Ewing Surname Y-DNA Project – Article 12

David Neal Ewing (+1 505.764.8704, *DavidEwing93 at gmail dot com*)

This is the twelfth in a series of articles about the Ewing Surname Y-DNA Project. The previous eleven articles have appeared in the last eleven issues of the *Journal of Clan Ewing*. They are also available online through links at the *Clan Ewing* web site (*www.ClanEwing.org*). Extensively cross-linked results tables, project participant lineages, group relationship diagrams and network diagrams are also available on the *Clan Ewing* web site.

Update on the Ewing Y-DNA Project

In the last Y-DNA Project article, I promised to discuss more fully the DNA results of project participants that I have referred to in previous articles as "singletons." On reflection, I realize that "singletons" was a pretty sorry choice of words, and it does not at all convey what it was I wanted to discuss in this article. What I want to discuss in this article is the results of those Ewing Surname Y-DNA Project participants who are *not* in the "large closely related group of Ewings" to which we have paid so much attention in preceding articles. As of October 12, we have results on seventy men in the Ewing surname Y-DNA project. Of these, forty-eight are in Haplogroup R1b1c7, and all but one of these are almost certainly descended from a common male ancestor, who lived near the limit of genealogic time, maybe 500 years ago or so. We do not (yet!) know who he was. The remaining twenty-two men in our project are plainly not related within a genealogic timeframe to the men in R1b1c7, and except for some sub-groups among them, they are not related to one another, either. This article is dedicated to a discussion of their results.

Grouping Ewing Y-DNA Data

Let us review for a moment, in Figure 1, how we are sorting the Ewing Y-DNA data that we have collected so far. We have results on 70 project participants, and 37-marker results on all but five of

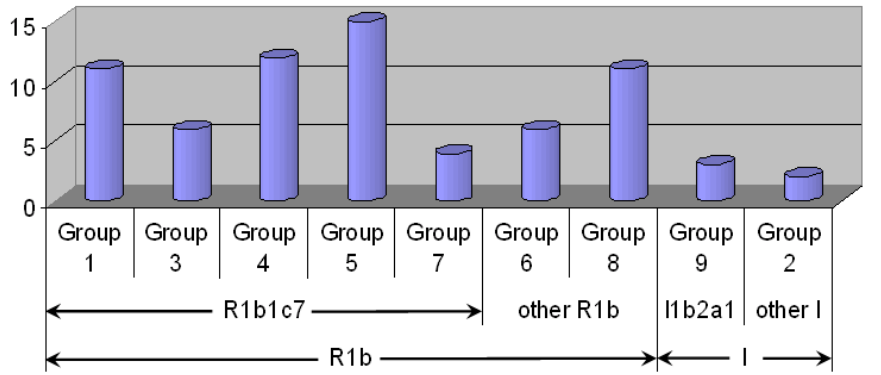


Figure 1: Ewing Y-DNA Project Groups

these. Of the 70 participants, 48 are in the R1b1c7 subclade of haplogroup R1b.[1] As you can see in Figure 1, Groups 1, 3, 4, 5 and 7 are all in R1b1c7, Groups 6 and 8 are in haplogroup R1b but not in R1b1c7, and Groups 9 and 2 are in different branches of haplogroup I (the letter "I" not the number "1").

- Group 1 includes eleven men, ten of whom[2] have DNA results close enough to the others in R1b1c7 that we believe them to be close relatives, but we do not have conventional genealogical evidence proving this.
- Group 3 includes six descendants of James Ewing of Inch.
- Group 4 includes twelve descendants of John Ewing of Carnashannagh.
- Group 5 includes five descendants of Nathaniel Ewing and his half-brothers, plus ten other men who have similar haplotypes but are not known to be related on conventional genealogic grounds.
- Group 7 includes four descendants of James Ewing, born circa 1720/25.

The seventeen participants in Groups 6 and 8 are in haplogroup R1b, but not in the R1b1c7 subclade.

- Group 6 is a cluster of six men that appear closely related to one another based on their DNA results. Three of them have conventional genealogy documenting their relationship.
- We have changed the definition of and membership in Group 8 beginning with this article. Formerly, Group 8 consisted only of Stephen Lee Ewing (SL) and Mark Edwin Ewing (ME), who are known third cousins of one another, but are not related to any of the other men in the project. Now, Group 8 includes another nine men that are in R1b but not R1b1c7, most of whom were formerly in Group 2, Singletons. There is a second pair of related men (RL2 and PT) in the new Group 8, but the others are not related to one another in a genealogic timeframe.

Finally, we have five participants that are in haplogroup I:

- Group 9 consists of three descendants of William Ewing of Rockingham County, Virginia, known relatives who have very similar haplotypes.
- Group 2 consists of two men in a completely different branch of Haplogroup I.

As you can see, the Ewing Project groups are not defined in a uniform way. Some are defined on the basis of kinship groups, others on the basis of genetic clustering, Groups 5 and 6 use a hybrid of the two methods, and some groups consist of men who don't fit into the other groups. In this article we are discussing the results of the twenty-two men in Groups 6, 8, 9 and 2, and will be trying to understand the relationships among these "unrelated" men.

---

[1] A current, detailed diagram of the subclades of haplogroup R is posted on the ISOGG web site at:
http://isogg.org/tree/ISOGG_HapgrpR07.html
along with defining SNPs and extensive references. When we say "subclade," we could as easily say "sub-haplogroup;" it means the same thing—a branch off a branch of a tree, all of the branch points of which are defined by SNPs. When we are speaking of subgroups characterized by specific patterns of STR haplotypes, we usually speak rather of "clusters."

[2] Our participant TD is in R1b1c7, but is further from the Ewing modal than the other participants (TD is genetic distance 9 from the Ewing 37-marker modal, while those in the "related group" are genetic distance 5 or less) and up to now we have not considered him part of the closely related group. For the purposes of this discussion, I am going to include TD with the other men in R1b1c7, even though he may not be related to them in a genealogic timeframe.

I am going to try to stay away from technical details in this article to the extent possible, but just to give you a "quantitative" idea of the relative degree of relatedness among these various groups, consider these gross genetic distances with respect to the Ewing modal haplotype. The men in "the large closely related group" of Ewings are within genetic distance five (GD 5) of the modal and most of them are at GD 3 or less. TD, our "borderline" case, is at GD 9. Group 6 is at GD 19-20; Group 8 ranges from GD 15 to GD 27; Group 9 is at GD 43-45; and lonesome old JD in Group 2 is at GD 54. Squint your eyes, cock your head and don't look too closely, but notice that men at GD 4-5 are "related," men at GD 20-30 are in the same haplogroup, and men at GD 40-50 are in different haplogroups.

Distinguishing Haplogroups

Actually, all men on the planet are related, because all are thought to descend from a single common male ancestor, who lived something like 60,000 years ago in Africa.[3] By comparing the Y-Chromosomes of living men, scientists have determined the relationships among several main branches of the family tree of this man's descendants and have estimated the times at which various branches originated. They have done this by comparing "SNPs," genetic markers of a different kind than we study in the Ewing project.[4] Have a look at a somewhat simplified version of the family tree in Figure 2.[5]

The twenty-one letters shown along the bottom of the tree are the haplogroups. (Don't let it bother you that these are "branches" and the "root" is at the top.) Haplogroup C and all of the haplogroups shown to the right of it share a mutation[6] that Haplogroups A and B do not share.[7] Haplogroups A and B are found today only in Africa (and in recent immigrants from Africa, of course), and all human populations outside of Africa share the mutation that distinguish Haplogroup C and those to the right of it in the diagram from Haplogroups A and B. This is part of the reason we think human beings originated in Africa, but discussion of this is beyond the scope of this article. Similarly, Haplogroup P and all of the haplogroups to the right of it share a mutation that is not found in the haplogroups to the left of P. Haplogroups R, R1a and R1b all share a mutation not shared by the others, and R1a and R1b each have their own mutation as well. Perhaps you can see that Haplogroup R1b has the mutations that set

---

[3] Actually, they have common ancestors going back a lot further than that, too, and what I am speaking about is the ancestry of the Y-chromosome rather than of the men themselves. It turns out we can estimate how long ago it has been to the "common ancestor" of even individual genes, and many of these existed long before the human race came into being.

[4] SNP stands for "Single Nucleotide Polymorphism." Discussion of the meaning and implications of this sort of marker is beyond the scope of this article, but you can read something about this in Ewing Surname Y-DNA Project: Article 7, *J. Clan Ewing*, Vol. 12, No. 3 (August 2006), and more in a separate article, *Haplogroups, Haplotypes and Clusters for the Flustered*, both of which are available on the *Clan Ewing* web site (*www.ClanEwing.org*)

[5] For a really cool, full color PDF file showing both mtDNA and Y-DNA trees, as well as the world-wide distribution of haplogroup members, go to *www.scs.uiuc.edu/~mcdonald/WorldHaplogroupsMaps.pdf*. If you can not or do not want to download a PDF file, you can see these diagrams in your browser by just leaving off the ".pdf" part.

[6] Please do not be misled by the common usage of the word "mutation." SNP and STR mutations have no effect whatsoever on the appearance or functioning of men who have or inherit them, but the fact that they are passed on from generation to generation allows us to keep track of who is related to whom.

[7] The mutations (SNPs) are not shown on this diagram, but you can tell where they would be if they were because of the shape of the diagram. If you want to see a somewhat outdated version of the Y-DNA tree that does include the mutations, have a look at *http://ycc.biosci.arizona.edu/nomenclature_system/fig1.html*.

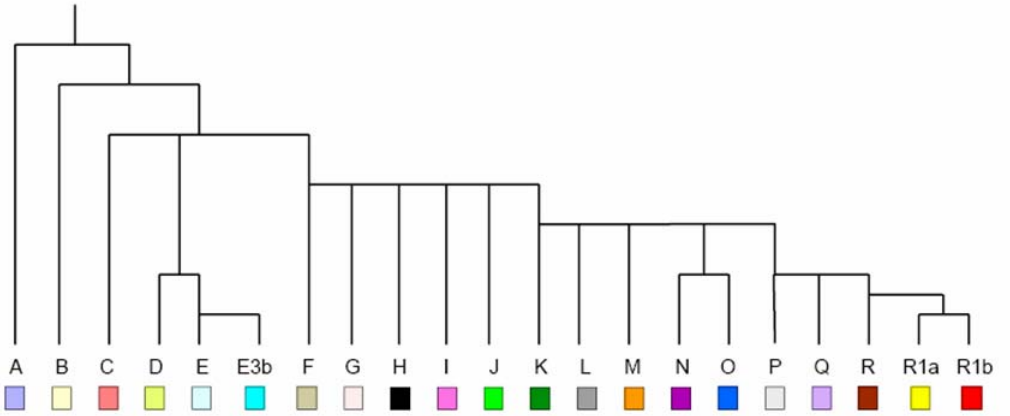


Figure 2: Simplified Tree of Y-Chromosome Haplogroups
(*www.scs.uiuc.edu/~mcdonald/WorldHaplogroupsMaps.pdf*, page 3)

off Haplogroups C, F, K, P and R in addition to the one that is unique to R1b. We would say that in some sense, these haplogroups are ancestral to Haplogroup R1b, whereas Q, for example, is an "uncle" or a "cousin" rather than an ancestor.

But I digress. I am trying to give you an idea of how distant the relationship is between the men in haplogroup I and the men in haplogroup R1b. Well, try this: a goodly fraction of American Indian men are in haplogroup Q. They are more closely related to our R1b men than the men in haplogroup I are. Amazing, isn't it? The most recent common ancestor of haplogroups I and R lived something like 45,000 years ago in the Middle East.

Men from both haplogroup I and haplogroup R are thought to have been in Europe before the last ice age 20,000 years ago, or so. Haplogroup R is thought to have arisen about 30,000 years ago in Central Asia north of the Black Sea from haplogroup P, and then to have spread across into Europe. (Haplogroup Q also arose from haplogroup P in Central Asia and then spread eastward across Siberia and eventually to the New World.) Haplogroup I seems to have arisen from haplogroup F in the Middle East about 25,000 years ago, and then representatives migrated up the Danube river valley into Central Europe. Then the ice came. People survived only in a few areas in Southern Europe south of the huge ice sheet that made most of Europe uninhabitable during the ice age. These areas are known as "refugia," places where humans and other species took refuge from the ice. We think there were three such areas, that each developed its own genetic signature, that Europe was re-populated from these refugia as the ice began to recede 15,000 years ago or so, and that this is what is responsible for the main part of the regional structure we see in the haplogroups of modern-day Europe.

Glacial Refugia

As shown in Figure 3 on the next page, the three glacial refugia were located on the Iberian peninsula (in modern-day Spain), in the Southern Balkans (in modern-day Macedonia/Serbia/Bulgaria maybe) and just north of the Black Sea (in modern-day Ukraine). There were also people in Italy through the last ice

age, but they seem not to have contributed to the repopulation of Europe. The lineages that survived in Spain were mostly in haplogroup R1b, the lineages that survived in the Balkans were mostly in haplogroup I, and the lineages that survived in the Ukraine were mostly in haplogroup R1a. You will find this idea stated as fact in many places on the web, but my experience is that anything that can be summed up this tidily is probably wrong—the "real deal" almost always turns out to be more complicated than anything we can diagram so neatly. Still, an idea like this has value, because it gives us a starting point from which to argue.

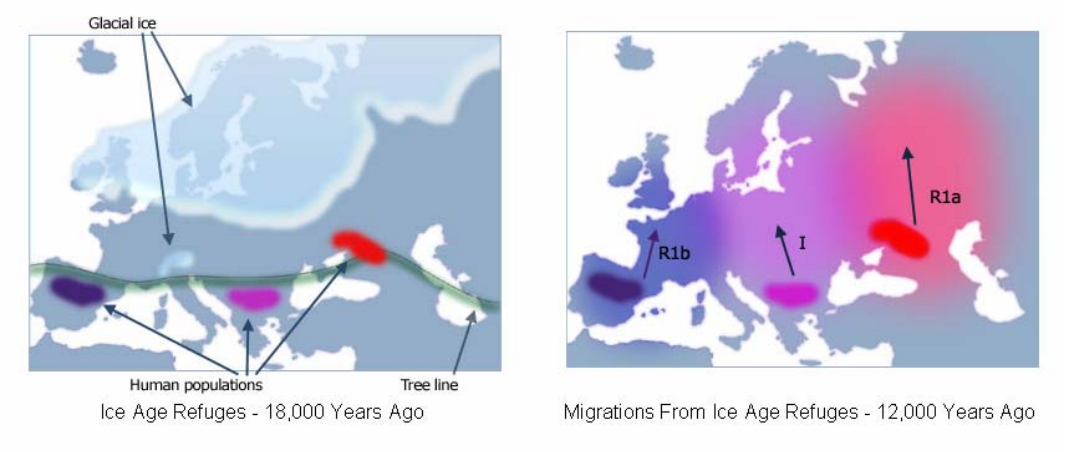


Figure 3: Migrations from Ice Age Refuges
(*www.freerepublic.com/focus/f-news/1888093/posts*)

As the ice receded, haplogroup R1b expanded from the Iberian refugium northward along the Atlantic façade;[8] haplogroup I expanded from the Balkan refugium northward through central Europe to Scandinavia; and, haplogroup R1a expanded to the north and west from the refugium north of the Black Sea into Eastern Europe and Scandinavia—interestingly, R1a is better represented in modern-day Norway than in the other Scandinavian countries.

So far, no one in the Ewing Surname Y-DNA project has been found to be in haplogroup R1a, notwithstanding that nearly 10% of Scots are in this haplogroup. Five Ewing men are in haplogroup I and all the rest are in R1b. This is neither surprising nor terribly informative: see the discussion under The Capelli Study, below.

We should also mention that based on archeological findings, the spread of farming into Europe beginning eight thousand years ago in the Neolithic seems to have followed two routes, one along the north coast of the Mediterranean and up the Atlantic façade, and the other from the Middle East up the Danube into Central Europe and points north—so just the routes that R1b and I would have taken from

[8] The R1b1c7 subclade of R1b probably emerged on the order of 4000 years ago, and most scholars think it probably emerged in northwest Ireland, though both the date and location are controversial.

the Iberian and Balkan refugia, respectively, six thousand-odd years before. Also, the most commonly cited theory of the spread of Indo-European languages into Europe a couple of thousand years later argues that they were brought by invading people, who originated in the vicinity of the Black Sea refugium (the putative R1a homeland).[9] Of course, this all creates fertile ground for discussion and dispute about which haplogroup, and especially about which specific branch of which haplogroup, came from where and when. You will find folks who are just beginning to learn about this stuff claiming that the R1b haplogroups in Britain must have come from Spain and the I haplogroups must have come from Scandinavia or Germany, but things are not that simple. Some sub-groups of Haplogroup I were probably in Britain before it became an island nine thousand-odd years ago, and there is some reason to believe they may have come up the Atlantic coast with the R1b folks. Furthermore, some of the invaders from the continent must surely have been in R1b.

Haplogroup I

Group 2

JD and our only Ewan participant are the only two project participants remaining in Group 2 after the recent reshuffling. We have only 12-marker data on Ewan, so even though he matches JD exactly at all twelve markers, we cannot speak with any great confidence about whether or not they may have had a common ancestor in a genealogic timeframe. JD knows of no Ewans in his background and we do not have lineage information on Ewan.

JD traces his Ewing lineage with some confidence to his second great-grandfather, James D. Ewing (1773-1850) who married Mary McCleary. This man appears in Fife's Chapter 26 on the descendants of Samuel Ewing (1705-1758),[10] who is the fifth son of "I believe his name was William," the progenitor of our Group 5, Part 1. Fife must have a mistake though, because JD's Y-DNA is completely different than the other descendants of this man.

I have corresponded with Ken Nordtvedt, an expert on haplogroup I, and he thinks that JD is probably in the I1b2* subclade of haplogroup I, because he has some of the characteristic STR markers of this group. To be certain, JD would have to be tested for the specific SNP that defines this group. Dr. Nordtvedt says, "This variety is found well-dispersed in continental Europe from Italy and Iberia, in France and Germany, and up through Denmark."[11] Obviously, this does not narrow things down much. JD's ancestors could be among the Anglo-Saxons who invaded Britain beginning in the fifth century, or the Danes who invaded a couple hundred years later, or even among native Britons who were there thousands of years before any of the recorded invasions.

One thing is for sure: as new Ewings join the project, it will be easy to distinguish JD's close relatives, because he has a completely distinct type of DNA from the other Ewings.

---

[9] This summary statement is laughably abrupt. Every claim in it is subject to endless controversy. Please do not quote it without a big disclaimer—I do not really believe or disbelieve a word of it; we simply do not have evidence adequate to reach definitive conclusions.

[10] Fife, Margaret Ewing. *Ewing in Early America*, Chapter 26, p. 220.

[11] I got this material from *www.northwestanalysis.net* , but this page has been taken down and the material has not been re-posted. Dr. Nordtvedt's diagrams and calculations can be seen at *http://knordtvedt.home.bresnan.net*.

Group 9

All three of the men in Group 9 are known on the basis of their conventional genealogies to be descended from William Ewing (c1696-1794) of Rockingham County, Virginia, whose lineage appears in Fife, Chapter 32. Interestingly, Family Tree DNA (FtDNA[12]) initially predicted that WM, the first member of this group, would be in haplogroup G. Haplogroup G is found in a very small percentage of modern-day British, and it is much more prevalent in Georgia—not Georgia next to Alabama, but Georgia next to Azerbaijan! We had some fun with that one, let me tell you. WM initially thought that this rather "far out" result confirmed his suspicion that his third great-grandfather had been adopted into the Ewing family. Then HN joined the project. He had also traced his lineage to William Ewing of Rockingham, who was the fifth great-grandfather of both him and WM. Their results were almost a perfect match, so it turns out that WM's third great-grandfather was not adopted after all; he was the biological grandson of William of Rockingham.

Shortly afterwards, VC joined the project. He is known on the basis of conventional genealogy to be a third cousin of HN, and their results match closely, not surprisingly. What is surprising to the uninitiated is that VC's results do not match HN's results as closely as WM's do. It is a little hard to get used to the idea that we are sometimes going to find sixth cousins who appear on the basis of their DNA to be more closely related than third cousins, but such are the vagaries of random events like mutations.

Furthermore, FtDNA predicted that HN and VC would be in haplogroup I. Something was wrong. It is simply impossible for close relatives with practically identical DNA results to be in different haplogroups, but FtDNA had predicted that WM would be in haplogroup G. What happened? The problem turns on the meaning of the word "predicted." What is to predict? They were tested, were they not? Well, they were tested for STR markers, but haplogroups are defined on the basis of SNP markers. SNP mutations happen so infrequently that for all practical purposes they are unique and permanent, so they are the gold standard for tracing lines of descent. A good number of SNPs have been identified, but most of them are very old—on the order of tens of thousands of years old—so they are not helpful in distinguishing recent branchings, and they are all but useless in genetic genealogy. For the purposes of genetic genealogy, we test STR markers, which mutate about 100,000 times more often than SNP markers. This makes them much more useful for distinguishing branchings in a short timeframe, but it introduces a considerable degree of uncertainty when trying to sort out branchings that are a few thousand years old. This is because STR markers are subject to "back mutations," where a marker can mutate away from the ancestral value and then mutate "back" to the original value in a subsequent generation, and "parallel" or "convergent mutations," where two men can have the same value at a marker by coincidence rather than because they are descended from a common ancestor.[13]

You can imagine that the man in whom an SNP mutation first occurred would not only pass that on to his descendants, he would also pass on all of his STR values. Over many subsequent generations the SNP would persist (remember: SNPs are permanent for all practical purposes), and gradually STR mutations would begin to occur and would come to distinguish various lines among his descendants. Still, each of his descendants would have most of the original STR values, and there would be a

[12] *www.FamilyTreeDNA.com*

[13] The STR markers do not all have the same rates of mutation. Back mutations and convergent mutations become even more of a problem in the more rapidly mutating markers like CDYa/b.

considerable degree of overlap in the various lines. Of course, there will be indeterminate cases in which not enough markers match for us to be sure, and there will be cases where we are simply mistaken, especially if we are relying on only a very few markers. We could go on about the technical details at considerable length here, but the bottom line is that we can use certain characteristic patterns of STR marker results to "predict" what SNPs we would find in a man if we tested him, but we might make a mistake.

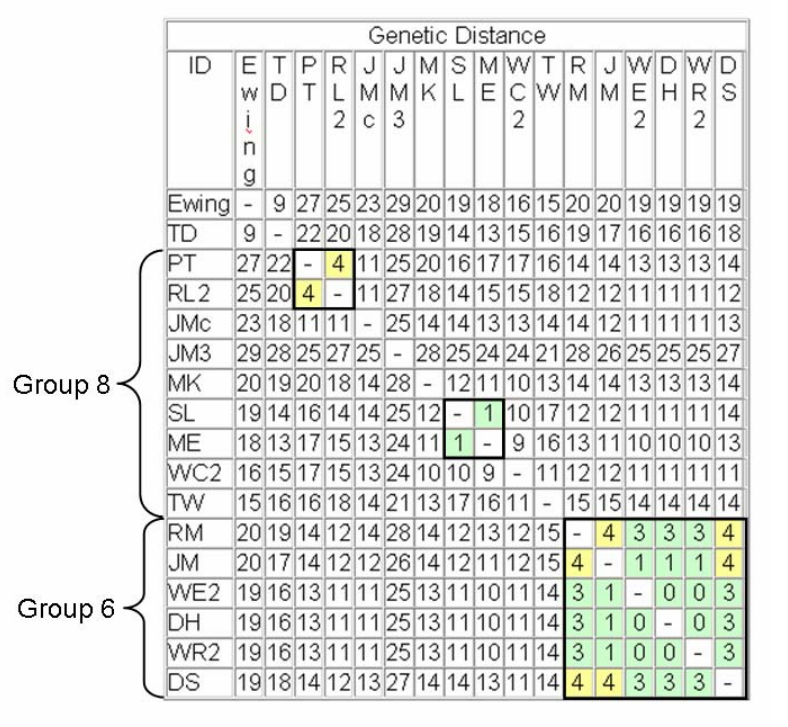
I called up FtDNA and brought their attention to the aberrant results of their haplogroup predictions for WM and his known relatives, HN and VC. They tested all three for the SNP that defines haplogroup I and found that it was present in all three. WM was not in haplogroup G after all. The poor guy keeps getting his hopes up that he is a completely odd-ball Ewing, and results keep coming back that suggest he is a regular Ewing. I have become well enough acquainted with him that I can certify with some confidence that he is definitely an odd-ball, regardless of what the results may say. Of course, the argument could be made that being an odd-ball is a fairly wide-spread trait amongst Ewings, but maybe we had best not explore that too deeply.

In any case, the three men in Group 9 are in haplogroup I, but they are not at all closely related to JD in Group 2.[14] Dr. Nordtvedt tells me that he thinks they are most likely in what he calls the "I1b2a1-Isles-Eng" sub-group of haplogroup I. He says, "I1b2a-Isles is found almost exclusively in the British Isles, and heavily from Scotland at that…I1b2a1 is a candidate haplogroup which may have arrived in the British Isles in pre-Roman times, and perhaps directly from more southwesterly Europe instead of Anglo-Saxon or Scandinavian sources." As with JD, any new Ewing Surname Y-DNA Project participants that are closely related to the men in Group 9 will be easy to identify, because their DNA is quite different from the DNA of everyone else in the project.

[14] JD is genetic distance 36-39 from the men in Group 9.

Haplogroup R1b

Have a look at the genetic distance table in Figure 4 to get help orienting to the relationships among the Ewing men in haplogroup R1b outside of R1b1c7[15] and to see why we say they are "unrelated."



	Genetic Distance																	
ID	Ewing	TD	PT	RL2	JMc	JM3	MK	SL	ME	WC2	TW	RM	JM	WE2	DH	WR2	DS
Ewing	-	9	27	25	23	29	20	19	18	16	15	20	20	19	19	19	19
TD	9	-	22	20	18	28	19	14	13	15	16	19	17	16	16	16	18
PT	27	22	-	4	11	25	20	16	17	17	16	14	14	13	13	13	14
RL2	25	20	4	-	11	27	18	14	15	15	18	12	12	11	11	11	12
JMc	23	18	11	11	-	25	14	14	13	13	14	14	12	11	11	11	13
JM3	29	28	25	27	25	-	28	25	24	24	21	28	26	25	25	25	27
MK	20	19	20	18	14	28	-	12	11	10	13	14	14	13	13	13	14
SL	19	14	16	14	14	25	12	-	1	10	17	12	12	11	11	11	14
ME	18	13	17	15	13	24	11	1	-	9	16	13	11	10	10	10	13
WC2	16	15	17	15	13	24	10	10	9	-	11	12	12	11	11	11	11
TW	15	16	16	18	14	21	13	17	16	11	-	15	15	14	14	14	14
RM	20	19	14	12	14	28	14	12	13	12	15	-	4	3	3	3	4
JM	20	17	14	12	12	26	14	12	11	12	15	4	-	1	1	1	4
WE2	19	16	13	11	11	25	13	11	10	11	14	3	1	-	0	0	3
DH	19	16	13	11	11	25	13	11	10	11	14	3	1	0	-	0	3
WR2	19	16	13	11	11	25	13	11	10	11	14	3	1	0	0	-	3
DS	19	18	14	12	13	27	14	14	13	11	14	4	4	3	3	3	-

Figure 4: Genetic Distances

You can see the genetic distance between any two individuals in this chart. "Ewing" is the modal haplotype for "the large group of closely related Ewings." You can see the number 9 where TD and Ewing intersect on the chart—TD is genetic distance 9 from the Ewing modal. He is the only man in the R1b1c7 subdivision of haplogroup R1b shown in this chart. (The genetic distances from the Ewing modal to the other R1b1c7 Ewings, who are not shown in this chart, range from 0 to 5.) Notice that the genetic distances from the Ewing modal to the other men in this chart range between 15 for TW to 29 for JM3. We have been saying that anything over genetic distance 5 counts as "unrelated," so they are all plainly unrelated to "the large group of closely related Ewings." But now have a look at the genetic distances between them. I have drawn dark boxes around clusters of men who are (or appear to be) related to one another. The large box shows the genetic distances between the men in Group 6—these

[15] We have included in this analysis only those men who have 37-marker results.

range from 0 to 4, so we can consider them related to one another. The two smaller boxes show the genetic distances between the men in each of the two pairs of known relatives within Group 8: sixth cousins PT and RL2 are at genetic distance 4 from one another, and third cousins SL and ME are at genetic distance 1 from one another. Otherwise, everybody in the chart is at least genetic distance 9 from all the others.

Group 6

Group 6 consists of six men so far. Three of them have identical 37-marker results and a fourth is different from them at only one marker. The other two differ at three or four markers from one another and the others. This cluster is tight enough that we can consider these men to be related to one another in a genealogic timeframe. Indeed, three of them (JM, WE2 and WR2) have conventional genealogies tracing their lineages to a common ancestor: William Ewing (c1730-1774) who married Eleanor Thompson (c1738-1774).[16] (Both died in Conemaugh Township, Indiana County, Pennsylvania.) WR2 and WE2 have identical test results and JM differs from them at only one marker. A fourth man in Group 6 has identical test results with WR2 and WE2, but he has been able to trace his conventional genealogy back only as far as his second great-grandfather, Thomas Ewing, born between 1808 and 1811 in New Castle, Mercer County, Pennsylvania, married Rebecca Burke (c1815-c1898), and has not been able to establish a connection with the William Ewing born circa 1730 line. We consider it to be a good bet on the basis of his DNA results that he is connected to this line, too, though.

The other two men in Group 6 (DS and RM) are each genetic distance three from the three men who have identical results. They share one marker, but each has two that the other does not, so they are genetic distance four from one another. The fact that they share a marker that the other men in this group do not have raises the possibility that they are more closely related to one another than they are to the others, but the marker they share[17] is a relatively quickly mutating one and it would not be impossible for this to have occurred by coincidence. It is also not impossible that they are also descended from William Ewing born circa 1730, but we think this is unlikely. DS traces his lineage back to John Ewing born 1759 in Pennsylvania, m. Elizabeth Gardner, died 1817 in Montgomery County, Ohio. He is the right age to be a son of William Ewing b. c1730, but William already has a son John born circa 1765 (the common ancestor of WR2 and JM), so this is unlikely. RM is a Canadian Ewing, whose immigrant ancestor was born in Ireland, probably Donegal, c1768 and came to Nova Scotia in 1833. One of his grandsons, Orlando Chester Ewing (1855-1926) moved to Massachusetts and has several living American Ewing descendants, but none has participated in the project. Unaccountably, the lineage for RM we have had posted on the web site shows William Ewing (circa 1730-1774) as a sibling of his third great-grandfather, but this information is not in the material RM sent us, so it is mistaken. I have an idea that it resulted from some fooling around and wishful thinking that I was doing in the database I maintain on project participants. We will fix this as soon as possible.

---

[16] Interestingly, this family seems not to appear in Fife, but it does appear in Joseph Lyons Ewing's *Sketches of the Families of Thomas Ewing and Mary Maskell, William Ewing and Eleanor Thompson, James Ewing and Eleanor Rhea and their Descendants, with Historical Data and Reminiscences*, The Stratfor Commercial Job Printery, Stratford, New Jersey, 1910, which is available as a facsimile reprint from *www.HigginsonBooks.com*.

[17] DYS 458 = 17

Group 8

We have redefined Group 8 beginning with this article. Formerly, it consisted of only two men (SL and ME), who are third cousins, both descended from Charles Alonzo Ewing (1836-1877) who married Mary Ellen Funkhouser (1841-1907). Both were born in Illinois and died in Wayne County, Illinois. Now, we have added a number of unrelated men to Group 8, because we have changed the membership criterion for this group to "R1b but not R1b1c7 and not Group 6." Group 8 has become a catch-all group. Undoubtedly, as time goes along and we identify larger kindreds within this group, we will break them out into groups of their own.

SL and ME

Steven Colson is a DNA researcher who has been working on a closely related cluster of haplotypes that he thinks is specific to Strathclyde. He contacted us when he discovered that SL and ME are close matches for this cluster, and we have persuaded him to write an article about his research, which appears in this issue of the *Journal*. You may recall that the Y-DNA article in the last issue of the *Journal* contained a fair amount of speculation that the Ewings originated among the Brythonic Celts of Strathclyde—well, here is some DNA evidence that at least two of them did! We are still scratching our heads trying to figure out why so many of us have DNA that is more closely associated with NW Ireland.

PT and RL2

Two other men in Group 8 are also related to one another. This is actually one of the most interesting stories we have found in the DNA project, and it illustrates the power of the DNA project to illuminate old genealogical puzzles. PT and RL2 are sixth cousins, who have traced their lineages to their common fifth great-grandfather, John Ewing (1695-1751). He is the subject of Fife, Chapter 31; Fife believed him to be the eldest son of "I believe his name was William," the patriarch of our Group 5, Part 1. Now, PT and RL2 have DNA results that match closely enough to support the notion that they are sixth cousins, but their results are nothing like the other men in Group 5, so they cannot have had a recent common ancestor with the men in Group 5. There are many possible explanations for this result, but an off-hand statement in Fife provides a clue that suggests a very plausible explanation. Fife says,

> John Ewing 'age about 55 in 1745' (Q.A. 2:299) Deposition taken in Queen Annes Co., MD-1745. He died there in 1751 without a will. His children's names and ages are in the settlement of his estate. I believe his age is in error and he was half-brother to Nathaniel. [See Chapter XXXI]. He, too, lived in Sadsbury twp. Lancaster Co., PA 1738-40.[18]

The man I have ponderously referred to as "I believe his name was William" married first a woman who may have been named Elizabeth and had with her one son, Nathaniel born 1693. He married second a woman who may have been named Ann. Fife believed that they had five sons. She thought the deposition mentioned above must be in error, because if John was "about 55 in 1745," he would have been born in about 1690—before Nathaniel. She supposed that since John was the son of William's second wife, he must have been born after the son of his first wife, so she supplied 1695 as his date of birth. The next son, also William, was born about 1700, then Joshua circa 1704, then Samuel circa 1705, then James 1712, then George circa 1715. But we have very little hard information about "I believe his name was William" and his wives; we don't even know their names for certain. I think we

[18] Fife, Margaret Ewing. *Ewing in Early America*, Chapter 24, p. 188.

should consider that the deposition in 1745 may have been correct, and if John was "about 55" then, he must have been "about 3" when Nathaniel was born and "about 5" when his mother had her first child with Nathaniel's father. This all makes perfect sense if Nathaniel's father married John's mother when he was a small child, adopted him and raised him as his own. This would also explain why the DNA of PT and RL2 does not match the rest of the men in their family.

The Seven Others

The seven other men in Group 8 are not related to one another or to anyone else in the project. About all that we have said about them so far is that they are all in haplogroup R1b (and not in R1b1c7 or Group 6), but that does not narrow things down a bit. Haplogroup R1b is far and away the most common haplogroup in Europe, and even more so along the Atlantic coast, reaching 90% among the Basques and in parts of Ireland. We said above that the men in Group 8 are at genetic distances ranging from 15 to 29 from the Ewing modal haplotype. But they are also at genetic distances ranging from 9 to 28 from one another. Is there a way we can talk about how long ago their ancestral lines diverged and where they came from?

Subdividing R1b

As we discussed above, SNPs are the gold standard for identifying deep branches in the family tree. This is a rapidly evolving area of study and new SNPs are continually being discovered. As of this writing, nineteen different SNP-defined branches of R1b have been identified. Only four men in the Ewing project have had detailed SNP testing. Three of us in "the large closely related group" (one each in Groups 1, 3 and 5) have had this testing; all were found to be "M222+," which is the definition of subclade R1b1c7. It would be a waste of money for other men in the closely related group to have this testing, because the result is a foregone conclusion. PT is the fourth man to have had detailed SNP testing. He did have the SNPs defining R1b1c, but did not have M222 or any of the other SNPs defining sub-branches of R1b1c. Now, RL2 is PT's sixth cousin, so he would certainly have the same results as PT, but we don't know about the other men in Group 8 or Group 6. We could start a campaign to get the others SNP tested, but this would cost them $79 apiece, and once we had the results, we still could not get too far, because we do not have enough men that have been tested this rigorously for us to use as a comparison group.

A fair amount of work has been done using an alternative strategy, based on looking at "STR clusters." Remember that we can speak with some confidence about the branch structure over maybe ten or fifteen generations using STRs, but if we try to go back the several thousand years since the various sub-branches of R1b arose, we are apt to make errors, especially if we use just a few markers.

The Capelli Study

In 2003, Cristian Capelli and fourteen collaborators published an interesting study[19] that was intended to investigate the extent to which the "indigenous" peoples of Britain were replaced by populations of Anglo-Saxon, Danish and Norwegian Viking invaders. This group collected Y-DNA from 1772 British men, who were selected systematically from twenty-five small towns that were more or less evenly

[19] Capelli, Cristian et al. A Y Chromosome Census of the British Isles, *Current Biology*, Vol. 13 (May 27, 2003), pp. 979-984. A PDF of this paper is available for free download at *www.ucl.ac.uk/tcga/tcgapdf/capelli-CB-03.pdf*.

distributed across Britain. Only men who were living in the same town their paternal grandfathers had been born in were used as subjects. The Y-DNA was tested for eleven SNP and six STR markers.

The specimens collected were categorized into fourteen different "types." The fourteen types consisted of eleven SNP-defined haplogroups (including some "sub-haplogroups" if you insist). Only three of these eleven[20] amount to more than 5% of any of the British or comparison populations. What do you suppose the three best represented haplogroups were? Of course, they were R1b,[21] I[22] and R1a. Capelli divided each of these three into two different "types" for the purposes of his analysis by breaking a haplotype cluster out from each. For Haplogroup R1b, he broke out a haplotype cluster he calls "AMH+1," which means "Atlantic Modal Haplotype[23] plus its one-step neighbors," or "everyone within genetic distance one of the Atlantic Modal 6-marker STR haplotype."

Percentages of each of the fourteen different types present were tabulated for each of the towns, and were compared with percentages of these types that had been found in three comparison populations. One of the comparison populations was taken to represent the population in Britain before the Roman conquest, and consisted of men from Castlerae, Ireland, (a town in Central Ireland thought never to have suffered Anglo-Saxon or Viking occupation) and Basques. These two groups did not have significantly different percentages of the types of DNA identified, so were pooled. A second comparison group thought to represent Norwegian Vikings was composed of the pooled data from two towns in western Norway that were not significantly different from one another. The investigators had hoped to be able to distinguish Anglo-Saxons from Danes, so they compared data from a general Danish collection with some data from the Schleswig-Holstein region of northern Germany that is thought to be the homeland of the Anglo-Saxons, but these two populations were not significantly different from one another, and the third comparison group ("German/Danish") was composed of the pooled data from these two regions.

The vast majority of genetic variation in the British samples was within the individual towns and only 3.65% of the variation could be apportioned to differences between the towns. This is another way of saying that there was no striking difference in the mixtures of types found among the locations sampled in Britain, but if you look really closely with powerful statistical techniques, you can see that some locations have a little more of this or that type. Furthermore, some patterns can be seen in the differences. The Capelli group concluded that there was overall less genetic input from invading

---

[20] The eight haplogroups that were not found at greater than 5% in any of the populations sampled included E3b, J, J2 and N3. Their distribution is of some interest, but none of the Ewing men fall into any of these haplogroups, so I have decided to omit any discussion of them here.

[21] Capelli's data table does not show Haplogroup R1b as such, but rather "R1xR1a1," which means "everyone in R1 *except* those in R1a1." This would amount to pretty much the same thing. The reason he uses this ponderous locution is that M173 (which defines haplogroup R1) and M17 (which defines Haplogroup R1a1) were tested, but M343 (which defines R1b) was not tested.

[22] I'm glossing over something here. Actually Capelli found very few men in Haplogroup I1b2 but a good number of them in "IxI1b2," which means "I but *not* I1b2." Though Haplogroup I1b2 was found in only a small number of British men, those who had it were all in areas thought not to have received significant continental input and it was not found at all in the Norwegian or German/Danish comparison groups. The short story (based on other sources) is that I1b2 is a subclade of Haplogroup I that has a modern-day distribution and probably history more like Haplogroup R1b than the rest of Haplogroup I.

[23] The Atlantic Modal Haplotype is an STR cluster within Haplogroup R1b that includes the majority of haplotypes on the Atlantic fringe of Europe, and is found at its highest percentages in the Iberian peninsula and Ireland.

populations into Britain than has been thought by many, but that there is evidence that traditionally "Celtic" areas of Britain had the least continental input, the Central-Eastern part of England had the most, and Orkney and Shetland had the most Norwegian input.

So where do the Ewings fit into this data? There are no R1a participants in the Ewing project, so far. No one in the "large closely related group" of Ewings is in the Capelli AMH+1 group,[24] but fifteen of the seventeen Ewing men in R1b outside of R1b1c7 do fall into the AMH+1 group and five of these match the 6-marker AMH exactly: PT, RL2, JMc, WC2 and JM3. Five of the six members of Ewing Group 6 (all except RM) match one another exactly and are genetic distance one from the AMH; SL, MK, ME and Ewen match one another exactly and are genetic distance one from the AMH at a different marker; and, TW is genetic distance one from the AMH at yet another marker. RM is genetic distance two from the AMH and Js is three from the AMH, so they are not in the AMH+1 cluster, but they are in R1b. The Ewing project has two clusters of participants in Haplogroup I: the three-man kindred in Group 9 and two men not known to be related to one another in Group 2, JD and Ewan. These two clusters are at considerable genetic distance from one another; neither is in the I1b2 sub-haplogroup and neither is in the STR cluster that Capelli broke out of the balance of Haplogroup I.[25]

And what does this tell us? Well, precious little, in my opinion. Those Ewings in R1b that are not in R1b1c7 are almost all in the AMH+1 STR cluster of R1b, which is the most commonly found cluster not only in Britain, but in all of Western Europe. Those Ewings in Haplogroup I are in the part of it that Capelli does not differentiate from Haplogroup I at large. Can we get any more specific than this?

Oppenheimer's Data

In his book *Origins of the British,*[26] Stephen Oppenheimer used the Capelli and other data he found in the public domain to a total of 3084 Y-DNA samples and came up with his own set of clusters. Twenty-one of them subdivide R1b. This tree has the virtue of making some very specific assertions, not only about the relationships among clusters, but also about the timing of the emergence of these clusters. Sadly, it does not have the virtue of being nearly as precise as it appears or, for that matter, correct. Before we criticize, let us have a look at Figure 5 on the next page to understand what the tree says.

---

[24] The 6-marker Atlantic Modal Haplotype is 12 13 13 14 24 11 in Capelli's order, which is DYS 388, 393, 392, 19, 390 and 391. The Ewing modal is genetic distance three from the AMH, and the Ewing Group 5 modal is genetic distance four from the AMH.

[25] I believe the 6-marker I1b2 modal is 13 13 11 16 23 10 for the six markers in Capelli's order. Note that the terminology has changed since the Capelli article and this haplogroup is now called I1b1b—still, it is the haplogroup defined by the M26 SNP. Though our haplogroup I men are probably in branches of I1b2 in the "new" terminology, they are not in Capelli's I1b2. The modal for the 2.47+1 STR cluster that Capelli broke out of the part of I not in I1b2 (IxI1b2) is 14 13 11 14 22 10.

[26] Oppenheimer, Stephen. *The Origins of the British, A Genetic Detective Story: The Surprising Roots of the English, Irish, Scottish, and Welsh*, Carroll & Graf, New York, 2006.

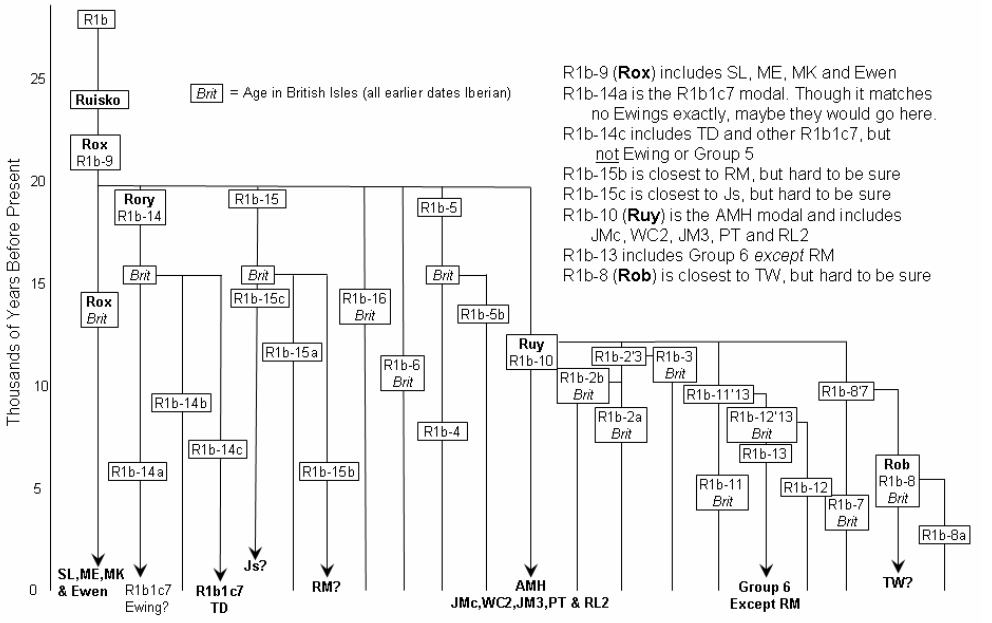


Figure 5: Ewings in Oppenheimer's R1b Cluster Tree

I have copied Oppenheimer's tree, but have added arrows and designations across the bottom of the figure. The scale on the left side of the diagram shows the number of years before present and the blocks with cluster names in the tree appear across from the number of years ago Oppenheimer has calculated that they originated. For example, R1b-14a appears in the diagram across from about 6,000 years ago, and R1b-10 (Ruy) appears across from 12,000 years ago, or so. Oppenheimer has given some of the clusters male names beginning with "R," such as Rusko, Rox, Rory, Ruy and Rob, thinking this might be more interesting and easier to follow for some folks than using names like "R1b-10,"[27] but has about driven me nuts. When the British variety of a cluster emerged at a different time from the parent cluster, Oppenheimer has made the note, *Brit*. The dating of the time of origin of these clusters is not as precise as it appears because the confidence intervals are huge. For example, Oppenheimer estimates the age of the R1b-14a cluster at 5,449 years, but the standard deviation is 2,090 years. Round numbers, this means that there is a 70% chance that the actual age of the cluster is between 3,359 and 7,539 years.

Maddeningly, Oppenheimer has not published the modal haplotypes for his clusters, perhaps because he has a deal with a commercial DNA testing company to test people and send them a certificate telling them what cluster they belong to. If you already have the test results, they will "interpret" this for you and send you the certificate for seventy bucks. This has not set too well with the genetic genealogy

[27] Incidentally, both these names and the letter/number designations for the clusters are used only by Oppenheimer. Sykes also uses names, but different ones. Other workers use different naming systems for the clusters they have identified. Keeping all of this straight can drive you crazy.

community, let me tell you. Folks have been collecting Oppenheimer test results and correlating these with known haplotypes and also with SNP-defined subclades unknown to Oppenheimer. Dennis Wright has kindly sent me his tabulation of a number of these. This has allowed me to place some of our Ewings on the Oppenheimer tree with full confidence, but has left me entirely in the dark about where to put others. Notice that I have put R1b1c7 below the R1b-14a cluster. I did this because Oppenheimer has assigned this cluster to a man, who has the same 6-marker haplotype (which also happens to be the modal haplotype for R1b1c7) as forty men who are SNP-proven R1b1c7. No Ewing project participant has this exact haplotype.[28] Oppenheimer also certified a man as belonging to cluster R1b-14c who has the same 6-marker haplotype[29] as seven other men who are SNP-proven R1b1c7. The only Ewing man who has this exact haplotype is TD. R1b-14b is almost certainly also in R1b1c7, but we don't have confirmation of that. Although we know the "large group of closely related Ewings" to be in R1b1c7, we do not know for sure where Oppenheimer would put them, and although I have an idea that he would put most of them in R1b-14a, I have no idea where he would put the men in Group 5. This is because the Ewing modal is genetic distance one from a 6-marker haplotype Oppenheimer certified as R1b-14a, but the Group 5 modal is genetic distance two from this haplotype and is no closer than genetic distance two to any of the other Oppenheimer clusters.[30]

Oppenheimer's cluster R1b-10 (Ruy) exactly matches the 6-marker Atlantic Modal Haplotype, and also five of the Ewing men, but of these only PT and RL2, who are known to be related to one another on the basis of their conventional genealogies, have 37-marker haplotypes close enough to consider them related in a genealogic time frame. Oppenheimer's cluster R1b-9 (Rox) includes Ewing project participants SL and ME,[31] who are known to be third cousins, but also MK,[32] who is genetic distance 12 and 11 respectively from them on the 37-marker panel, so cannot be related in anything like a genealogic time frame. Oppenheimer's cluster R1b-13 matches all of the men in Ewing Group 6, except for RM. I am just guessing on the placement of RM, Js and TW; none of them exactly matches any of the Oppenheimer clusters we have figured out, but they seem closest to those that I put them under.

Now, I said above that this tree is not "correct." Why do I say that? Remember that these clusters are based on 6-marker STR haplotypes. As Oppenheimer himself says,

> STR types are less reliable than [SNPs] for building trees because they are made up of a combination of rapidly mutating genetic sites which can mutate forward and backwards, thus introducing ambiguity. None the less, they are used for this purpose, and their rapid mutation has the advantage of facilitating analysis of shorter time periods.[33]

---

[28] The Ewing 6-marker modal differs from R1b1c7 only at DYS 19 = 15 (where the modal is 14), and the Ewing Group 5 modal differs from the R1b1c7 both at DYS 19 and at DYS 391 = 10 (where the modal is 11).

[29] This differs from the R1b1c7 modal only at DYS 390 = 24 (where the modal has 25).

[30] DYS 391 is one of the six markers that Oppenheimer considers. You will recall that the criterion for Ewing Group 5 membership is matching the Ewing modal closely, but having DYS 391 = 10 instead of the Ewing modal DYS 391 = 11. The other marker that both the larger Ewing modal and the Group 5 modal share but Oppenheimer's cluster R1b-14a does not share is DYS 19 = 15, where Oppenheimer (and the R1b1c7 modal) have DYS 19 = 14.

[31] See Steven Colson's article, page 55 in this issue of the *Journal*, on an interesting sub-cluster to which they belong.

[32] And also Ewen, but we have only 12-marker data on him.

[33] Oppenheimer, Stephen. *The Origins of the British, A Genetic Detective Story: The Surprising Roots …,* p. 436.

There are a good eighteen or twenty SNPs now used to define subclades of R1b that Oppenheimer did not have available when he did his analysis, including the M222 SNP that defines R1b1c7. A tree accurately reflecting genetic kinship will have these SNP-defined subclades on its terminal branches, so a tree based on STR clusters will be "correct" to the extent that it matches the SNP tree. Oppenheimer's tree does not stand up too well to this kind of scrutiny. We are forced to consider only those haplotypes for which we have been able to figure out the Oppenheimer cluster, and for which we also have at least one man with that haplotype who has been SNP tested. We have this information for sixteen of the twenty-one Oppenheimer R1b clusters: 2b, 4, 8, 8a, 9, 10, 11, 12, 13, 14a, 14b, 14c, 15a, 15b, 15c and 16 (I have left off the "R1b-" prefix in each case). In eight of these there are men who have been SNP tested and found not to have any of the R1b sub-clade SNP markers; that is, they are R1b1c*. Four, widely separated Oppenheimer clusters (9, 10, 11 and 8) have haplotypes that men who are SNP-proven R1b1c6 also have. SNP-proven R1b1c7 haplotypes are best represented in Oppenheimer clusters 14a and 14c, but 14a and 14b also have one man each in 9*, and R1b1c7 also appears in clusters 15a and 11. Perhaps not surprisingly, Oppenheimer cluster R1b-10, which includes the Atlantic Modal Haplotype, has men in at least six different SNP-defined subclades: c*, 6, 9, 9*, 9a and 10.

How could this have gone so wrong? The problem is that Oppenheimer used so few STR markers to define his clusters that the resolution is abysmal. Folks in the genetic genealogy community have been calling these 6-marker haplotypes "bikini haplotypes," presumably because the cover so little.

Sykes' Data

In his book, *Saxons, Vikings, and Celts*,[34] Bryan Sykes explains in a simplified and interesting way the implications of his analysis of some data from his Oxford Genetic Atlas Project.[35] His project included mtDNA data, but we are considering only the Y-DNA data, here. He collected DNA from English and Scottish men, which he analyzed for up to ten STR markers, and he borrowed a little Irish data.[36] He talks about "50,000 DNA sequences to work with," but I had a look at the data posted on his web site,[37] and as near as I can tell he derived his clusters and bases his arguments on a total of 2,322 Y-DNA haplotypes. Of these, only about 1,500 haplotypes have full 10-marker data; most of the rest of the

---

[34] Sykes, Bryan. *Saxons, Vikings, and Celts: The Genetic roots of Britain and Ireland*, W.W. Norton & Co., New York, 2006. This book was originally published in Britain under the title *Blood of the Isles: Exploring the Genetic Roots of Our Tribal History*, Bantam Press, 2006. It is intended for a general audience and does not include enough scientific detail to allow checking Sykes' facts or methods. Kevin Campbell (see footnote 42) has written an excellent paper analyzing Sykes' data, and fills in a lot of the detail missing from Sykes' book.

[35] Ibid., p. 119. For discussion of his data, see *www.bloodoftheisles.net/ogap.html*.

[36] The markers he considers are DYS 19, 390, 391, 392, 393, 389i, 389ii, 388, 425, and 426. Seven-marker data was on the first seven of these. Capelli's markers consisted of the first five and the eighth of these. The order of the makers in Sykes' data tables is not the same as Capelli order and not the same as FtDNA order, but Campbell (see below) puts them in FtDNA order for his discussion, and also seems to have substituted DYS 439 for 425—perhaps this is just a typo. Since DYS 425 is not a part of the FtDNA 37-marker panel (it is in the 67-marker upgrade), most Ewing men have not been tested for this, so I just used their DYS 439 values when comparing them with Campbell's OGAP chart. This may have introduced a confound, because while all but one of the OGAPs have twelve repeats at DYS 439, the Ewing modal and three of the men in Ewing Group 8 have thirteen repeats at this marker. Virtually all R1b men and all eleven of the Ewing men who have been tested have twelve repeats at DYS 425. Furthermore, the 389i/ii reporting convention Sykes uses is different from FtDNA. It is enough to drive you crazy trying to compare data across these different sets.

[37] For a PDF of his actual Y-DNA data, see *www.bloodoftheisles.net/OGAP_yDNA.pdf*.

haplotypes have 7-marker data. As it happens, the "additional" markers were not all that informative, anyway.[38] He grouped his data by regions, including eight regions in Scotland: Northern Isles, Hebrides, Highland, Grampian, Argyll, Strathclyde,[39] Tayside & Fife, and Borders. England is divided into seven regions. His map of regional borders[40] shows four regions in Ireland and three in Wales, but there is very little Irish data (only twenty-two men) and the data table on his web site does not distinguish regions within Ireland or Wales. He put haplotypes into regions based on the reports of DNA donors about where their paternal grandfathers were born.

Sykes' most general conclusions are based on the distributions of haplogroups and are very much the same as Oppenheimer's: mostly, modern British Y-DNA appears to be descended from ancestors who were natives of Britain before the historically attested invasions of the Romans, Anglo-Saxons, Norwegian Vikings, Danes or Normans. The exceptions to this are that there is are a significant number (but still a minority) of descendants of Norwegian Vikings in the Northern Isles, and of continental types (remember that with this marker set, it is impossible to distinguish Anglo-Saxons and Danes) in East Anglia (20%) and other parts of what was once Danelaw, but interestingly much less so in the south of England (10%), the supposedly traditional stomping grounds of the Saxons and Jutes. To me, one of Sykes' most interesting conclusions is that there is little genetic evidence that the "Celts" of Britain were particularly closely related to the Celts that originated in Central Europe,[41] but there is plenty of evidence for affinity with peoples of the Iberian Peninsula.

Sykes makes some claims that depend on his having performed STR cluster analysis of his data, but like Oppenheimer, he does not publish the definitions of his clusters. For this, we need to rely on the impressive analysis of Sykes' data by Kevin Campbell.[42] This is really a terrific paper; I encourage everyone with more than a passing interest in these matters to read and study it. Campbell did the exercise of calculating modal haplotypes for each of the regions in Sykes data, but this was not informative because the modal haplotype of each region matched the Atlantic Modal Haplotype, the most commonly found haplotype in all of Western Europe.

Campbell was most interested in learning about the distribution of R1b STR clusters, so he made a list of all the R1b haplotypes in Sykes' data and tallied how many of each there were. Once he had done this, he named the most commonly found haplotype in the R1b data "OGAP 1,"[43] the second most common haplotype "OGAP 2," and so on. There were a total of fifty unique R1b haplotypes in Sykes' data. The most common single haplotype (OGAP 1) was the Atlantic Modal Haplotype, not surprisingly, which accounted for a little over 16% of all the R1b haplotypes. Together, the twenty most common

---

[38] In the R1b subset of his data, 97% of these three markers all had the same value (twelve repeats). Campbell, Kevin D. Geographic Patterns of Haplogroup R1b in the British Isles. *J. Genetic Genealogy*, Vol. 3, No. 1 (Spring 2007), p. 3.

[39] The data from Strathclyde includes 120 haplotypes. The best represented cluster in Strathclyde is OGAP 8 (which corresponds with R1b1c7) and the next best is OGAP 4, which is a distinctively Scottish cluster and may have originated among the Picts.

[40] Sykes, Bryan. *Saxons, Vikings, and Celts: The Genetic roots of Britain and Ireland*, W.W. Norton & Co., New York, 2006, p. xvi.

[41] Ibid., p. 281.

[42] Campbell, Kevin D. Geographic Patterns of Haplogroup R1b in the British Isles. *J. Genetic Genealogy*, Vol. 3, No. 1 (Spring 2007) which is available for free download at *www.jogg.info/31/campbell.pdf*.

[43] "OGAP" for "Oxford Genetic Atlas Project."

haplotypes accounted for a little over 60% of the haplotypes, and the twenty least common accounted for a little over 32%. Only the top nine were found in 2% or more of the R1b haplotypes. To give you a sense of scale, 33 haplotypes amounts to 2% of the data, round numbers.

Next, Campbell had a look at the "regional affinity" of the top twenty haplotypes; that is, he tabulated the number of each of the haplotypes that were found in each region, and then normalized these to correct for the very different sample sizes. Though in a couple of instances the relative imbalance was striking, for the most part even the differences one plainly sees do not have sharp boundaries. The most striking regional specificities were for:

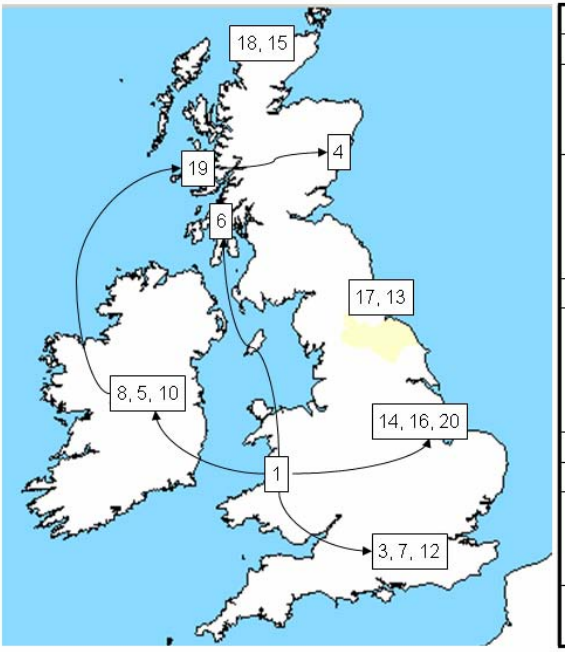
- OGAP 15 and 18 in the Northern Isles
- OGAP 8 in Ireland and Argyll, and to a lesser extent in other Scottish regions (and Northumbria) that probably also had significant Dal Riata input. OGAP8 corresponds to the Ui Neill cluster, which in turn correlates highly with R1b1c7.
- OGAP 10 in the Hebrides and the Isle of Man, and to a lesser extent in Ireland
- OGAP 19 in Ireland and the Scottish Highlands
- OGAP 17 in North England, and to a lesser extent in East Anglia
- OGAP 16 in the Isle of Man, and to a lesser extent in Northumbria.
- OGAP 4 did not achieve the striking prevalence of OGAP 8, 10 or 19 in any individual Scottish region, but it was found at relatively high rates in all of the Scottish regions except the Northern Isles and the Borders, and it was found only in Scotland. We have not yet discovered an SNP shared by members of this cluster, but it corresponds with McEwan's R1b cluster STR47Scots, and is thought perhaps to be characteristic of haplotypes of Pictish descent.

Campbell put these and a few less strikingly regionalized OGAP clusters on a map of the British Isles. There was no especially prevalent OGAP cluster in Wales, but OGAP 1 (the AMH modal) was well represented there, it is thought to be the ancestral haplotype of all the other clusters, and Wales is a reasonable guess for where the R1b Paleolithic settlers of Britain first established themselves, so he put OGAP 1 on the map in Wales. Then he drew arrows from the OGAP clusters "specific" for each region where he had identified these to haplotypes in other regions that differed by at most genetic distance one.[44] He gave the arrows direction based on his (non-genetically based) understanding of the paths and direction that settlement spread through Britain. In Figure 6, I have redrawn and considerably modified his diagram. Although I had left only the OGAPs that appeared in his diagram on the map, I have added modals for Ewing, and Ewing Groups 5 and 6 to the data table.[45] Notice that these are

---

[44] Campbell, Kevin D. Geographic Patterns of Haplogroup R1b in the British Isles. *J. Genetic Genealogy*, Vol. 3, No. 1 (Spring 2007), Figure 4. This is not true of the arrow that leads from OGAP 19 to OGAP 4, which is genetic distance 2 away. I have an idea he intended for the arrow to lead to OGAP 4 from OGAP 6 rather than from OGAP 19, because OGAP 4 & 6 are only genetic distance 1 apart.

[45] I put Ewing with the Irish group because on this 6-marker "affinity map panel," it is genetic distance one from OGAP 8 (and also OGAP 9 and 10), but it is no closer than genetic distance 4 (GD 4) on the full 10-marker Sykes panel to any of the OGAP clusters (including OGAP 8, which has the R1b1c7 modal values). GD 4 on 10-markers is a long way, but we know Ewing is R1b1c7, so this is probably the best place to put it. Ewing Group 5 differs from Ewing at DYS 391 = 10, so it is GD 1 from Ewing on the 6-marker "affinity map panel" and also happens to be GD 1 from OGAP 17, 27 & 31. On the full 10-marker Sykes panel, Group 5 is still GD 1 from Ewing, but no closer than GD 4 from any of the OGAP clusters, these include 8, 17, 27 & 49. I put Group 5 next to OGAP 17 in the North



DYS #	393	390	19	391	389i	389ii
OGAP 1	13	24	14	11	13	16
OGAP 3	13	23	14	11	13	16
OGAP 7	13	23	14	10	13	16
OGAP 12	13	24	14	10	14	16
OGAP 14	13	24	14	11	12	16
Group 6	13	24	15	11	12	15
OGAP 16	13	24	14	11	13	15
OGAP 20	12	24	14	11	13	16
OGAP 6	13	24	14	11	13	17
OGAP 8	13	25	14	11	13	16
OGAP 5	13	24	14	11	14	16
OGAP 10	13	25	14	11	13	16
Ewing	13	25	15	11	13	16
OGAP19	13	25	14	11	13	17
OGAP4	13	24	14	10	13	17
OGAP 17	13	24	15	10	13	16
Group 5	13	25	15	10	13	16
OGAP 13	13	23	14	11	14	16
OGAP 18	13	25	14	11	14	16
OGAP 15	13	24	13	10	13	16

Figure 6: Geographic Affinities in Britain

6-marker haplotypes, but not the same as Capelli's six markers, and not the first six in the order that Sykes lists them.[46]

How do the Ewing men in R1b fit into this scheme? Remember that most of the men in the Ewing project are in a "large closely related group," which includes fifteen men in Ewing Group 5 that have a different value at DYS 391 (a marker considered in all of the data sets we are discussing here) than the rest of the men. I do not want to bother with re-calculating how these men fit in the 6-marker scheme used in the map, but in the full Sykes 10-marker panel:

- The Ewing modal is genetic distance 1 from OGAP 8, which is the same as the R1b1c7 modal and the modal for the cluster variously known as the NW Irish cluster of the Ui Neill cluster.
- The Ewing Group 5 modal is genetic distance 1 from OGAP 17 and also genetic distance 1 from the Ewing modal, and even though it is genetic distance 2 from OGAP 8 (the R1b1c7

English group, but I believe it really belongs next to Ewing, whether that be in the Irish group or somewhere else. Ewing Group 6 is GD 1 from OGAP 14 and 41 on both these panels and is no closer to any other OGAP, so I put it next to OGAP 14. Following Campbell's logic, I could have put the Group 6 modal up in Strathclyde, and drawn an arrow up to it from the Central English group

[46] I do not know why Campbell chose to do this, but among other things, it has resulted in OGAP 8 and 10 being identical in this table, because they are distinguished in Sykes' data at DYS 392, where OGAP 10 has the AMH modal thirteen repeats and OGAP 8 has fourteen repeats.

modal), we know that at least one man in Group 5 is SNP-proven R1b1c7, so we can conclude with some confidence that all of them are.

- PT, RL2, and WC2 all exactly match OGAP 1; JMc and JM3 are genetic distance 1 from OGAP 1 and match no OGAP cluster more closely than that.
- SL & ME exactly match OGAP 2. Remember that OGAP 1 and 2 are the most commonly found clusters in all of the data, and are pretty well evenly distributed over the "Celtic" areas of the British Isles, so do not give much in the way of regional information.
- The Ewing Group 6 modal is genetic distance 1 from OGAP 9.
- RM (the only Group 6 man who does not exactly match the Group 6 modal for these markers) is genetic distance 2 from OGAP 9 and 11.
- Our only Ewen participant matches OGAP 30 exactly, but this cluster was found in only nine men (0.06%) of the data, too small a sample size to say anything about regional specificity.
- The three remaining Group 8 men came no closer than genetic distance 1 to any of the clusters.
    - Js is genetic distance 1 from OGAP 7, 26 and 49.
    - TW is genetic distance 1 from OGAP 19.
    - MK is genetic distance 1 from OGAP 4.

Bottom line: based on an analysis of the markers in Sykes' data, the men in Ewing Group 8 (so, in R1b but not in R1b1c7) are definitely not related to the men in R1b1c7 or to one another, except that Ewing Group 6 consists of a cluster of six men, some of whom are known relatives of one another and all of whom appear to be related based on their DNA results, SL and ME are known third cousins, and PT and RL2 are known sixth cousins. In the Sykes data, the Ewing modal is probably most closely identified with Ireland, and although the Group 5 modal appears most closely identified with North England, we do not believe this because of additional data in the STR markers Sykes did not consider and the SNP testing on one Group 5 member. PT, RL2, WC2, JMc and JM3 are in or closest to OGAP 1 (the AMH modal), which has no regional specificity. SL and ME are in OGAP 2, which has no regional specificity, but see Steven Colson's article in this issue of the *Journal* for an interesting discussion of 37 and 67-marker data that he claims shows regional specificity for Strathclyde. Group 6 does not match an OGAP haplotype exactly; it is closest to OGAP 9, which has no striking regional specificity but is best represented in the Northern Isles, the Borders and North England. We can not come to any conclusions about regional specificity for Ewen, Js or TW. MK is genetic distance 1 from OGAP 4, which is the quintessential "Scottish" haplotype and may represent Pictish heritage.

McEwan's Data

Ewing Surname Y-DNA Project participant John McEwan is a prominent member of the genetic genealogy community, and may very well be the ranking expert on clusters within R1b. In 2005, McEwan gathered nearly four thousand 37-marker haplotypes, including 2,553 R1b haplotypes, from YSearch and other sources and performed a cluster analysis.[47] He identified fifty clusters within R1b and designated them R1bSTR1 through R1bSTR49. (You thought I said fifty? I did—those forty-nine

[47] *www.geocities.com/mcewanjc/p3modal.htm*

plus R1bSTR25a make fifty.) Let us see, in table below, how the fifteen R1b Ewings outside of R1b1c7 that have 37-marker data fit into McEwan's STR clusters. For the sake of comparison, I have also included the Ewing, Ewing Group 5 and Ewing Group 6 modals.

R1BSTR-	8	19	12	13	15	21	27	37	39	40	43	49
Ewing modal		7										
Group 5 modal		8										
PT			11		11					11		
RL2			9							9		
JMc				9						9		9
JM3								9				
WC2							6			6	6	
MK	8											
SL											6	
ME											5	
TW							9		9			
Group 6 modal/ WE2, DH, WR2						7						
RM						6						
JM						8						
DS						7						

The STR cluster names are across the top; the numbers below them in the table indicate the genetic distance of each participant (or modal) from the nearest STR cluster or clusters. You can see that the Ewing and Group 5 modals are genetic distance 7 and 8, respectively, from R1bSTR19, and PT is genetic distance 11 from each of three different clusters, R1bSTR12, R1bSTR15 and R1bSTR40.

On first consideration, it may be confusing that although Campbell has identified fifty 10-marker OGAP haplotypes and McEwan has defined fifty 37-marker STR clusters, these sets can not be mapped onto one another one to one. Indeed, twenty-two of McEwan's clusters exactly match OGAP 1 (the AMH modal) at the ten markers Sykes uses; differences among these McEwan clusters are in the twenty-seven markers that Sykes does not consider. On reflection, we should probably have been able to predict as much. What is a little more surprising, perhaps, is that eight of the OGAP haplotypes do match only one of McEwan's STR modals, and thirty-five of them do not match any! I am afraid that full discussion of this could easily get too long winded, but basically, I think this is a result of considering the larger panel of markers—what counts as a cluster is not going to consist of a group of exact matches. Loosely, a cluster will consist of haplotypes within a specified, relatively small genetic distance of a modal for the cluster.

We can see that none of the Ewing men or modals exactly match any of McEwan's STR cluster modals, but that is not surprising or distressing—remember how unlikely it is to find an exact 37-marker match. For 37-marker modals, we should expect to allow some slack. For example, we have defined our "large closely related group of Ewings" cluster on the basis of being within genetic distance five of the Ewing modal. On this diagram, we see that the Ewing modal is genetic distance seven from the R1bSTR19 modal—but we have said this a dozen times; this is the R1b1c7/Ui Neill cluster modal. Most of our participants are between genetic distance six and nine from candidate STR modals. This means that

any such group they belonged to would probably not represent kindreds in a genealogical time frame; for that, we would expect to see a genetic distance of less than five.

What can we make of all this? We can get some general idea of how distantly these Ewings are related to one another, and we can see roughly where they might fit in the big R1b phylogeny tree McEwan has constructed,[48] which will in turn allow interested people to see the surnames of other men in these clusters. It would be terrific if we could correlate these clusters with specific geographic areas, but for the most part this has been impossible. This is because these data have been collected from genealogy hobbyists who have been tested by commercial labs and have voluntarily uploaded their haplotypes to public data bases, where some of them have given their best guess as to where their oldest known paternal line ancestor lived. The vast majority of these folks are Americans of British or Irish extraction, and most of them are not at all certain of exactly where their remote ancestors in Britain may have lived. Many have not ventured a guess, and many of those who have ventured a guess might very well be mistaken. For example, different men in the Ewing project might very well answer the question "Where did your earliest known paternal line ancestor live?" by saying Missouri, Pennsylvania, Ireland or Scotland. As you can easily see, this sort of data is certainly not going to allow us to make conclusions about whether the earliest Ewings and their relatives lived in Argyll or Strathclyde.

To Join or Get More Information

If you are ready to join the project, go to *www.familytreedna.com/public/ewing* and click on *Join this group* at the top of the blue section at the left of the page. Participation by Ewing women is also welcome; they can get valuable genealogic information by persuading a male relative to submit a specimen. You can see results tables showing participant haplotypes on the *Clan Ewing* web site. There are also links on the FtDNA web site 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 dot com*.

David Neal Ewing has been a member of Clan Ewing in America *since 1996 and has served as its Chancellor since 2006. He previously served as Chair of its Board of Directors from 2004-2006. He is also Administrator of the Ewing Surname Y-DNA Project, which he founded in 2004, and he is a regular contributor to the* Journal of Clan Ewing*. Dr. Ewing has a private practice in clinical geriatric neuropsychiatry in Albuquerque, New Mexico. He received his M.D. degree from the University of New Mexico and did his residency training at the University of Michigan Hospital in Ann Arbor, Michigan.*

[48] This occupies four PDF files that are available as follows:

R1bSTR1-15: *www.geocities.com/mcewanjc/37strallhapr1bone.pdf*
R1bSTR16-27: *www.geocities.com/mcewanjc/37strallhapr1btwo.pdf*
R1bSTR28-35: *www.geocities.com/mcewanjc/37strallhapr1bthree.pdf*
R1bSTR36-49: *www.geocities.com/mcewanjc/37strallhapr1bfour.pdf*