

English or Irish?: A Smith Family DNA Case Study

Ugo A. Perego, Ph.D looks at the genetic origins of a large Smith family in Utah

BEGINNING IN 2000, DNA samples and pedigree data were collected as part of the worldwide Molecular Genealogy Research Project. The main objective of this study was to build the world's foremost database of correlated genetic and genealogical data to assist people in discovering lost and hidden relationships through the power of DNA. The current size of the database includes more than 110,000 DNA samples linked to nearly 8.5 million names and it is accessible at www.GeneTree.com and at www.SMGF.org.

One of the first tasks that needed to be accomplished was to select a number of families willing to contribute genealogical records and biological samples to assist in calibrating the genetic methods employed in building the correlated database. One of these pilot groups was a large Smith family in Utah, whose ancestors were among the first pioneers arriving in the Great Salt Lake Valley starting in 1846. These progenitors are survived today by a legacy of more than 15,000 descendants. As often happens in genealogical research, this group was not immune to

genealogical brick walls and family history ambiguities.

Genealogical Evidence

An article published in 1991 by Elaine Nichols, an expert on the Smith family history, summarized what was actually known up to that time about the origin of the Smith family and their New World beginnings, concluding that the information available to make a valid connection with England was indeed quite sketchy and unreliable. According to Nichols,

the earliest documented Smith ancestor in the United States was a young boy named Robert, possibly born in 1626 somewhere in the proximity of Kirton, Lincolnshire, in England. Robert Smith shows up in records for the first time at age 12, as a witness in a trial that took place in Massachusetts in 1638. At that time, he was recorded as an indentured servant to John Whittingham, a wealthy landlord with properties in the Kirton area. No additional information is available about

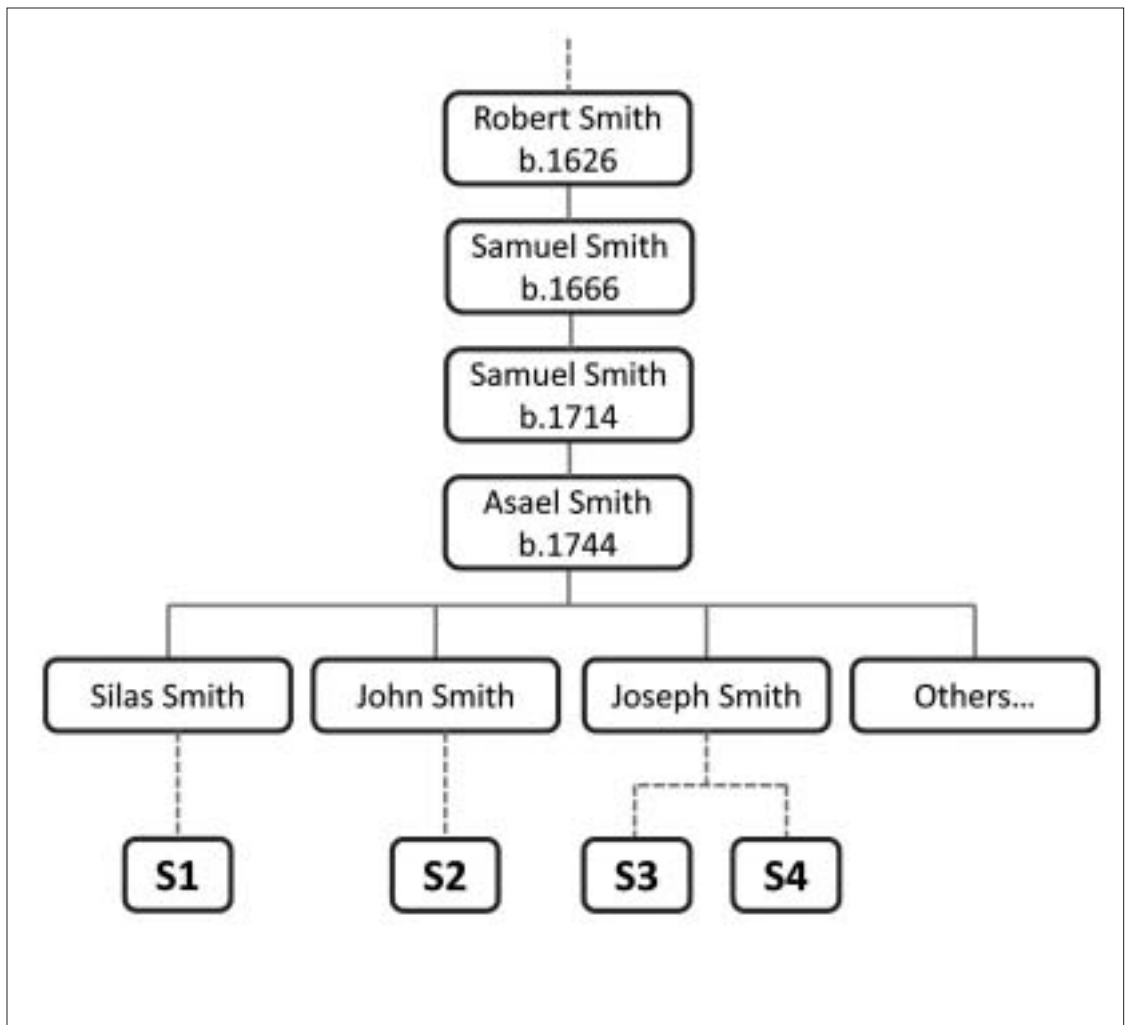


Figure 1: A schematic paternal pedigree for the Smith family. S1-S4 provided DNA samples for this study. (Names removed for privacy reasons)

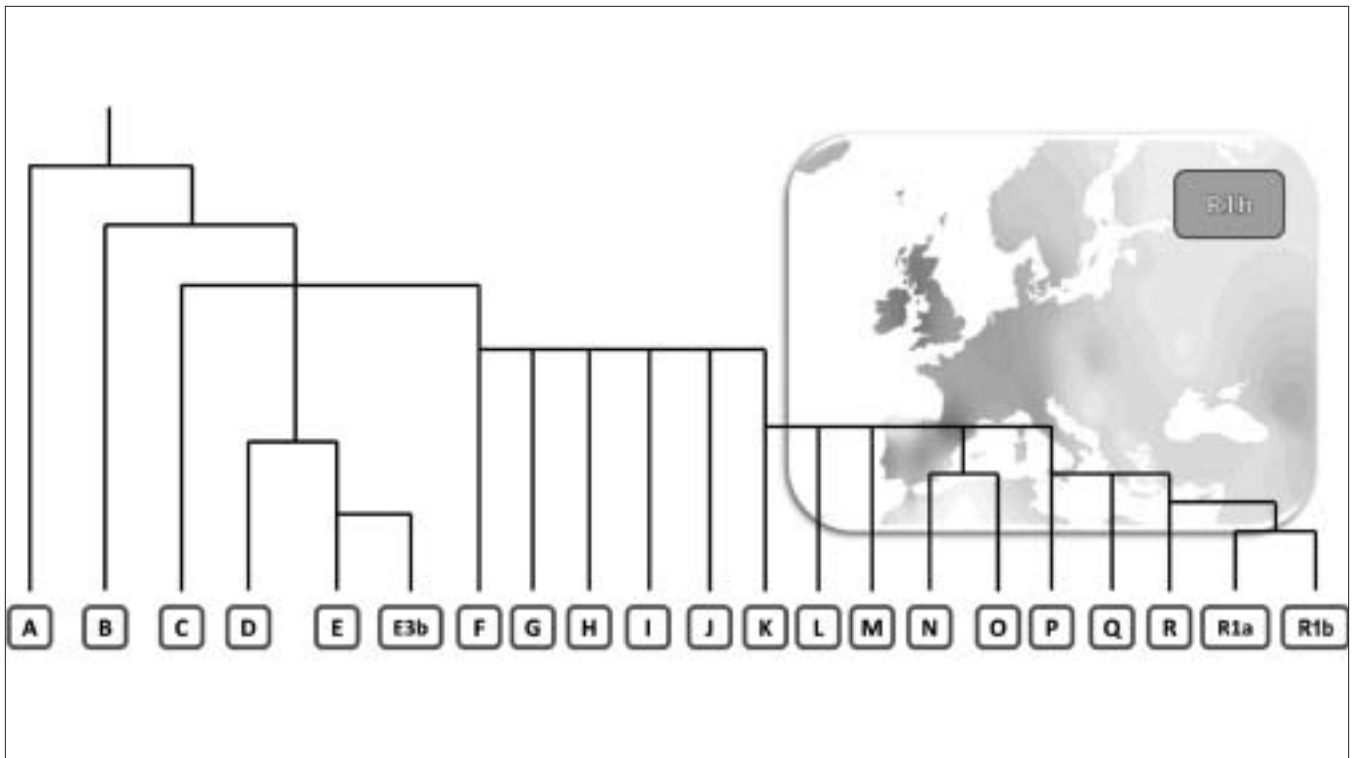


Figure 2: Schematic Y chromosome haplogroup tree with the frequency distribution map of haplogroup R1b in Western Eurasia.

Robert, and his numerous descendants eventually concluded that if Whittingham came from Lincolnshire, then most likely that was also the place where Robert lived prior to departing for the Western Hemisphere. However, no parish or other records documenting Robert Smith's birth in that geographic region have been found to date.

Genetic Evidence

As DNA samples and pedigree information were collected from the Utah descendants of this particular Smith family, the idea was proposed to employ DNA to help identify where in England Robert Smith came from. Although it was a long shot, the hypothesis could be tested. Because surnames are passed from fathers to their sons in western countries, the genetic data from the male-inherited Y chromosome (Ycs) could be used to identify potential distant Smith relatives in England. This is accomplished by first reconstructing the Ycs profile for the Utah Smith family and then comparing it with other individuals from England whose last name is also Smith.

Figure 1 is a schematic pedigree chart representing five continuous generations of the Smith paternal line (Robert, Samuel, Samuel and Asael, plus three of Asael's sons).

Since the Ycs is inherited along the direct male line (from a father to his sons), the first step was to identify a small number of living male descendants of Robert Smith and collect DNA samples from them. Ideally, these samples should come from separate paternal lines tying back to Robert Smith. However, only descendants of his great-grandson, Asael, were identified and tested (S1-S4 in Figure 1). Therefore, Asael Smith is the most recent paternal common ancestor (or MRCA) of all the individuals tested in this study.

An extended Ycs profile (called haplotype) was generated for each descendant tested and then compared to the others with the objective of reconstructing the genetic profile for their MRCA, Asael Smith. Testing more than one of Asael Smith's descendants was a necessary step in order to rule out possible undocumented non-paternity events (NPEs, such as

adoptions and illegitimacies), as well as resolving ambiguous situations involving random genetic changes (called mutations), thus inferring the most accurate Ycs profile for him. Of course, this would not rule out all the possibility for a NPE in the generations separating Asael and Robert Smith, or even earlier. This Ycs signature is just as accurate as if a DNA sample was taken directly from Asael Smith himself. This information turned out to be quite helpful in unlocking previously unknown details about the origin of the Smith line in Europe.

The Smith surname is the most common in the United States and England, accounting for about 1 percent of the total population in both of these two countries. It is traceable to a specific trade genre, that of workers of any kind of ore. Therefore, locksmiths, blacksmiths, etc. were all referred to as Smiths and their posterity eventually adopted this name as their identifier. Also, the most common first name in England in the 17th century was Robert. Looking for a connection involving a Robert Smith in the New and the Old World is like looking for a needle

DNA

in a haystack.

However, although both first and last name are very common, the genetic profile that was reconstructed for the Smith line was unique to the descendants of Asael (and presumably also of his great-grandfather, Robert Smith). When surveying thousands of records in databases containing Ycs profiles associated with Smith surnames, no meaningful matches were found for the Smith family in Utah.

In other words, every Smith individual with the same or very similar Ycs haplotype as of Asael Smith could be linked genealogically to him, but to no other Smith in the United States or England.

Irish or English?

Since using two variables (the Smith surname and the Ycs profile for Asael Smith) greatly limited the possibility of finding significant matches, and since the surname Smith is too common to be used as a single identifier, the next approach was to use, exclusively, the DNA markers to investigate whether DNA would link the Smith family from Utah to other family names.

A survey of genetic matches (leaving the surname Smith out of the equation) using the Sorenson Molecular Genealogy (SMGF) Ycs database at, www.smgf.org, provided the first relevant clues about the possible origins for the Ycs of the Smith's in Utah. The closest matches found were individuals with Asael Smith as their MRCA,

which was not surprising. The next close matches were mostly with people whose paternal ancestors were born in Ireland. Matches with ancestors from England and Scotland were found in much smaller numbers. A large number of close matches had genealogical data with a paternal ancestor that

are relevant in ancestral studies because they can be placed in a hierarchal tree (called phylogeny), with each branch showing clear geographic patterns of distribution. For example, Ycs haplogroup "I" is most common in the Scandinavian Peninsula, haplogroup "C" is found, typically, in certain areas

of Asia, haplogroups "A" and "E" are mostly found among males with African ancestry, and so on (see Figure 2). Haplogroups can be inferred from Ycs haplotypes or confirmed from testing individual mutations on the Ycs (called Single Nucleotide Polymorphisms or SNPs). Since it was assumed that Robert Smith came from Northeast England, it was important as part of this study to confirm the origin of his family Ycs by learning something about his haplogroup affiliation and frequency distribution. Using two free online haplogroup predictors, found at, www.GeneTree.com



Figure 3: The county of Donegal in Northern Ireland, where the Y chromosome for the Irish royalty was found at the highest frequency.

could not be traced outside of the United States. However, these close matches bore surnames that were typical of Ireland, such as Flanagan, McDowell, Dowda, etc.

To confirm a possible Irish origin for the Smith family in this case study, a deeper analysis of their Ycs was performed. Ycs profiles can be grouped in larger clusters called haplogroups based on genetic similarities. Haplogroups

and at, www.hprg.com/hapest5, the Ycs haplotype for Asael Smith was classified as belonging to haplogroup R1b, the most common Ycs lineage in Western Eurasia, with the highest frequencies observed in the Iberian Peninsula and in Great Britain (see Figure 2). The fact that R1b is very common in Great Britain provided an initial confirmation that the Ycs from the

Smith's in Utah could have indeed originated there. Next, a number of SNPs were tested for the known sub-lineages of R1b, until the deepest attainable branch was identified through a SNP called M-222 (a detailed list of SNPs and corresponding Ycs sub-lineages for haplogroup R1b is found at, http://www.isogg.org/tree/ISOGG_HapgrpR.html). Ycs DNA's testing positive for SNP M-222 belong to a sub-haplogroup called R1b1b2a1a2f2, which is a branch of the R1b family with its highest frequency (20 percent of the population) observed in the county of Donegal in Northern Ireland (Figure 3). Thus, both matches obtained from querying the SMGF Ycs database and SNP testing performed on the DNA of Asael Smith descendants strongly suggested that the Smith family from Utah could have originated in Northern Ireland, rather than England as previously thought.

Additionally, a paper published in 2006 by a group of scientists from Trinity College dealt with a particular Ycs haplotype that was found prevalently in Northern Ireland and that has been associated with the legendary warlord Niall of the Nine Hostages.

Presumably, Niall lived during the fifth century and among his posterity are the rulers of Ireland for the successive five centuries. The researchers involved in that study believed that the high frequency of this particular genetic signature in Northern Ireland is the DNA legacy that survived the royal Irish family and their offspring.

Comparison of the reconstructed Ycs profiles for Niall of the Nine Hostages and the Utah Smith family revealed a strikingly similarity. Out of 43 Ycs markers compared, only two one-step mutations were observed. Often, when a mutation occurs on the Ycs, it increases or decreases the marker value by a factor of one. Then, after few more generations, it can change again by the same factor, resulting in a two-step mutation.

Although not identical, the two Ycs profiles are most likely

closely related, indicating a possible common paternal ancestry within a relatively small number of generations. The Trinity College researchers included in their paper a list of common Irish surnames that have been associated with the Irish royal family and the genetic signature they proposed for Niall of the Nine Hostages. These surnames are very similar to those found in the SMGF database when queried using Asael Smith Ycs haplotype, thus providing further evidence that the Smith family in this case study likely has an Irish paternal ancestry.

Conclusions

Based on the three levels of genetic evidence described in this article (SMGF query, SNP testing, and Niall's Ycs haplotype), it is fairly evident that the genetic signature reconstructed for the Smith family in Utah does not have English, but Irish, origins. The haplotype reconstructed for this family is not very common and it is not found among other Smith DNA surname studies.

However, an abundance of this particular genetic profile is found in the Donegal region of Northern Ireland and it is shared by individuals who apparently descended from Irish royal rulers of the fifth to 10th centuries. This evidence, and the fact that Robert Smith was only 12 years old when he was brought to America as an indentured servant, provide the basis for further investigation into the family history of this particular Smith lineage.

A possible explanation is that Robert Smith was indeed Irish and that his family moved from Ireland due to economic distress. Once in England, the family could have changed their surname and adopted one of the most common English ones: Smith. Not being able to provide for their children, they were either abandoned in orphanages or sold as indentured servants to neighbor's landlords.

This would explain why young Robert did not have any family members with him when he began a new life in New England. Of course, these are all

assumptions, but the genealogical, historical, and genetic evidence taken together provide a synergistic view into the past of this family who, for generations, believed their ancestry to be English based on limited genealogical data.

The steps outlined in this article provide strong support toward a combined approach to family history where the joint use of traditional man-made documents and the information encoded in our DNA could reveal greater and previously unknown details about our pasts.

FC

Ugo A. Perego, Ph.D, is a senior researcher with the Sorenson Molecular Genealogy Foundation and a scientific consultant for GeneTree.com. You can contact him at, ugo@smgf.org