRCA) of the individuals in this shaded cluster likely was born between 800 and 1200 years ago.¹⁵ Most likely, the MRCA lived in Iberia, with one or more of his descendants migrating east to Ashkenazi countries at some point soon after, and at least one descendant remaining in Iberia and migrating to what is now Turkey when the Jews were expelled from Spain and then Portugal.

On the far right edge of Figure 1 is one other individual with known paternal Jewish ancestry. Although his MDKA lived in Belarus, his surname indicates descent from a well-known Sephardic family. His MRCA with the other known Jewish cluster in the center was likely born between 3,500 and 5,600 years ago.¹⁶ Locating more genetic matches will help refine the estimate.

Iberian Ancestry

As far as we know now, FGC20747 has at least five branches (see Figure 1).

- FGC20764 from which the Jewish cluster is descended, and also a branch with three men from Portugal with possible Sephardic surnames

- Z1899. One from the Netherlands with no known paternal Jewish ancestry and one from Brazil with a possible Sephardic surname

- CTS11005. One from Puerto Rico with a possible Sephardic surname

- Not yet determined. One from Spain with a possible Sephardic surname and the aforementioned one from Belarus with known paternal Sephardic ancestry

- Not yet determined. One from England with no known paternal Jewish ancestry

We expect that future analysis and more data samples will identify additional SNPs that will create additional branches.

Because the R-DF27 haplogroup is predominantly found in Spain and Portugal, the individuals outside the known Jewish cluster are predominantly of Spanish and Portuguese

descent, as expected. There is also a high incidence of possible Sephardic or converso surnames. It is important to note, however, that a surname used in Spain or Portugal by Sephardic Jews or conversos does not necessarily prove Jewish ancestry. Both Jews and non-Jews used many of the same surnames.

Other Ancestry

In addition to known Jewish and/or Iberian branches of the FGC20747 tree, we also have at least two descendants from other parts of Europe. One man traces his oldest known ancestor to England in the 18th century and the other to the Netherlands at about the same time. We do not know yet if we will ultimately discover a connection to Spain or Portugal.

The major question raised by this study is: what historical scenario or scenarios explain a long-ago potential Iberian common ancestor for present-day descendants who overwhelmingly appear to be of Ashkenazi, Sephardic, or possibly converso descent on multiple branches?

Tracing up from the bottom of the chart in Figure 1, as each branch joins the one above it, there is an MRCA of everyone below—the man who first had that particular mutation, or SNP. A branch is created each time one group has a mutation or mutations that the others do not have. All descendants on the male line will carry that mutation to the present day. It is possible, in theory, for a mutation to be lost or changed again in a future generation, but it is extremely unlikely statistically.

Although it seems clear that the shaded cluster in the middle of Figure 1 has a Jewish ancestor about 1,000 years ago, we do not yet know if that ancestor was born in Spain, Portugal, or elsewhere. Looking at the common ancestors prior to 1,000 years ago (the ancestor of the FGC20767 branch, the FGC20764 branch, and back to FGC20747), we see 11 descendants with known recent Portuguese or Spanish ancestry. This is a strong indication that one or more of the ancestors of the Jewish cluster lived in Iberia at some time.

Future Research

One important question to answer is how many of the participants of non-Jewish Iberian ancestry can be traced to a paternal converso ancestor? At present, none of the non-Jewish men in the study has documentary evidence that their paternal line was Jewish. Nearly all, however, either suspect that they have Jewish ancestry or have surnames that appear in databases of known Sephardic/converso surnames.

A further area for research is the acquisition of more Y-DNA data. This can come from testing more individuals from Spain, Portugal and their former colonies, as well as more Sephardic individuals. It would also be useful to identify more R-M269 men already in the FTDNA database who belong in our subclade, but have not yet tested individual SNPs (via Big Y or SNP Pack) that would place

them in the correct area of the R1b tree.

Final Thoughts

The lesson of this study for Jewish genetic genealogists is that a bird's eye view, above and beyond the limitations of the Y-DNA STR match lists provided by FTDNA, may reveal an MRCA farther back with non-Jewish males whose ethnicity can shed further light on Jewish history and migration.

Although R-DF27 is most common in Spain and Portugal, we did not have a concrete genetic connection between the Ashkenazi lineage and individuals of Spanish and Portuguese ancestry until we looked farther back in time, using NGS testing. We only knew about the Ashkenazi cluster of STR matches, related to each other fairly recently (within the last 1,000 years).

We hope that our study can be a model of a new approach to placing Jewish DNA clusters in a broader historical and geographical context. Jewish migration throughout the world is something we all take for granted, but it is not possible for most genealogists to find documentation of their families' routes to Germany and Eastern Europe during the last 1,000 years. By studying Jewish clusters as haplogroup subclades, genealogists might find clues to their ancestors' geographical origins even before they came to the Ashkenazi countries.

Notes

1. Next Generation Sequencing (NGS) is an advanced test for the Y-chromosome. Family Tree DNA's Big Y and Full Genomes Corporation's Y Elite 2.1 are examples of this type of direct-to-consumer test (as opposed to medical lab tests).

2. Y-DNA is passed down exclusively through the male line, from father to sons, relatively unaltered over many generations.

3. Haplogroups are branches of the human phylogenetic tree, designated by capital letters branching out from the "A" group (hypothetical "Adam"). A new haplogroup is created when a single location mutation occurs in an individual that is then passed on to his descendants. Mutations that occurred many millennia ago and are now shared by vast numbers of men define the main branches of the human haplotree (alphabetically A through T).

4. The original naming schema for haplogroups and their sub-branches (subclades) was to use a capital letter for the main branch and then a number for the next branching point, then a lower-case letter, and then alternating number and letter as each new branch was discovered. For example, main branch "R" was divided into R1 and R2. R1 was divided into R1a and R1b. R1b was further divided into R1b1 and R1b2. Over time, so many new branches were created as new SNPs were discovered, that the names became long and unwieldy. In the new naming schema, haplogroups are named using the main branch capital letter followed by a dash and then the SNP that defines the haplogroup. For example, what used to be known as R1b1a1a2a1a2a is now called R-DF27, because the common mutation all men in that branch have is called DF27. In the new naming schema, J1 is J-M267, J2 is J-M172, and E1b1b1 is E-L117. The SNP names are shorter, but one disadvantage of the new system is that there is no "breadcrumb trail" of branching embedded in the name. It's now necessary to refer to several online sources to determine which part of the haplotree a particular haplogroup belongs to.

5. SNP, or Single Nucleotide Polymorphism, is a mutation of a single nucleotide (A, C, G, or T) from the reference value at a

given location.

6. A subclade is a subdivision of a haplogroup. Within a subclade, the individuals are more closely related to each other, with a more recent common ancestor, than the larger haplogroup.

7. A Converso is a person who converted to Catholicism from Judaism prior to or during the Inquisition in Spain or Portugal or any of their former colonies.

8. This SNP is located at position 13829071, and is a mutation from the reference C to A.

9. Based on the estimates of the YFull Corporation's Experimental Tree version 4.02 <http://yfull.com/tree/R-FGC20747/>

10. An STR or short tandem repeat is a repeating nucleotide pattern that can be counted, e.g., "AGAGAG" would be counted as three repeats of the pattern "AG." In specific locations, men who are descended from a common male ancestor will usually have the same number of repeats. See the previous AVOTAYNU article, "Interpreting Y-DNA Markers: A Primer" by Rachel Unkefer, Spring 2014.

11. Alex Williamson is a volunteer "citizen scientist" who has created a public haplotree for anyone who has undergone "Next Generation Sequencing" (NGS) and has been identified as belonging to R-P312 or any subclade below it. This effort to aggregate individual data for thousands of participants has been instrumental in illuminating the structure of this section of R1b. <http://ytree.net/>.

12. SNP names are assigned by the individual, company or organization that first identifies them. This SNP happened to be named by the Full Genomes Corporation, hence the prefix "FGC" and the number. These numbers are simply labels and do not have any other meaning in themselves. The naming convention can be found on the International Society of Genetic Genealogy (ISOGG) website, <http://isogg.org/tree/index.html>.

13. www.sephardicgen.com/databases/indexSrchFrm.html

14. An individual's terminal SNP is the one on the lowest level of the tree that he shares with the smallest number of other men. In other words, it is the SNP most specific to him and his closest patrilineal relatives.

15. Based on the estimates of the YFull Corporation's Experimental Tree version 4.02 <http://yfull.com/tree/R-FGC20747/>.

16. *ibid.*

Rachel Unkefer has more than 30 years experience in German and American Jewish genealogy. In addition to being a co-administrator of the R-FGC20747 Project, she is the administrator of the Bacharach DNA Project and co-administrator of the Jews of Frankfurt DNA Project and the Suwalki-Lomza Jewish DNA Project. She lives in Charlottesville, VA. Email for project: fgc20747@gmail.com.

J.B. Royal, a passionate software architect, was an instrumental contributor to the design and development of Lotus Development software programs and the Leading Edge and Wang Word Processors. A valued problem solver, he now employs the same intensity and zeal in his genealogical research using traditional methods and DNA analysis. He lives in the Boston, Massachusetts area.

Wim Penninx received his PhD in Astronomy in 1990. Penninx created and maintains several websites on genetic genealogy, among which is jewishdna.net, on the origins of various Jewish Y-DNA lines, and a Dutch language website genografie.nl. He lives in Delft, the Netherlands.