Ewing Surname Y-DNA Project Article 8

This is the eighth in a series of articles about the Ewing Surname Y-DNA Project. The previous seven articles have appeared in the last seven issues of the *Journal of Clan Ewing*. They are also available on-line via the Ewing Surname Y-DNA Project link on the Clan Ewing WebSite's Home Page (*http://www.ClanEwing.org*).

DNA Talks at the Gathering

I was sixteen kinds of nervous before giving the DNA talks at the gathering, because as hard as I've worked on this, I am still a rank beginner. The topic is so vast and intricate that no one can fully master or adequately explain it, even without time constraints. I was relieved that the talks were well received, at least by everyone who commented aloud. I want to thank equally those who spoke and those who remained silent for their kindness. I was chagrined to discover that many folks who understood the talks have found my DNA articles in the *Journal* too difficult to understand. Joe Neff Ewing, Jr. stood up after the first talk and said something like, "Well, David, I have to say that I feared the worst after reading your articles, but with the talk you gave here today, I think I'm beginning to understand how DNA can be useful in genealogy."

I accept full responsibility for not being clear, of course, but the task of explaining this rich field is more difficult than you might imagine. I tried to give some background with the first few articles and to build on that in subsequent articles. The problem is that the *Journal* comes out only every three months and very few members have had the time, interest and patience to keep rereading the background articles. I plan now to begin working with William E. Riddle, our WebMaster, to post a kind of DNA tutorial on the WebSite that will allow interested folks to easily find background information and answers to specific questions in a way that will let everyone learn at his or her own pace. Also, I am always happy to receive EMails or phone calls from anyone who wants to discuss these issues privately.

New Results

We gave out nine new collection kits at the gathering, six of which have already been returned and are being processed; and, we have another couple of new members who joined through the FtDNA WebSite and are waiting on results. We have partial results on one of these men, but it makes sense to me to wait until the next issue of the *Journal* to report more results. We will be updating the WebSite with results when they come in; most are due by the middle of December.

Questions

In addition to the talks at the gathering, I've given an introductory genetic genealogy talk to the Albuquerque Genealogical Society since returning home from the gathering in Fort Wayne. One question that keeps coming up involves what can be learned from different kinds of DNA tests. Everyone has read about the use of DNA in criminal investigation, in proving paternity, and in diagnosing, predicting, and even treating disease. Many people have read about DNA tests in very old archeological specimens and DNA tests that are used to reconstruct ancient migration paths of human beings around the world. Some folks have read about DNA tests that can show what fraction of several different ethnic groups may be represented in the DNA of a person who is tested. This is all interesting stuff, but it is not what we are doing in the Ewing Surname Y-DNA Project

To understand what we are doing, and how it differs from other uses of DNA testing, it is important to understand that we don't just "test" DNA and get a result that will serve all of these purposes. Rather, a DNA test is something like asking a specific question of the DNA. If you don't know what you are asking, you won't get a helpful answer. All DNA tests depend on looking at specific places in DNA where there are known to be different possible results. We call the results we find in these places "markers." Imagine that! There are over three billion base pairs in the human genome, and DNA testers can check to see if a specific one of them is where it is supposed to be.

Although this may sound impossible, it is not even all that difficult. But explaining how it is possible is not so easy unless you are talking to someone with a pretty good background in microbiology. The important thing to understand is that when we check the DNA for something, we are not just browsing around; we have to know exactly where to look and what we are looking for. Mostly, what we are looking for are either SNPs or

STRs.[1] When we look for an SNP, we are looking at exactly ***one*** of the over three billion base pairs in the DNA to see which of four possible "letters" is in exactly that spot. When we look for an STR, we look at a very specific spot on the DNA and count how many repeats of a short sequence of "letters" we find there. In either case, we call what we find in the spot we are looking at a "marker." This very specific and focused inspection for certain markers is what we are doing in all DNA tests, regardless of what kind of a DNA test it is that we are doing.

The DNA sequence in all human beings is remarkably similar. After all, this is what makes us human. Consider that DNA sequences differ between humans and chimpanzees by less than two percent, and the DNA sequences of any two human beings differ by less than one tenth of one percent. All of us have practically identical DNA. But since we are talking about 3.2 billion places along the DNA where we could differ, a tenth of a percent still allows for each of us to differ from the next guy at over 3 million places. You mathematicians out there might want to run the permutations and combinations on that mind-boggling prospect, but the bottom line is that no two of us have perfectly identical DNA. We are 99.9% identical to one another, but still all different. Incredible, isn't it?

Identity and Paternity Testing

So how do we prove that hank of hair we found at a crime scene belongs to a suspected bad guy? What if we checked one base pair in the hair sample and one in the bad guy and found an exact match? This would tell us exactly *nothing*, because the chance of a random individual matching the bad guy at a random location on the DNA is 99.9%. Finding a match at a random marker would be something like telling the court that the suspect had five fingers, which is useless information because nearly everybody has five fingers. So, what if we completely sequenced the DNA in the hair sample and in the bad guy and compared the results? That would give us positive identification, but it would take months or years of work and would cost about a bazillion dollars.

On the other hand, if we knew of a spot to look where there were five different possibilities[2] of equal likelihood, whatever result we got there could rule out 80% of people as suspects, because only one in five people would have the same marker at that spot. In fact, what is done is that a relatively short list of markers[3] that are known to vary a fair amount in different individuals are checked. This list of places has been selected so that there is a vanishingly small likelihood of two individuals matching at all of them.

Paternity tests use the same principal, but the math is a little more complicated — you have to divide by two to start with because the mother has contributed fifty percent of the child's DNA, and then you have to take into account that the mother's and father's DNA has been "mixed" in a complicated way and also that there is a small chance of a new mutation having occurred. Even so, one can choose a list of places to look at so that the odds of a child and a man matching on a given fraction of them are virtually zero unless the man is the child's father.

I think you can see how it would be easier to prove that an individual is <u>not</u> the bad guy or <u>not</u> the daddy than it is to prove that he is. And you can see how it would be harder to determine which of two brothers is the father of a child than which of two unrelated men is the father. I'm no expert in forensic testing, but my guess is that if we are trying to distinguish brothers, we will need to check many more places than we otherwise might have to do.

Fractional Ethnic Ancestry

At least two different outfits offer DNA testing and return results that are supposed to be able to tell you what percentage of your ancestors were members of a list of different ethnic groups. In this discussion "ethnic group" is not so specific — we are not going to be able to distinguish Norwegian from German or Spanish from Italian, but we can distinguish Sub-Saharan African from Native American from European from East Asian.

Each of us has eight great-grandparents. Remember that Y-DNA is passed only from father to son, and mitochondrial DNA (mtDNA) is passed only from mother to child. Y-DNA can tell a man something about his father's paternal grandfather, and mtDNA can tell a person something about his or her mother's maternal

[1] There is a lengthy discussion about STRs (short tandem repeats) and SNPs (single nucleotide polymorphisms) in my DNA Article 7, which was published in the previous issue of the *Journal* and is available on the *Clan Ewing* WebSite.
[2] Plainly, there can't be five possibilities at a SNP; this would have to be an STR.
[3] We're talking about something like 20 places. And I am oversimplifying here, too, because most or all of the "places" checked are not single base pairs, but rather STRs. For forensic work like this, the STRs checked are not all on the Y-chromosome as in our project — they are on several different chromosomes.

grandmother, but neither Y-DNA nor mtDNA can tell us anything about a person's other six great-grandparents. To learn anything about them, we need to have a look at the autosomes.

"Autosomes" are the 22 pairs of nuclear chromosomes that are not sex chromosomes. Because of a phenomenon called "crossing over," which occurs as chromosomes are copied to make sperm and eggs, autosomes get mixed up and specific lines cannot be traced. Even so, when a marker is found in an autosome, we can be sure that it came from one of the eight great-grandparents, even though we may not know which one.

Some markers are found in a relatively high percentage of people in one ethnic group and a relatively low percentage in other ethnic groups. Testing of this kind requires quite a number of markers, so costs a little more — the test from *DNA Print* that claims to be able to differentiate among some European populations costs about $400.

The Ewing Surname Y-DNA Project tests only the Y-chromosome and so gives no information about any part of participants' ethnic backgrounds except their strict paternal lines.

Genetic Disease

There is no single DNA test that can be done to answer the question, "Does this person have any defective genes?" We can only check the genes one at a time, though of course we could have a look at ten of them, one at a time.

We could in principle look at all 30,000 of the known human genes one at a time, but as a practical matter this would be far too expensive and time consuming to even consider doing. And even so, we could not check any of the maybe 50,000 or so other genes we think exist but that have not been fully characterized, without first learning exactly where they are and what they are supposed to look like.

Tests for genetic disease are only done when there is a reason to suspect a specific genetic problem. For example, the location and nature of the genetic defect that causes Huntington's disease is known. Each child of a person who has Huntington's disease has a 50% chance of having the disease. Children of two parents who do not have Huntington's disease have no chance of developing Huntington's disease. You can see that it might make sense to test the child of a person with Huntington's disease for the gene, but it would not make sense to test the child of parents who do not have the disease.

The ethics of doing such testing and what to do with the results obtained are complex and interesting, but discussing this is beyond the scope of this article. What is important to realize is that the testing we do for genealogical purposes has essentially zero medical applicability — in the Ewing Surname Y-DNA Project, we are specifically looking at markers which are <u>not</u> in genes and have <u>no</u> effect on the health of the person who carries them.

Maternal Line Testing Using Mitochondrial DNA (mtDNA)

Mitochondrial DNA (mtDNA) is passed from mother to child, both boys and girls. Men have mtDNA identical to that of their mother, maternal grandmother, etc., but they do not pass it to their children. Only women pass on their mtDNA.

The mutation rate in mtDNA is so slow that it has very little use in genealogy, though it can give interesting and informative results regarding deep ancestral origins, and it could be used to disprove a putative relationship between cousins thought related in the maternal line.

What is actually tested in mtDNA is the so-called "D-loop" or "hyper-variable region" of mtDNA. This area may be "hyper-variable" compared to the rest of the mtDNA, but the mutation rate is still over a hundred times slower than the mutation rate of the STRs we look at on the Y-chromosome.[4] As a result, if two people have an exact match at all 1050 tested mtDNA markers, there is only a 50% likelihood that they have a common female ancestor within the last 12 generations and a 90% likelihood that they have a common female ancestor within the last 36 generations or so. A 1049-out-of-1050 match implies that there is a 50% likelihood that the common female ancestor lived within the last 26 generations or so.

[4] The mutation rate at each of these 1050 places is about 0.00003, which is 0.003% – one per 33,333 generations. Just to give you a frame of reference, the mutation rate at the STR markers we are testing in the Ewing Surname Y-DNA Project is about 0.4%, which is one in 250 generations, over one hundred times faster. Wowsa! We have 1050 tested mtDNA nucleotides being copied each generation, so the chance of a mutation at any one of them at each generation is 1050 x 0.003% = 3.15%. In the Ewing Surname Y-DNA Project, we have been mostly testing 37 markers, so there is a 37 x 0.4% = 14.8% chance of a mutation at any of the 37 markers each generation.

You can see that we are not talking about a genealogical time frame here. In my own case, mtDNA testing revealed that I am in haplogroup T2, which means that my maternal lineage originated in the Near East more than 45,000 years ago, and came into Europe about the time that agriculture did around 10,000 years ago. The conventional genealogy for my maternal line reaches back only to my great-grandmother, Maria Frederike Hjelm, who emigrated from Sweden at the end of the 19th century, though probably her ancestors had lived in Sweden from time immemorial. I probably have exact mtDNA matches with people all over Europe, but they certainly wouldn't be close cousins.

DNA Testing in Archeological Specimens

Like any living thing, DNA gradually deteriorates after death. Mitochondrial DNA (mtDNA) is simpler, more abundant, and less subject to degradation than nuclear DNA, but even mtDNA gradually breaks down and disappears.

No one is ever going to be able to reconstitute dinosaur DNA from hundreds of millions of years ago, no matter what you read in *Jurassic Park.* Sometimes, though, under special conditions (being frozen or buried in a cool, dry, alkaline limestone cave) some DNA can remain in recognizable form for hundreds or even thousands of years.

Even when some DNA remains, in a very old specimen most of it will have disappeared. This means that the tiny amount remaining will have to be artificially copied and "amplified" many orders of magnitude in order to perform an analysis, and this creates the danger that a trace amount of modern DNA from the fingerprint or bit of dandruff of a researcher will have contaminated the archeological specimen and get amplified in its stead.

Claims have been published that fragments of DNA have been recovered from Neanderthals, the last of whom died out some 30,000 years ago. This work is fascinating, but what we are doing in genetic genealogy doesn't depend on analyzing old specimens of DNA.

Even if we knew for sure where John Ewing of Carnashannagh is buried (and some folks think it could be at the family burial plot in Stevens City, Frederick County, Virginia, near Winchester where the 2008 Gathering of *Clan Ewing* will be held), and even if we had the temerity to dig up his bones (which we emphatically do *NOT*), it is very unlikely that a useful specimen of DNA could be recovered now, 261 years after his death.

Fortunately, the fact is that because we now have collected DNA specimens from twelve of his living descendants, we can know his Y-DNA haplotype just as surely as if we could test his DNA directly. We can do this by remembering that mutations (changes in markers) are rare events. We compare the results of our living participants and construct the haplotype of their common ancestor such that the differences among them can be accounted for by the minimum number of mutations possible. Mathematically, this is the "modal" haplotype of a group of related men,[5] and the odds are very good that this matches the haplotype of their common ancestor.

Conclusion

DNA testing may be used to address many different issues, but it can only give answers to very focused and well-designed questions.

In the Ewing Surname Y-DNA Project, all the questions we ask pertain solely to the paternal line and are designed to give us as much information as possible about how long ago the common male ancestor of two living men may have lived.

We learn nothing of medical interest and we get no information about anything but the surname lineage, so no information about fractional ethnic ancestry or the like.

To Join or Get More Information

If you are ready to join the Ewing Surname Y-DNA Project, go to:

www.familytreedna.com/public/ewing

and click on "Join this group" at the top of the blue section on the left of the page. Participation by Ewing women is also welcome; they can get valuable genealogical information by persuading a male relative to submit a specimen. You can see result tables showing participant haplotypes, genetic distances and time to

[5] You can read about "haplotype" and "modal haplotype" in more detail in the Ewing Surname Y-DNA Project article in Vol. 11, No. 3 of the *Journal of Clan Ewing* (August 2005), which is posted on the *Clan Ewing* WebSite (*www.ClanEwing.org*) as DNA Article 3.

the most recent ancestor estimates expressed as the number of generations on the *Clan Ewing* WebSite. There are also links on the *www.FamilyTreeDNA.com* WebSite to articles and FAQs. If you want to ask questions, call me at +1 505.764.8704 in the evening, or EMail me at *DavidEwing93 at gmail.com.*