Ewing Surname Y-DNA Project — Article 3

This is the third in a planned series of articles about the Ewing surname Y-DNA project. The first and second articles appeared in the last two issues of the *Journal of Clan Ewing*, and they are also available on-line at
http://www.clanewing.org/Y-DNA.html.
Understanding this article will require that one has a reasonable understanding of the information in the first two articles.

Markers

Recall that what we are testing on the Y-chromosome is the number of "microsatellite repeats" at each of up to 37 loci (places). Microsatellite repeats are also called "short tandem repeats" (STRs), but we call them "markers." It is not important to understand exactly what these are, but it is important to know how they are reported and how they change when a mutation occurs. Each marker has been assigned an arbitrary name. The name is often "DYS" followed by some numbers or other letters and numbers. These are just names of markers. There are 37 markers in our study; each has a different name.

Haplotypes

Each person will have a specific number at each marker. As we receive them, these data are posted online at the bottom of the page at
http://www.familytreedna.com/public/ewing/.
Periodically, we also update results tables that are posted on the website of Clan Ewing and are easier to read and interpret. As I write this, we have received results on 15 participants in the Ewing surname Y-DNA project, as shown in the Table on Page 3.

Across the top of the table,[1] you can see the names of each of the markers. Down the leftmost column are the ID numbers and initials of the participants.[2] If you follow one of the participants horizontally across the page, each number you come to is the number of repeats at the marker that is named at the top of the column. Take a look at the data on my Y-DNA results on the row labeled "26605 DN." Reading from left to right, you can see that I have 13 repeats at DYS 393, 25 repeats at DYS 390, 15 repeats at DYS 19, 11 repeats at DYS 391, and so on all the way across the page to a total of 37 markers. That list of numbers is called my "haplotype." You can see that Chancellor George Ewing (26607 GW) has exactly the same haplotype as I do—his numbers are exactly the same. A

[1] Both tables in this article were prepared using the Y-DNA Comparison Utility offered for free use by Dean McGee at http://www.mymcgee.com/tools/yutility.html. It was an ENORMOUS help. These tables have differences highlighted in color, but as the print version of the Journal is in black and white, they will be a little harder to read. You might prefer to read this article on line at http://www.clanewing.org/Y-DNA.html, where the type is larger, the colors will show and the links will work.

[2] If you check the FTDNA website for new data, be aware that I have changed the order of participants in the table here a little.

haplotype is just a list of numbers that correspond to the number of microsatellite repeats at each marker.

Mutations

Ordinarily, when a father copies his Y-DNA, he will pass on to his sons exactly the same number of repeats that he has at each marker; so his son will have exactly the same haplotype as he does. Once in a great while a copying error is made at one of the markers. What is "a great while?" It is estimated that on average an error is made at a given marker once in every 500 generations! So how in the world can that give us any genealogic information? Well, we are looking at 37 markers, each of which can have a mistake once in 500 generations, so a mistake will be made in one of the 37 on average in 500/37 = 13.5 generations (note that 14 generations separate 6th cousins—seven generations up to the 5th ggf and seven generations back down to the 6th cousin). And please notice that I said "on average." Some markers are known to have a higher mutation rate. The markers shown in red on the data table on the FTDNA website[3] are known to have a higher than average rate of mutation—maybe something like once in 250 generations. And we think that older fathers have a higher mutation rate than younger fathers. And some families may have a higher mutation rate than others. The bottom line on mutation rates is that geneticists are still trying to work this out quantitatively, and for now we are stuck with working with estimates, based on an "average" marker, even though we have evidence in our own data that mutations have occurred at a rate much faster than average.

Modal haplotype

Let me introduce another concept, the "modal" haplotype. We have 15 haplotypes, one each on 15 men (remember, a haplotype is just a list of numbers telling how many microsatellite repeats there are at each marker). Let's have another look at the data table above. Each DYS locus has a column of numbers under it that shows how many repeats each man has at that locus. We can see which number turns up most frequently in each column. Under DYS #393 there are all 13s, so that is a no-brainer—the modal number is 13. Under DYS #390 there are four 24s and eleven 25s—the modal number is 25. We can repeat this process clear across the page, and make a note of each modal number in turn. When we get done we have a list of modal numbers, which is the modal haplotype, winner of a genetic popularity contest. Now, this modal haplotype is just a construct; it doesn't belong to any specific man. An argument could be made that the modal haplotype is an approximation of the haplotype of the common male ancestor for the whole group, but the reason I've done this is just to find a place in the center of things to stand so I can see how far from the center each of the project participants lies.

[3] http://www.familytreedna.com/public/ewing/

DNA Results Comparison																																					
ID	DYS393	DYS390	DYS19/394	DYS391	DYS385a	DYS385b	DYS426	DYS388	DYS439	DYS389-1	DYS392	DYS389-2	DYS458	DYS459a	DYS459b	DYS455	DYS454	DYS447	DYS437	DYS448	DYS449	DYS464a	DYS464b	DYS464c	DYS464d	DYS460	GATA-H4	YCA-IIa	YCA-IIb	DYS456	DYS607	DYS576	DYS570	CDYa	CDYb	DYS442	DYS438
modal	13	25	15	11	11	13	12	12	13	13	14	29	17	9	10	11	11	25	15	18	31	15	16	16	17	11	11	19	23	18	16	18	17	37	38	11	12
30341 JMc	13	24	14	11	11	15	12	12	13	13	13	29	18	9	9	11	11	25	15	19	29	15	15	15	15	11	11	19	23	16	14	19	18	36	36	12	12
34979 JM	13	24	15	11	11	15	12	12	12	12	13	28	18	9	10	11	11	25	15	19	29	15	15	17	17	11	12	19	23	16	14	18	17	36	38	12	11
29038 DS	13	24	15	11	12	15	12	12	12	12	13	28	17	9	10	11	11	25	15	19	29	15	15	17	18	11	12	19	23	16	14	18	17	36	38	12	12
32056 TD	13	25	14	11	11	13	12	12	12	13	14	29	18	9	10	10	11	25	15	18	30	15	16	16	17												
35063 WC	13	25	15	11	11	13	12	12	12	13	14	29	17	9	10	11	11	25	14	18	31	15	16	16	17	11	11	19	23	18	16	17	17	37	38	11	12
27497 EN	13	25	15	10	11	13	12	12	12	13	14	29	17	9	10	11	11	25	15	18	31	15	16	16	17	11	11	19	23	18	16	17	17	37	38	11	12
28271 RL	13	25	15	11	11	13	12	12	12	13	14	29	17	9	10	11	11	25	15	18	31	15	16	16	17	10	11	19	23	18	16	19	17	37	38	11	12
34634 EG	13	25	15	10	11	13	12	12	13	13	14	29	17	9	10	11	11	25	15	18	31	15	16	16	17	11	10	19	23	18	16	18	17	37	39	11	12
32942 WR	13	25	15	10	11	13	12	12	13	13	14	29	17	9	10	11	11	25	15	18	31	15	16	16	17	11	11	19	23	18	16	18	17	35	38	11	12
26860 RC	13	25	15	10	11	13	12	12	13	13	14	29	17	9	10	11	11	25	15	18	31	15	16	16	17	11	11	19	23	18	16	18	17	37	38	11	12
33681 WK	13	25	15	11	11	13	12	12	13	13	14	29	17	9	10	11	11	25	15	18	31	15	16	16	17	11	11	19	22	18	15	18	17	37	38	11	12
35883 BE	13	25	15	11	11	13	12	12	13	13	14	29	17	9	10	11	11	25	15	18	31	15	16	16	17	11	11	19	22	18	16	18	17	37	37	11	12
26605 DN	13	25	15	11	11	13	12	12	13	13	14	29	17	9	10	11	11	25	15	18	31	15	16	16	17	11	11	19	23	18	16	19	17	37	38	11	12
26607 GW	13	25	15	11	11	13	12	12	13	13	14	29	17	9	10	11	11	25	15	18	31	15	16	16	17	11	11	19	23	18	16	19	17	37	38	11	12
28942 RA	13	24	15	11	11	13	12	12	13	13	14	29	17	9	10	11	11	25	15	18	31	15	16	16	17	11	11	19	23	18	16	19	17	37	38	11	12

Distance from reference:	Zero	One	Two	Three+

Genetic Distance

A concept crucial to understanding the analysis of the test results presented in this article is "genetic distance." There are at least two ways to define genetic distance. The first, the "infinite alleles" model, looks at two haplotypes and just counts how many of the markers differ—the genetic distance is just the number of differences, and the sizes of the differences are ignored. The other method takes into account that almost all microsatellite mutations are one step by counting not only the number of differences, but also the size of the differences. In both methods, a one step mutation is counted as a genetic distance of 1. In the first method, a two step mutation (or two one step mutations at one marker) also counts as a genetic distance of 1, but in the second method this would be counted as 2. Even more complicated schemes have been devised to take into account the relative frequency of two step mutations and the differences in the rates of mutation at different markers, but we don't need to concern ourselves with that here. It turns out that for our purposes it doesn't matter which of these methods we use. Whichever method is used will give us numbers for genetic distances such that the closer two haplotypes are to one another, the smaller the genetic distance will be. 37-marker haplotypes at genetic distance of 1, 2 or 3 are likely to be related within the period of genealogic interest; those at genetic distance of 6 or more are not.

Please keep in mind that genetic distance is not the same thing as "genealogic distance." We can calculate the probability of two men having a common male ancestor within a certain time frame based on genetic distance, but we cannot be certain. Fourth cousins George (26607 GW) and Roger (28271 RL) are at genetic distance 2 from one another, but while George and I (26605 DN) are at genetic distance 0 from one another, we know that we are not related within the last seven generations.[4] It is not only possible for a father and son to be separated by a greater genetic distance than separates some distant cousins, this has certainly happened at least twice in the 10 generations that separate George and Roger. Further, it is possible for "back mutations" and "recurrent mutations" to occur.[5] When George and I first got our identical results, I joked with him that now I had proof he was my grandpa, and I wanted him to put me in his will. In fact, it is possible that our ancestors had different haplotypes and there were mutations of this kind that coincidentally made George and I appear to be more closely related than we actually are.

Analysis of Results

First of all, all the participants are in haplogroup R1b.[6] This group is the most common group in the westernmost parts of Europe, and appears at the greatest

---

[4] Though if we can ever prove that his ancestor John of Carnshanaugh and my ancestor James of Inch Island were brothers, we will be 7th cousins.

[5] This article is already getting too long. If you want to read more about these concepts, go to http://www.familytreedna.com/ftGroupQIGuide.html.

[6] You can see a big chart of the world's haplogroups at http://www.familytreedna.com/haplotree.html.

percentage in the parts of Ireland that were never occupied by other European invaders. Further, they appear to be in a subgroup John McEwan has named "R1ba," which is most frequent in Gaelic Irish…and Dal Riada Celts.[7] Now, let's have a look at their differences from one another. Here is a table showing the genetic distance between each pair of participants, and between each participant and the modal haplotype.

ID	modal	30341 JMc	34979 JM	29038 DS	32056 TD	35063 WC	27497 EN	28271 RL	34634 EG	32942 WR	26860 RC	33681 WK	35883 BE	26605 DN	26607 GW	28942 RA
modal	37	18	16	15	5	3	3	3	3	2	1	2	2	1	1	2
30341 JMc	18	37	11	12	11	20	20	19	20	19	19	19	19	17	17	16
34979 JM	16	11	37	4	10	17	17	17	18	17	17	17	18	17	17	16
29038 DS	15	12	4	37	12	16	16	16	17	16	16	16	17	16	16	15
32056 TD	5	11	10	12	25	5	5	4	6	6	6	5	5	5	5	6
35063 WC	3	20	17	16	5	37	2	3	6	5	4	5	5	3	3	4
27497 EN	3	20	17	16	5	2	37	3	4	3	2	5	5	3	3	4
28271 RL	3	19	17	16	4	3	3	37	6	5	4	5	5	2	2	3
34634 EG	3	20	18	17	6	6	4	6	37	3	2	5	4	4	4	5
32942 WR	2	19	17	16	6	5	3	5	3	37	1	4	4	3	3	4
26860 RC	1	19	17	16	6	4	2	4	2	1	37	3	3	2	2	3
33681 WK	2	19	17	16	5	5	5	5	5	4	3	37	2	3	3	4
35883 BE	2	19	18	17	5	5	5	5	4	4	3	2	37	3	3	4
26605 DN	1	17	17	16	5	3	3	2	4	3	2	3	3	37	0	1
26607 GW	1	17	17	16	5	3	3	2	4	3	2	3	3	0	37	1
28942 RA	2	16	16	15	6	4	4	3	5	4	3	4	4	1	1	37

Related | Probably Related | Possibly Related

FTDNA's Interpreting Genetic Distance for 12 Markers (http://www.familytreedna.com/gdrules_12.html)

FTDNA's Interpreting Genetic Distance for 25 Markers (http://www.familytreedna.com/gdrules_25.html)

FTDNA's Interpreting Genetic Distance for 37 Markers (http://www.familytreedna.com/gdrules_37.html)

- Infinite allele mutation model is used
- Values on the diagonal indicate number of markers tested

This table lists the participants in a column down the left side and also in a row across the top, and it shows the modal haplotype as the first entry. The numbers on this table are not marker values; they are genetic distances.[8] Again, have a look at my entry (26605 DN), starting at the left, third row from the bottom. The number by my ID under the modal column is 1—this means that I am at genetic distance 1 from the modal haplotype. Check back on the first table, the results

[7] You can see his work at http://www.geocities.com/mcewanjc/index.htm.

[8] On the diagonal from upper left to lower right, where each participant intersects with himself, the number shown is the number of markers tested rather than the genetic distance. Obviously, any participant would be genetic distance 0 from himself.

table. Sure enough, there is only one difference between my haplotype and the modal haplotype, at DYS 576, so the genetic distance is 1. Continuing to the right in my row, the next number is 17 in the column under 30341 JMc. This shows that I am genetic distance 17 from John McEwan. By consulting this table, you can see the genetic distance between any two participants. Have a look at George (26607 GW) and Roger (28271 RL). You can see that there is a 2 in George's row, the second row from the bottom, in the column under Roger's ID, eighth data column from the left. George is genetic distance 2 from Roger. Check the results table again and you can confirm that there are differences at DYS 439 and DYS 460.

You can see that 11 of the men are at genetic distance 3 or closer to the modal haplotype and three are at 15 or further. Family Tree DNA's experts tell us that men who are at a genetic distance of 5 on the 37-marker profile are "possibly related;" those of distance 4 are "probably related;" those of distance 2 or 3 are "related;" those of distance 1 are "tightly related;" and, those of distance 0 are "very tightly related." It is exceedingly unlikely that men who are at a genetic distance of 6 or greater from one another are related in the period since surnames have been in use. As you can see, eleven participants are much more closely related to one another than to the others.[9] *[Note: For those of you reading this online, I have added an additional table at the end of the article. This is derived mathematically from the genetic distance data and shows the number of years to the most recent common male ancestor of each pair of men. A number of assumptions were made in doing this calculation, including that we wanted a 50% probability that the most recent common male ancestor lived at or more recently than the time shown on the chart.]*

We are fortunate in more ways than one to have had John McEwan join our project. He is a sheep and cattle geneticist in New Zeeland, and I'm hopeful that an article he has written about genetics and Scottish surnames will appear in this issue of the journal. He is at genetic distance 18 from the Ewing modal haplotype, a distance of at least 16 from all but three of our Ewing participants and a distance of 11 or 12 from the other three. So John is plainly not a recent relative. Of course, at some level all human beings are relatives, and as both McEwan and Ewing haplotypes suggest that both lines originated among the Dal Raida Celts, we are more closely related to him than to many. That said, there is only a 50% likelihood that our common male ancestor lived more recently than two or three thousand years ago, long before surnames came into common use in Scotland sometime around 1000 years ago.

[9] We have 37-marker data on everyone except Thomas Dale Ewing (32056 TD), who had the 25-marker study. He is at genetic distance 5 from the 25-marker modal haplotype, which means it is unlikely that he is "related" to the other 14 men, and he would probably find another couple of differences if he had the remaining 12 markers tested. He may be the sole representative so far of a third family of unrelated Ewings.

Dean Scott Ewing (29038 DS) and James Morgan Ewing (34979 JM) are at genetic distance from the modal haplotype of 15 and 16 respectively, so they are not significantly more closely related to the other Ewings than the other Ewings are to John McEwan. They are at genetic distance 12 and 11 respectively from John, still too far to be considered related. But they differ from one another by only 4. This is "probably related." I have put Dean and Jim in touch with one another, but I'm not sure whether they have worked out the connection, yet. They would appear to be members of one distinct Ewing family.

Let's have a look at the 11 men who are within genetic distance 3 from the Ewing modal haplotype. Five of them (RC, WK, BE, DN and GW) are all within a genetic distance of three or less from one another. A sensible initial hypothesis is that these five men constitute one family. But we have more information about our project participants than just their haplotypes; we have conventional genealogic information. WK and BE are 4th cousins; that's cool, both fall within this hypothetical family. But GW and RL are also 4th cousins. And RC and EN are 3rd cousins. This causes us to revise our hypothesis and include RL and EN in the hypothetical family. If we then include every man that is within genetic distance 2 or less from at least one member of the revised group, we end up including all 11 of these men in one family.

Correlation with Conventional Genealogic Data

Let's look at some of the conventional genealogic relationships a little more closely. Four of our participants are fifth great grandsons of John Ewing of Carnshanaugh. Wally Ewing (33681 WK) and Ben Ewing (35883 BE) are descended from his grandson "Swago Bill" Ewing. George Ewing (26607 GW) and Roger Ewing (28271 RL) are descended from his grandson John Ewing. So Wally and Ben are 4th cousins, George and Roger are 4th cousins, and the each pair of men is 6th cousins with the other pair. Ben and George are genetic distance 1 from the modal haplotype, Wally is 2 and Roger is 3, but that is not so interesting. What is a little more interesting is that 4th cousins Wally and Ben are genetic distance 2 from one another; and, 4th cousins George and Roger are also genetic distance 2 from one another. George is genetic distance 3 from Ben and Wally (so far, so good—sixth cousins are not as closely related as fourth cousins) but Roger is genetic distance 5 from Ben and Wally! It's plain that mutations have been happening in this family tree a fair amount more frequently than "average." Now I think that is interesting, but I could understand why a genealogist might still be asking, "Where's the beef?"

Well, here's the beef. Ben and Wally both differ from the modal haplotype at YCA IIb, where they have 22 repeats instead of 23. George and Roger both differ from the modal haplotype at DYS 576, where they have 19 repeats instead of 18. This suggests that a value of 22 at YCA IIb may be specific in Ewing men for descendents of "Swago Bill" Ewing (or perhaps his father, James). The same reasoning suggests that a value of 19 at DYS 576 may be specific for descendents of John Ewing (1714-1832) (or perhaps his father, William), but we

run into a little snag here. This is that Robert Alan Ewing (28942 RA) and I (26605 DN) also have 19 at DYS 576.[10] We are known not to be descended from John Ewing (1714-1832), and we do not have genealogic evidence of a relationship between us. The explanation for this is probably that there have been independent ("recurrent") mutations at DYS 576 in other lines. As DYS 576 is one of the rapidly mutating markers, this would not be too surprising. Data on more men in this family group will help us to understand this more clearly.

Conclusion

The Y-DNA Ewing surname project is off and running. We now have over twenty participants and have reported an analysis of the results on fifteen. All participants so far fall into a subgroup of haplogroup R1b thought to be associated with the descendents of Neolithic Irish Gaels. We have identified two and possibly three distinct, unrelated Ewing families, one of which is represented by a majority of our participants. Within the known descendents of John Ewing of Carnshanaugh, we have proposed identifying mutations for two branches of that family. As we continue to accumulate data and begin to approach our goal of having 100 project participants, we expect to be able to identify characteristic haplotypes for many branches of these Ewing families, and will be able to suggest lines on which to focus for project participants who have hit "brick walls" in their conventional genealogic research.

To Join or Get More Information

If you are ready to join the project, go to

http://www.familytreedna.com/public/ewing,

then click on "Join this group" at the top of the blue section on the left of the page. You can also see a table of our results there if you will scroll down to the bottom of the page, but results tables that are in an easier to read format are available on the Clan Ewing website. There are also links on the FamilyTreeDNA website to articles and FAQs. If you want to ask me questions, e-mail me at:

davidewing93 at gmail.com.

or call me at 505-764-8704 in the evening.

David Neal Ewing
Albuquerque, NM

Addendum

The following chart did not appear in the version of this article published in the *Journal of Clan Ewing*, and has been added here for readers of the online version because we have fewer space constraints. The numbers in the body of the chart are estimates at a 50% confidence of the time to most recent common ancestor (TMRCA) for each pair of men. Remember, we are dealing with probabilities here, and we know some of the numbers appearing here are not

[10] So does John McEwan, but since we know he is not in this family group, there is no danger of this confusing us.

correct. Although there may be a 50% likelihood that George (26607 GW) and I (26605 DN) have a common male ancestor in the last 50 years (2 generations), we know for a fact that we do not have a common male ancestor in something like 300 years. There are other places in this chart where we know that the actual TMRCA is *shorter* than the time estimate shown, too.

ID	modal	30341 JMc	34979 JM	29038 DS	32056 TD	35063 WC	27497 EN	28271 RL	34634 EG	32942 WR	26860 RC	33681 WK	35883 BE	26605 DN	26607 GW	28942 RA
modal	37	1600	1375	1250	725	250	250	250	250	175	100	175	175	100	100	175
30341 JMc	1600	37	875	950	1750	1850	1850	1725	1850	1725	1725	1725	1725	1475	1475	1375
34979 JM	1375	875	37	300	1550	1475	1475	1475	1600	1475	1475	1475	1600	1475	1475	1375
29038 DS	1250	950	300	37	1950	1375	1375	1375	1475	1375	1375	1375	1475	1375	1375	1250
32056 TD	725	1750	1550	1950	25	725	725	575	875	875	875	725	725	725	725	875
35063 WC	250	1850	1475	1375	725	37	175	250	450	375	300	375	375	250	250	300
27497 EN	250	1850	1475	1375	725	175	37	250	300	250	175	375	375	250	250	300
28271 RL	250	1725	1475	1375	575	250	250	37	450	375	300	375	375	175	175	250
34634 EG	250	1850	1600	1475	875	450	300	450	37	250	175	375	300	300	300	375
32942 WR	175	1725	1475	1375	875	375	250	375	250	37	100	300	300	250	250	300
26860 RC	100	1725	1475	1375	875	300	175	300	175	100	37	250	250	175	175	250
33681 WK	175	1725	1475	1375	725	375	375	375	375	300	250	37	175	250	250	300
35883 BE	175	1725	1600	1475	725	375	375	375	300	300	250	175	37	250	250	300
26605 DN	100	1475	1475	1375	725	250	250	175	300	250	175	250	250	37	50	100
26607 GW	100	1475	1475	1375	725	250	250	175	300	250	175	250	250	50	37	100
28942 RA	175	1375	1375	1250	875	300	300	250	375	300	250	300	300	100	100	37

0-225 Years	250-475 Years	500-725 Years	750-975 Years

- Infinite allele mutation model is used
- Average mutation rate varies: 0.0044 to 0.0054, from FTDNA derived rates
- Values on the diagonal indicate number of markers tested
- Probability is 50% that the TMRCA is no longer than indicated
- Average generaton: 25 years