Genetic Testing of Leonard Y-DNA
Brad Leonard, bradsport.com

DNA testing for genealogical purposes is in its early years. For non-scientists, it is difficult to understand and even more difficult to explain. Some of the difficulty with creating this new field of genetics and genealogy has to do with creating uniform systems and methodologies for laboratory testing so that there is consistency in terminology and uniform quality with results that can be compared. Some of it is in the need to build data bases and classification systems so that results can be meaningfully tabulated, organized, and interpreted. Human DNA consists of about 3 billion base pairs in a double helix, each of consists of a pair of amino acids and is labeled as to location on the double helix. On occasion (think: hundreds to thousands of years), the base pair at a given location changes -- mutates -- and will replicate forever after in that changed state. The basic data consists of patterns of mutations in DNA

Numbers of Y-DNA Markers Tested and What They Tell Us

Tests can be ordered for 12, 25, 37, 67, or 111 markers. A marker is a physical location on the Y-chromosome. Each marker is assigned a name by the scientific community (e.g., DYS 393). At each location a short DNA code repeats itself. An allele is a DNA sequence that repeats at a certain location. The number of times the sequence repeats is its value (e.g., DYS 393=12). The more the number of markers tested and compared, the greater the statistical accuracy in pinpointing how many generations back two people diverged from a common set of parents.

Matches – that is, two people having the identical values for each marker – at the level of 12 markers will indicate a likelihood of a common male ancestor if both bear the same surname. Matches at the 37, or 67 marker tests levels indicate a near certainty of a common male ancestor back in time. At the 25 marker level, a match indicates there is a 95% probability that the common ancestor was 16 generations or less back in time.

As noted above, mutations do occur, noted by a change in value of a marker of one (e.g., a value of 11 for marker DYS 393 becomes 12). And as noted above, these mutations occur at statistically derived rates (for example, once every 500 generations, although some occur more frequently) and from those statistics, the range in the number of generations that have occurred since the mutation can be estimated.

Once a mutation has occurred, all of the descendants of that ancestor will carry that mutation, that change in value, for a particular marker. Over time, other mutations will occur in descendant lines. From those changes and the patterns they establish, a descendant tree can be drawn, and the branches of a given family can be identified.

Since mutations are occurring in different markers in different branches of a family, a genetic distance between branches is being established that is measured by the sum of the number of marker value differences between the two family members. For example, in the Leonard table at the end of this article, CB Leonard and RJ Leonard have a marker value of 12 in DYS 389/2, while the other Leonards in this group have a marker value of 11 in DYS 389/2. The genetic difference is therefore 1. CB Leonard and RJ Leonard have a common ancestor, Nathaniel, seven generations back. Comparing CB Leonard and FM Leonard, one finds they differ by 1 at marker DYS 389/2 and by 1 at marker GATA H4. Their total genetic difference -- genetic distance -- is therefore 2 at the 37 marker level, signifying a common ancestor. Genealogical research indicates that their most recent common ancestor was James, as they each descend from different sons of James.

Generally speaking, a genetic distance of 1, 2, 3, or even 4 in a test of 37 markers or more indicates the likelihood of a common ancestor back in the family tree. Beyond 4, it becomes in doubt, at least for the time period where we can trace ancestry; and at 6 and beyond, it seems unlikely that there is a common ancestor since we of English ancestry began using surnames about 1100 to 1200 AD.

As an aside: a history course taken recently indicated that the Church was discouraging people from marrying if they were 1st, 2nd, 3rd, 4th, 5th, or 6th cousins in the years around 1000 AD. How people then who were known by only one name, say, Robert, could keep track of their family trees with sufficient detail to know who their 6th cousins were challenges one's imagination!

This information is being gathered at the level of 12 markers. Further genetic testing can be done at the level of 25, 37, 43, 67, or even 111 markers, which can be useful in narrowing down family relationships in genealogically meaningful timeframes (e.g. back to 1600, or back to when surnames became prevalent in England around 1200 AD).

Haplogroups

A haplogroup is a term used for particular patterns of markers that identify a major population group.
"Each of us carries DNA that is a combination of genes passed from both our mother and father, giving us traits that range from eye color and height to athleticism and disease susceptibility. One exception is the Y-chromosome, which is passed directly from father to son, unchanged, from generation to generation.
"Unchanged, that is, unless a mutation occurs – a random, naturally occurring, usually harmless change. The mutation, known as a marker, acts as a beacon; it can be mapped through generations because it will be passed down from the man in whom it occurred to his sons, their sons, and every male in his family for thousands of years.

"In some instances there may be more than one mutational event that defines a particular branch on the tree. What this means is that any of these markers can be used to determine your particular haplogroup, since every individual who has one of these markers also has the others.

"When geneticists identify such a marker, they try to figure out when it first occurred ("sequencing"), and in which geographic region of the world. Each marker is essentially the beginning of a new lineage on the family tree of the human race. Tracking the lineages provides a picture of how small tribes of modern humans in Africa tens of thousands of years ago diversified and spread to populate the world.

"A haplogroup is defined by a series of markers that are shared by other men who carry the same random mutations. The markers trace the path your ancestors took as they moved out of Africa."[i]

Haplogroups are broken down into subclades and labeled by the same process. The National Geneographic Project, a five-year worldwide effort sponsored by the National Geographic Society, IBM, and the Waitt Family Foundation, is testing populations in all countries on all continents to gather data on haplogroups, their subclades, and individuals. The origins of the J haplogroup are in the Middle East, perhaps in what is now Iraq, 10,000 years ago. The mutations that formed subclade J2 occurred later, perhaps 5,000 years ago, with further mutations creating subdivisions forming subclade J2ah4 2,000 years ago around the Mediterranean in Turkey, Greece, Lebanon, and the Holy Land. This sequencing of mutations is going on in all of the haplogroups, requiring new and changing nomenclature to identify them. Indeed, one of the haplogroup R1b's in the Leonard group has been identified as an R1b1a2a1a1a4a1a1 in the latest haplogroup tree. Because such designations were getting unwieldy, haplogroup subclades are now abbreviated to the last SNP mutation, for example, J-L70.

Leonard Haplogroups

There have been over 100 Leonards in FamilyTreeDNA's Leonard surname project who have had their y-dna tested for at least 12 markers,[ii] and four from other testing services whose results would place them in the same haplogroups as the 100+.

Those 100+ sharing the Leonard surname have been distributed through testing into the E3a, E3b, I, T, J2, and R1b Haplogroups. About half, or 35, are in the subclade R-M269. The R haplogroup is the most common in Northern Europe, with about 120 million or more living members. For example, in some Irish towns, 85% of the male population are in haplogroup R1b.

It would appear that Leonards of Irish, Welsh, English, and Scotch descent generally fall in the R1b modal haplotype, or, in the latest nomenclature, R-M269. This is the most common haplotype in the British Isles and Ireland.

There aren't enough Leonards tested in the E3a, E3b, and T haplogroups and not enough is known about their ancestry to say very much about them yet.

Seventeen English Leonards fall into the J, more specifically, J2 haplogroup. Fourteen of these whose genealogy has been traced back to the 1600's are descendants of James Leonard, 1620-91, of Taunton, Massachusetts, or his brother, Henry, 1618-aft. 1678, of New Jersey. One descendant of John Leonard of Springfield has tested in haplogroup J2. Two living in England are descendants of an ancestor of James and Henry who never left England.

The J2 haplogroup is relatively rare in the United Kingdom and Ireland. "Today, descendants of this (the J2) line appear in the highest frequencies in the Middle East, North Africa, and Ethiopia, and at a much lower frequency in Europe, where it is observed exclusively in the Mediterranean area. Approximately 20 percent of the males in southern Italy carry the marker, along with 10 percent of the men in southern Spain."[iii]

Given the relative rarity of the J2 haplogroup among the English, Welsh, Irish, Scotch, German and Scandinavian populations, this marker may be particularly helpful in identifying the common ancestry of those who trace their descendancy from James Leonard. What we have learned or can learn from genetic testing:

- Henry Leonard's descendants should also be J2's. Henry was James' brother. Two of Henry's direct male descendants has been tested, and they are, indeed, J2's.
- Five of Solomon Leonard of Duxbury's direct male descendants have been tested. They are in haplogroup I2b, subclade I-M223. This would indicate that they are not related to James and Henry Leonard.
- If direct male descendants of John and William Leonard can be found who are willing to get their y-dna tested, we will learn whether they have a common ancestry with James and Henry or stem from other early Leonards.
- We may be able to establish that other branches belonged to other haplogroups (e.g., R1b1) and subclades within those haplogroups easily identifiable through Y-DNA testing, thereby simplifying genealogical research. We have identified the American descendants of Robert Leonard 1745 of Scotland, for example, as well as both the American and English descendants in R1b1 of an early Samuel Leonard of Somerset, England.
- Branches of American Leonard families still residing in England or other countries of origin have been identified. Stuart Leonard of Birmingham and several of his relatives test as J2's. They must descend from a common ancestor before James and Henry.
- Contrary opinions about the ancestors of James and Henry may be resolved (whether they were effete English aristocrats or "truculent and profane" French ironworkers). This will depend upon whether we can find J2 Leonards with the same y-dna patterns in France or Belgium.
- Verification of genealogical research in the various branches of James Leonard's descendants as their subpatriarchs and the characteristic patterns of their branch lines are established. For example, the CB Leonard and RJ Leonard match described earlier indicates that the same Nathaniel Leonard had families by two different wives: one in Middleborough, MA, before 1807, and one in Ontario, Canada, after 1807.

There are several instances where people with other than the Leonard surname appear in the J2 haplogroup with a pattern of markers exactly or almost exactly matching those of James Leonard's descendants. In one case, for example, the person traces his ancestry back to a Gilbert Worden who lived in Becket, Massachusetts, in 1813. His y-dna pattern doesn't match the marker pattern of other Wordens. There appear to be at least four possible answers in this situation:

- The most recent common ancester was before the Leonards adopted the surname Leonard (which appears to have occurred in the 1400's or 1500's for James Leonard's ancestors).
- There was a name change in the family.
- There are other families with exactly the same pattern of markers as the Leonards.
- An ancestor of the person was actually a Leonard who was adopted by a Worden or other family. Formal and informal adoptions occurred more frequently in the past when people died at a younger age.
- An ancestor of the person was actually fathered by a Leonard in an extramarital relationship.

The latter two are what DNA Heritage (http://www.dnaheritage.com/tutorial4.asp) calls "non-paternity events." The latter occur at about 2 to 5% of births in every generation.

There are other Leonards who have traced their ancestry to James Leonard but find that they're not in haplogroup J2 nor do they have a close match of markers. This is perplexing because, as yet, we haven't figured out the cause, which could be as simple (or complex) as a mistake in tracing ancestry or as confounding as a non-paternity event. Given that the Leonards in America used the same given names and lived in the same areas, a mistake in the genealogy would be easy to make, yet hard to detect.

Matching Markers

Here is the 12 marker pattern for Leonard J2's tested:

Locus	DYS#	15 Leonards Alleles	J2	2 J2 Leonard Alleles
1	393	12		12
2	390	23		23
3	19*	15		15
4	391	9		9
5	385a	13		13
6	385b	17		17
7	426	11		11
8	388	18		18
9	439	11		11
10	389-1	13		13
11	392	11		12
12	389-2	29		29

The allele values in the first column are for Leonards who descend from James Leonard and 1 Leonard in England whose ancestry before 1750 is unknown, but whose ancestors were gunsmiths. The two Leonards in the second column trace their ancestry back to James but differ by one marker from the other Leonards. A descendant of Henry differs by one at another position not shown able: DYS458. As you look across the table on the referenced Leonard surname chart, you find a number of instances where there's a difference of one in almost every instance. These differences will begin to identify the various branches of Leonards descending from James and Henry and their antecedents.

The Leonard surname table referenced on the Internet shows markers out to 67. There is a problem for the J2 Leonards in the so-called Palindromic Region, with markers DYS464 a, b, c, d, e, f, and g. These differences are particularly evident in markers DYS464 d, e, and f. FamilyTreeDNA was using a lab that had a different procedure for identifying these markers. They later changed to a lab with a procedure that met the newly updated standard. So those who were tested before 2007 show a slightly different pattern than those who were tested or retested later.

These tests yield results which are usually measured in the probability that someone with a given pattern of matching markers will have a "most recent common ancestor" (MRCA) with someone else in x generations. Matching 12 of 12, for example, yields a probability that the MRCA was no more than 7 generations back and a 90% probability that the MRCA was no more than 23 generations back. Testing for more markers reduces the number of generations need to reach a 50% or 90% probability for the MCRA. For example, matching 37 of 37 markers would put the 50% probability at 4 generations back, while matching 67 of 67 would put it at 2 generations back. But mutations can occur at any time. I know of a case where brothers differ (mismatch) by 1.

Please note: These are statistical measures of probability. Based on your genealogical research, you may have identified the MRCA in 5, 10, or 15 generations back. The y-dna analysis supports what your genealogical research tells you (or perhaps tells you you are barking up the wrong family tree). It does not guarantee an ancestor, nor does it supply his name.

As more Leonards are tested for more markers (25, 37, 43, or 67 markers) beyond the basic 12 and can supply their genealogies back to James or before, we may be able to make some general hypotheses about what matches (e.g., 35 out of 37, 36 of 37, 37 of 37) mean in terms of the number of generations back a common ancestor may be found.

Most of us have used the FamilyTreeDNA laboratory in Houston, Texas (http://familytreedna.com). This lab offers competitive prices and services, sends their results to the National Genographic Project, and has a small discount for those in family surname projects (think "Leonard"). But there are other labs, such as AncestryDNA, that offer comparable prices and services.

In time, perhaps we will find out which, if any, characteristics (besides sex) are carried along by the Y-chromosome. The information stored on the Y-chromosome is the one piece we certainly share in common. Perhaps we will find that our intelligence, good looks, athleticism, charm, and, above all, modesty are all carried by the Y-chromosome to all us Leonards!

Updated 1 December 2013

---

[i] National Geographic web site – National Genographic Project – for more information, see http://www.nationalgeographic3.com/genographic/.

[ii] See http://www.worldfamilies.net/surnames/leonard/results.html for a full listing.

[iii] National Genographic Project