

A dated phylogenetic tree of M222 snp haplotypes

*Exploring the DNA of
Irish and Scottish surnames and possible ties
to Niall and the Uí Néill kindred*

WILLIAM E. HOWARD III
AND JOHN D. MCLAUGHLIN

ABSTRACT

Using an application program, Mathematica¹, we have produced a phylogenetic tree directly from a set of Y-DNA haplotypes of male testees who carry the M222 Single-Nucleotide Polymorphism (SNP). The tree has a time scale, derived from pedigrees², which is built into the program code³. Evolutionary relationships of the major features of the M222 phylogenetic tree are discussed. A new approach has been developed that dates a SNP by counting the number of lines of descent on a phylogenetic tree which appear at successive slices of time. We show that the M222 SNP originated between 1400 and 2000 BC and we explore living conditions in Ireland and Scotland at that time. Heuristic arguments point to an origin in Scotland, long before the reign of the Irish Chieftain Niall (of the Nine Hostages). After Niall's supposed reign, but before the year 1000, a proliferation of the SNP took place as evidenced by the many points on the tree at which the most recent common ancestors of current M222 testees lived. This situation probably occurred along with a general population explosion at the time just before surnames were adopted in Ireland. Although we have found no evidence from studies of phylogenetic trees that the

A DATED PHYLOGENETIC TREE

progeny of Niall or his supposed descendants, the Uí Néills, contributed significantly to this explosion, we have found a bimodality in the M222 histogram that may connect the ages of two concentrations of M222 with Irish history. From their positions on the M222 phylogenetic tree, we can show how a few groups of different surnames are related to each other and how the times of origin of an M222 line that later becomes a surname cluster can be derived from a progenitor whose testee-descendants carry many different surnames. In cases where there are two clusters with two different surnames that are closely located in a phylogenetic tree, we suggest that they may share not only a recent common ancestor but also may have lived in close geographic proximity.

INTRODUCTION

A new approach to analyze Y-DNA haplotypes has been published recently⁴. An initial database containing 172 testees whose Y-DNA carries the M222 SNP was briefly discussed. These articles develop a time scale, called a Revised Correlation Coefficient, or RCC that can be used to investigate the evolutionary relationships that tie genetics to genealogy, history, and anthropology over tens of thousands of years by analyzing various clusters of haplotypes. The time scale has been calibrated with over 100 pedigrees, so that one RCC is equivalent to just under 50 years. This paper uses the test results assembled by the Family Tree DNA's R-M222 Project to derive a phylogenetic tree directly from a more extensive list of testees' Y-DNA 37-marker sequences.

Y-DNA TESTING

A laboratory sends a test kit to a testee who brushes a toothbrush-like swab along the inside of his cheek and sends it back to the testing lab for processing. The Y-DNA test result consists of numerical marker values called a haplotype that is analogous to a rough fingerprint. Those marker values tend to identify the testee. All males carry the Y-chromosome; females do not. Every male inherits Y-DNA from his father and passes it on to his sons. Each of 37 markers contains a numerical identifier, and this set of numbers is passed along the

patrilineal line. Were it not for mutations among the markers, which occur at random, that marker sequence would be identical from generation to generation. However, at random times — sometimes every few generations — a marker will change its value. Over many hundreds or thousands of years, the marker values will slowly change as the haplotype passes along male lines from son to son. As generations pass, these values will change so that modern male testees along different lines of descent will have different values, although both men will share that same male progenitor. Although marker changes take place at random, we can estimate the time it took pairs of lines to mutate from the progenitor's haplotype to the current set of testees since mutation changes can be calibrated with lots of pedigrees. In this paper, we focus our attention on testees who share a particular set of markers, called the M222 SNP⁵. This SNP is interesting because its progenitors inhabited known areas of northern Ireland and Scotland many hundreds of years ago.

We should point out a possible misconception about how the Y-DNA trail may involve a particular ancestor. One popular misconception is that DNA evidence points to that ancestor as the origin of their Y-DNA. If a pedigree trail and the DNA evidence both point to a particular ancestor, it is highly probable that both the trail AND the DNA evidence merely go **through** that ancestor without originating there. In the case of a pedigree, the male line from the earliest man, originating many tens of thousands of years ago, merely passes through that ancestor, unbroken, to the present. The line does not die out as long as male children are among the descendants.

This series of papers has shown that it is possible to estimate the time when a Y-DNA line splits off an older line, evolving into, and forming another new branch. The M222 phylogenetic tree contains a very restricted set of haplotypes, which makes it easy to understand where such branchings took place.

M222 – BACKGROUND AND DESCRIPTION

The background of the M222 Haplogroup Project is given by one of us (JDM) and David Wilson, co-administrators of the R-M222 Haplogroup Project (formerly the R1b1c7 Project).⁶

A DATED PHYLOGENETIC TREE

Y-DNA is carried only by males and can be subdivided into various groups called haplogroups; M222 belongs to Haplogroup R, which can, in turn, be continuously subdivided. The M222 SNP is closely associated with a subhaplogroup of R1b1c whose ancestors appear to have come from north-west Ireland and perhaps lowland Scotland⁷. The M222 SNP has been suggested to have been the genetic fingerprint associated with the Uí Néill dynasty of Irish pre-history. However, it is not restricted to the Uí Néill as it is also associated with the closely related Connachta dynasties. Many emigrants from Ireland and Scotland have carried the M222 fingerprint to other parts of the world.

FIGURE 1: The Phylogenetic Tree Based on 320 M222 SNP-tested, 37-marker Haplotypes – see opposite page 22

FIGURE 2: The Phylogenetic Tree Generated from 184 37 Marker Haplotypes of Sixteen Similarly-Spelled or -Prefixed Surnames That Are Tightly Clustered – see opposite page 23

THE M222 DATABASE

A large database containing 684 testees has become available; all testees were purported to carry the M222 SNP, but only 320 of them had been confirmed to carry this SNP by special testing. Since we intended to analyze the haplotypes of the M222 group, we subjected the two groups to some tests and concluded that they were statistically the same, so we felt free to use either group in our analysis since the RCC time scale applies only to haplotypes, not to SNPs. The larger and smaller groups consist of 358 (52%) and 254 (79%) males whose surnames are not duplicated in the database, respectively. The remaining testees have at least one similarly-spelled or -prefixed surname (e.g., Daugherty, Doherty, Dougherty, O'Dogherty, etc). Figure 1 shows the phylogenetic tree that contains the 320 M222 SNP testees. We will discuss later when the progenitor/originator of the M222 SNP lived.

SIXTEEN SELECTED SURNAME CLUSTERS WHOSE MEMBERS CARRY THE M222 SNP

To simplify our study, to minimize the complexity of interpretation,

and to indicate a path to better conclusions when more testee results become available, we narrow our analysis in this section to 184 testees with sixteen surnames, each of which has **both** similarly-spelled names and prefixes **and** are concentrated in well-formed clusters on the 320-testee tree. The sixteen-surname tree appears in Figure 2. Our analysis of these 16 surname clusters can be used to show how surname positions can be interpreted on a phylogenetic tree. Although these particular haplotypes fall within the group that carries the M222 SNP, the same approach can be made on any group of haplotypes. Our conclusions are easier to reach when the haplotype marker values (i.e., their ‘fingerprints’) are relatively similar.

Figure 2 shows the Kit Number and surname of the 184 testees on the long axis of the tree⁸. The short axis of the tree gives the RCC time scale where 10 RCC is approximately 433 years. The tree is produced by Mathematica’s hierarchical clustering program using the 37-marker sequence in each testees’ haplotype. Mathematica looks at each haplotype and arranges them so that the closest markers are placed closer together on the tree. This use of the RCC technique and its associated time scale as input to Mathematica is arguably the first time a phylogenetic tree has been generated with a calibrated time scale.

HOW TO INTERPRET THE PHYLOGENETIC TREE

The original mutation that caused the M222 subclade occurred at the time the original progenitor lived long ago – a time equivalent to a value of RCC above 80, far to the left on Figure 2 as we shall see later in this paper. When he had sons, they carried the M222 line on toward the descendants of today. There were no surnames then. As time went on, and each son had sets of sons, they would carry the same haplotype as their M222 progenitor. But, along all the lines as they descended toward modern times to the right of the tree, mutations would occur occasionally at random, changing the marker values. When a marker changes, it shows up as a junction point on the tree. At a time near RCC ~ 20, males adopted surnames. As sons produce more sons, the number of descendant lines increases. Mutations responsible for changes in the marker values force more and more variance among the haplotypes of those descendants. It is the increase in variance in these

A DATED PHYLOGENETIC TREE

haplotypes with time that permits us to derive our RCC time scale and allows us to estimate the approximate times in the past when these mutations took place. Figure 2 is the phylogenetic tree of Y-DNA results that is derived from the haplotypes of a subset of surnames taken from the larger M222 databases. It shows only the surnames we selected for further analysis.

Since the M222 surname group is very narrowly defined, its phylogenetic tree is relatively uncomplicated. However, these surname clusters suffer from incompleteness because there is a significant correlation ($R^2 \sim 0.78$) between the time to the MRCA of each cluster and the number of testees that have been selected because of their proximity in the tree.

To show how the tree may be interpreted, let us consider the clusters of Howles and McGoverns between Kit. Nos. 64486 and 131412. There are two pairs of Howle haplotypes that are identical (70866 & 56154 and 155522 & 78168), and they are linked at RCC = 0. The first pair shares a most recent common ancestor (MRCA) with 69327 at RCC \sim 2.5; the three share a MRCA with 62896 at RCC \sim 3.3. The second pair shares a MRCA with 131412 at RCC \sim 5. All seven Howles have a MRCA at RCC \sim 4.5. A similar analysis of individual MRCAs of the McGoverns shows similar sets of junction points, culminating in the one at RCC \sim 15 when both the Howles and the McGoverns shared the same MRCA, or progenitor.

The group of McGoverns between 117924 and 177480 are distinctly different from those that have a closer association with the Howles. Those McGoverns are intermixed with the Gallaghers and their ancestry does not intersect with the McGovern-Howle group until the two groups share their most recent common ancestor at RCC \sim 17.

The surname groups below 30896 (McCord) and above 141621 (Gallagher) are distinctly different. They do not share a MRCA until RCC \sim 27, certainly before surnames came into use. Epochs of interest to genealogists seldom go back further than 1000 AD (RCC \sim 20). Both figures show branching activity before that time, with heightened branching activity starting around RCC 20–24 (900–1000 AD), at about the time when population explosions forced the Irish to adopt surnames⁹.

FAMILIA

SURNAMES THAT ARE LOCATED OUTSIDE A CLUSTER

Present day testees carry both surnames and the Y-DNA of their ancestors. There is not a one-to-one relationship between them over time. Our early ancestors had no surnames, but they carried markers that mutated, slowly changing their haplotypes. They were the progenitors of modern day surnames. When the lines of heredity reached 1000–1300 CE, surnames were adopted or assigned. Those surnames remained in the lines of their male ancestors down to the present day.¹⁰ Let's investigate what happens by going back to the time when surnames were chosen, near $RCC = 20$. We know that a 37-marker haplotype will have a mutation about every 137 years, or one mutation in about 5.5 generations, as the line goes down from progenitor to testee, i.e., from $RCC = 20$ to zero. There will be, on average, about 6.3 mutations along the line, but that can vary, perhaps by ± 2.5 mutations because mutations occur at random.

As we progress down various evolutionary lines from the progenitor to a present testee, the lines will 'fan out' due to mutations, placing present testees farther and farther apart on the tree although they share the same progenitor. The more mutations, the greater the distance apart they will be in RCC and on the tree. Tracing the line of a testee back in time can be a challenge, not only because of uncertainties in mutations, but also because surnames can change. Moreover, surname clusters are usually incompletely filled causing further uncertainties in the time when the most recent common ancestor of all its members lived. In summary, mutation uncertainties compounded with surname uncertainties complicate the analysis.

Of course, the problem is mutation uncertainty, and it will move people a considerable distance from other similar surnames on a tree if there are a large number of mutations along the way, and mutation differences between two lines from the same progenitor will just add to the complexity.

In a related study, we have compared a 67-marker tree with a 37-marker tree that contains the same testees. The 67-marker tree shows the same basic form as the 37-marker tree, although the exact positions of the testees will differ presumably because more markers lead to

A DATED PHYLOGENETIC TREE

higher resolution in the positioning of the testees on the tree. Tests indicate that the RCC time scale of a 67-marker tree does not differ significantly from that of a 37-marker tree.

OTHER SURNAMES THAT CARRY THE M222 SNP

Most surnames in a larger M222 phylogenetic tree do not appear in a major cluster. As more test results become available, these and other surnames will become ‘fleshed out,’ resulting in better-defined clusters whose evolutionary history will become more evident. A single surname entry in the tree may also simply indicate that few men with that surname have had their Y-DNA test results published.

A surname may be located on the tree close to a similar surname, but not in that surname’s major cluster. While not in the genetic line of the progenitor of the major cluster, he still shares a common ancestor with those cluster members, and that ancestor will probably be located closer in time than if their surnames were different.

A surname that is located within a major cluster whose members have a different surname may occur from Mathematica’s optimization process or from a misattributed paternity known as a non-paternal event (NPE). This situation may occur anywhere on the phylogenetic tree. An NPE in our tree may represent a son not taking the surname of the man who sired him. Causes of NPEs might include a formal or informal adoption, pregnancies occurring outside of a marriage, children who adopted a stepfather’s name or his mother’s maiden name, a change of name, aliases, clerical errors including mix-ups of babies at or near birth, and misattributions of test results.

EVOLUTION OF TEN SELECTED SURNAME CLUSTERS

Using our approach to interpreting the positions on the tree, we can estimate the times in the past when the most recent common ancestor (TMRCA) of members of selected groups of testees lived, as well as the evolutionary relationships that they have with each other. Figure 2 contains 16 different similar-sounding surnames grouped in clusters either together or close to another surname. Table 1 shows information about them.

FAMILIA

TABLE 1: Data on Surnames in the Phylogenetic Tree in Figure 2

NAME	INTERVAL ON TREE		COMMENTS
	BEGIN/END	NO.	
1 Cain	7739–67808	27/29	Mixed with 1 McLaughlin, 2 Gallagher, 4 McHenry
2 Cowan	19612–35133	14/14	No mixture
3 Doherty	8373–30455	11/13	No mixture but different groups
4 Davidson	22015–84787	4/4	No mixture
5 Dunbar	74862–115904	8/8	No mixture
6 Ewing	27497–105696	15/15	No mixture
7 Gallagher	N74755–150560	8/11	Mixed with 2 McGoverns
8 Greer	42131–129688	13/19	1+2+3+13
9 Howle	69327–131412	7/7	No mixture
10 McAdam	88644–88646	5/5	No mixture
11 McCord	27071–30896	8/8	No mixture
12 McGonigle	48772–63538	9/9	No mixture
13 McGovern	34985–64486	7/12	Five mixed w/Gallagher, Howle & Davidson
14 McLaughlin	84247–127244	16/22	1 Mixed with 1 Cain, 1 w/Cain & Gallagher, 2 w/ Cain, Greer, Doherty
15 McHenry	14276–27077	3/4	3 together + 1; all among Cain
16 Milligan	135550–191000	2+2/4	None closely associated; all among Greer

Key: Column 2 contains the beginning and ending Kit Number of the surname cluster
Column 3 shows the number of testees in the cluster and the total number of that surname on the tree

Column 4 shows the purity of the cluster

To simplify our analysis, we eliminated six surnames because they had more complicated presentations on the tree.

The following double entry table presents some results for the ten remaining groups of testees. The surnames are identified in the first column. Within the diagonal line the entries show the RCC and the corresponding estimated date when the progenitor of those particular testees lived^{11, 12}. The estimates of the cluster TMRCAs along the diagonal should be viewed as lower limits on when the progenitor of tested surname carriers lived, and the estimated error is about 300 years, standard deviation (SD).

A DATED PHYLOGENETIC TREE

At the surname intersections above the diagonal line we show the values of RCC where the most recent junction between the two surnames appears on the tree. This is the point in time when a mutation occurs that splits the evolution of the two surnames into two lines. Further splitting may occur on the downward path toward the current set of testees as more mutations occur, but this point is the MRCA of the two surnames that eventually appear. The dates at the surname intersections below the diagonal on Table 1 show the approximate dates that correspond to the values of RCC for the surname intersections above the diagonal line. These are the dates derived from the junction points where the MRCA of the testees in one cluster share a MRCA of the testees in another cluster. It is obvious that most of the MRCAs of each of these surnames lived before times of interest to most genealogists, but all the testees are closely related genetically. Because of the selection process they all carry the M222 SNP.

TABLE 2: The RCC and Evolutionary Time Relationships Among the Ten Major Surname Clusters

	Cain	Cowan	Davidson	Dunbar	Ewing	Greer	Howle	McAdam	McCord	McGon.
Cain	20/2000	27.1	27.1	26.7	33.4	36.2	27.1	27.1	27.1	26.6
Cowan	800	9.3/1500	21.5	27.1	33.3	36.2	21.5	24.6	24.6	27.1
Davidson	800	1000	4.3/1750	27.1	33.3	36.2	19.3	24.6	24.6	27.1
Dunbar	800	800	800	10.5/1450	33.4	36.2	27.1	27.1	27.1	22.0
Ewing	500	500	500	500	13/1300	36.2	33.3	33.3	33.3	33.3
Greer	400	400	400	400	400	9.5/1500	36.2	36.2	36.2	36.2
Howle	800	1000	1100	500	500	400	4/1750	24.6	24.6	27.1
McAdam	800	900	900	500	500	400	900	10.5/1450	13.8	27.1
McCord	800	900	900	500	500	400	900	1350	10.3/1450	27.1
McGon.	800	800	800	500	500	400	800	800	800	9.5

We caution the reader on one important point. If a tree is produced using a different set of haplotypes of M222 males with similar surnames, their tree and their clustering will look approximately, but not exactly, the same. The Mathematica hierarchical clustering routine minimizes the differences of all the haplotypes using the entire marker set in each haplotype. It does not group testees by minimizing the

distances of their near-zero markers. The result of this process means that the appearance of a testee on the tree may be near, but not within a cluster. Tests have shown that the junction points on the tree are not significantly affected and the RCC time scale remains valid. Date errors in Table 2 are probably of the order of 200 years (SD).

When a testee's position on the tree is compared with his position in a pedigree, the descendency relationship will generally not be identical because mutations will cause discrepancies in tree position but not in pedigree position. An example of this can be found in Figure 2 where Kit No. 33323, with two random mutations out of the ordinary, is placed well away from his surname cluster. If you have a good pedigree, trust it and not the position on the tree. Keep in mind, however, that some discrepancies can indicate that the pedigree is in error because of the occurrence of an unrecognized non-paternal event. DNA errors usually occur on the tree in regions less than $RCC \sim 9$ which is equivalent to three times the standard deviation of one mutation¹³.

SURNAME RELATIONSHIPS BETWEEN PEDIGREES AND CLUSTERS ON THE PHYLOGENETIC TREE

The link between a testee's pedigree and his position on a phylogenetic tree provides valuable cross-fertilization between what we know about the testee's family line and his association with others on the tree. That link is provided though the time scale that ties Y-DNA results with genealogy. Y-DNA cannot prove a pedigree, but it can provide credibility to, or destroy, the correctness of a pedigree. This link can lead to valuable genealogical relationships among the testees, and to more extensive and better pedigrees.

Pedigrees have errors that may run as large as a few percent per generation. Positions on the phylogenetic tree may be in error because of a combination of testing error, unknown mutations, and variations among the selection of haplotypes used to produce the tree. Those errors can cause plotting errors where a testee might be placed in a cluster in which he does not belong or outside a cluster to which he does belong. Errors in time on the tree average about 11% (SD).

There is one facet of these comparisons that needs further exploration – trying to find the geographic location where two families

A DATED PHYLOGENETIC TREE

whose DNA is closely associated might have lived closely together. We already know that testees who carry the M222 SNP have ancestors who clustered in Ulster and lowland Scotland. When we have found that two different surname clusters share a MRCA who probably lived within a time of interest to genealogists, it may be possible to locate the place where those two families lived close together. From Figure 2 the two leading candidates might be (1) the McCords and the McAdams, and (2) the McGoverns and the Howles whose RCC junction points (at $RCC \sim 13.8$ & 15.1 , respectively) are less than $RCC \sim 20$, a time after surnames had been adopted.

We also suggest investigating family geographical locations where surname clusters have an admixture of surnames within them. We searched the available phylogenetic trees for such groupings. Table 2 shows some candidate surnames for which geographical associations might be found. The ten cluster numbers are arbitrary; they enable the cross-links to be found among the clusters.

TABLE 2: The Association of Multiple Surname Associations Within Clusters on the Phylogenetic Tree

CLUSTER > SURNAME	1	2	3	4	5	6	7	8	9	10
Burns			5				3		1	
Cowan		1					3			18
Doherty	3	3		1	3	7	6	1	1	1
Davidson			4							
Forrest					3					
Howle			7				1			
McLaughlin	1	2		11	1				2	2
McCord									8	
McGonigle						9				
Wilson			1					3	1	

FAMILIA

Shaded intersections in Table 2 contain three or more surnames; non-shaded intersections containing fewer surnames may be the result of mutation errors. We note, in particular,

- Burns, Davidson and Howle who share Cluster 3
- Doherty and Forrest who share Cluster 5
- Dohertys and McGonigles who have major representation in Cluster 6
- Burns, Cowan and Doherty who share Cluster 7

SELECTED SURNAME HISTORIES

In this section we will summarize information on the surnames mentioned in this paper that might have a bearing on further connections among their genealogy and their DNA.

McAdam/MacAdam

The five testees in Figure 2 have separate pedigrees and their MRCA is undetermined. The pedigree of Kit 119699 traces to Arieniskill, near Arisaig, Scotland, and the progenitor left for Prince Edward Island in 1790. The progenitors of Kits 65678 and 88646 sailed to Prince Edward Island the same day as Kit 119699's progenitor and can be traced as tenants at Glenuig, near Arisaig, about nine miles from Arieniskill. All were listed under the surname Adamson, but in the 1798 Prince Edward Island census, they were under McAdam. The progenitor of Kit 88644 came to Cape Breton in 1814, and a land petition indicates he was born about 1794 in Prince Edward Island. He settled a few miles from the progenitors of Kits 65678 and 119699, who also came to Cape Breton around the same time. The progenitor of Kit 180243 arrived from Arieniskill in 1790. He is recorded as purchasing land from a soldier at Moidart, Antigonish County, Nova Scotia in 1791. These Highland families are quite well documented in Canadian genealogy sources. Thanks to DNA testing, a pattern appears to be emerging that ties together the Lowland McAdams, Griersons, and Milligans with these Highland lines. There is a MacAdam family who trace their lineage back to the McCaddams of Waterhead in the Parish of Carsphairn between Ayr and Kirkcudbright; their history can be traced with certainty back to the early 1500s.

A DATED PHYLOGENETIC TREE

Burns/Byrne/Byrnes/Bierne/O'Byrne

Over 40 men with these surnames are either SNP-tested or predicted to be M222 based on their markers. Three-quarters of them live in, or can be traced to, southwest Donegal, western Sligo, northern Roscommon, or Monaghan. The Sligo (n=6) and Donegal (n=9) contingents are distantly related and appear to descend from a progenitor who lived in the 13th century, after surnames came into common use. The Monaghan group (n=12) is of different descent, and their Byrne (and variations) surname may be coincidental. The Roscommon group (n=4) may have been part of the Sept O'Beirne of Roscommon, but, if so, this sept would not have had a single progenitor since most Beirnes are not M222.

Cane/Kane/McCan/McCane/McCann/McKane/McKean/O'Cahan/O' Kane and McHenry

'The O Cathains (O'Kane) were an important Irish sept descended from Owen, son of Niall. They descend from Ferghal mac Mael duinn, the Irish High King who died in 718 AD. Originally a part of the Clan Conor of Magh Ith on the Donegal/Tyrone border, the O Cathains later were chieftains of territories in Londonderry County under the ruling O'Neills. A branch of the O Cathains took the surname McHenry from Enri, a great-grandson of Magnusa catha duin (battle of Downpatrick), slain 1260 AD.' According to the authors of *The Ulster Clans* (T.M. Mullin and J.E. Mullan) the founder of the family, Enri or Henry, died in 1428.

Cowan/Cowen

The Cowans have a large DNA project but little to say on this group. The group in the M222 project is probably Scottish since many of them list Scotland as place of origin. Some McCords and Cowans lived south of Ayr in the Scottish lowlands at that time. The name is also common in parts of Ireland.

Davidson/Davisson

Is probably Scottish

FAMILIA

Devenney, Slavens

These names are Irish and probably Cenel Eoghain. MacLysaght (*Surnames of Ireland*) refers to the Slavens as a Cenel Eoghain sept but his source is unknown. The surname Devenney is similar to a tribe name that appears in the O'Clery Book of Genealogies, the clan Duibh enaigh which split off the line of the kings of the northern Ui Neill in the early 9th century.

Doherty/Daugherty/Docherty/Dougherty/O'Dogherty/O'Doherty

The Dohertys are descended from Conal gulfan, son of Niall. First appearing in the annals in 1180, they soon become Lords of Inishowen in Donegal, a territory they held until the early 1600s. As such they were sub-chieftains to the O'Donnells, kings of Tirconnell. The Dohertys share a descent with other Cenel Conaill septs of Donegal, the O'Donnells, O'Gallaghers, and O'Boyles. Kit 38173 was a sample donated by the brother of the current O Dochartaigh clan chieftain.

Dunbar

The Dunbar name is well known in Scotland. There was a royal family of the name but no one knows anything about the M222 Dunbars. One sample gives an origin in Tyrone, Ireland.

The M222 Dunbars are a mystery since they are only a small sub-set of all Dunbars in the public databases, most of which are not M222. Many Dunbars came to Ulster in Ireland during the Plantation period. Some were planters and servitors with land grants.

Ewing/Ewin

Ewings in the project mostly descend from ancestors who came from Scotland to Donegal and Ulster during the Plantation years but they do have matches from Scotland as well. The surname is thought to have originated in the vicinity of Loch Lomond in Lowland Scotland. The Ewings in Donegal are documented by listings in the 1630 Muster Rolls, the 1654 Civil Survey, 1665 Hearth Money Rolls and Presbyterian Church registers for Burt congregation.

McCord

Scottish origin. Some McCords and Cowans lived south of Ayr in the

A DATED PHYLOGENETIC TREE

Scottish Lowlands. The McCords first appear in Galloway in the reign of Robert the Bruce, when Nigel McCorrard was granted land in Wigtownshire and Kirkcudbrighshire. The principal seat lay at Skaith near Newton Stewart in Wigtownshire and they held this land until 1471, when another Nigel sold the land of Skaith to John Kennedy of Blarquhan in Ayrshire. The McCords became tenants of the Kennedy family. The McCords and the McAdams share a MRCA at RCC ~ 14 (1350 AD), much more recent than the McCord tie with the Cowans at RCC ~ 24 (900 AD).

McGonigle/McGonagill/McGonigall/Megonnigil

The McGonigles are also a known Donegal sept but with no pedigree. They are probably also Cenel Conaill. The McGonigles appear in Donegal records as herenaghs (ClanMagonegill) of Aghnis and Kilbegs parishes; they also appear as clerics in Raphoe parish records.

McGovern

The McGoverns were Uí Briúin, said to descend from half-brothers of Niall. Other familiar Irish surnames with the same descent are the O'Connors of Connacht, McDermots, O'Reillys and O'Rourkes. The tribe Niall belonged to was originally known as the Dal Cuinn, or descendants of Conn 'of the hundred Battles.' These related tribes are not descendants of Niall and were known to Irish history as the Connachta.

Greer/Grier/Grierson¹⁴

Scottish origin. Pedigrees on the Greer/Grierson site suggest these families are related and originated in the Nith valley of Dumfriesshire, in lowland Scotland and DNA studies are investigating the clustering. The principal family of Grierson were the lairds of Lag and they descend from Gilbert Grierson of Ard, who first appears about 1400. In 1408, Gilbert Grierson obtained a charter confirming his right to the lands of Lag, which had been sold to him by John McRath of Laucht. Some early mention of the families goes back to the 1200s in Galloway. In Figure 2 Kit Mathematica has placed No. 33323 well apart from the other 18 that are well-clustered on the tree. This glaring anomaly carries two random mutations out of the ordinary, but both

FAMILIA

are in fast-mutating markers. This is the cause of the misplacement; he is otherwise very much in the Grierson mainstream, with three out of four off-modal matches. See also Milligan.

Howle/Howell

The Howles in the M222 project descend from an Epaphroditus Howle, b. c. 1685, New Kent Co., VA, who came to the United States from England. According to some the name may be Welsh in origin (Ap Howell from the personal name Hywel). The surname may be Scottish; the Howles appear as a sept of the McDougals of Argyllshire.

McLaughlin/Laughlin/McGlothin/McGlothlen/McGlothlin/ McLauchlan

There are a number unrelated M222 McLaughlin samples in the project but the largest related cluster represent the old MacLochlainn of Tirconnell sept centered at Derry and the Inishowen peninsula of Donegal. Like the O Cathains, they were descended from Owen, son of Niall, and are well known to Irish history, furnishing Ireland with two high kings, Domhnall MacLochlainn, d. 1121; and Muirchertach MacLochlainn, d. 1166. In 1241 they were defeated at the Battle of Caim Eirge by the combined forces of the O'Neills (their kinsmen) and the O'Donnells and lost their political importance. In later centuries they were active in church affairs in the diocese of Derry and were herenaghs of churchlands in the Inishowen peninsula. In 1609 they appear in the state papers as counselors to O'Doherty.

There are some Doherty samples in the project that match up closely to the MacLochlainn cluster. That may be due to the fact that these two Irish septs lived side by side in the Inishowen Peninsula of Donegal for over 600 years. Kit 75919 has an ancestor born in the townland of Rathdonnell, Donegal. Several other samples in this cluster can also be traced to an origin in Donegal, Tyrone or Derry (43514, 84247, 37175, 153321).

Millican/Milligan/Millikan/Milliken/Millikin/Mulliken/Mullikin/ Mulligan

Members with this surname cluster nicely with the Griersons. Pedigrees show that they have pre-1600 origins, originating in the Nith

valley in Dumfriesshire, lowland Scotland. The Milligan/Milliken surnames were originally spelt as Amuligane. The principal family of Amuligane were the lairds of Blackmyre and they descend from McRath ap Molegan, who first appears in the Ragman Roll of 1296. Molegan, MacRath's father or possibly his grandfather, witnessed a charter by Edgar, lord of Nithsdale, granting the church of Morton in the Nith valley to the Abbey of Kelso about 1210. The McAdams may also be associated with this surname with connections about RCC ~ 27 (800 AD). Pedigree information indicates that the Griersons, Milligans, and Mac/McAdams knew one another and intermarried in the 1500s and 1600s. Some Milligans and Mac/McAdams were merchants in Ayr, Scotland. In Figure 2, Milligan 12068 descends directly from a Scottish family who settled in Co. Antrim early in the 1600s. In that Figure the Milligans and the Griersons appear to have MRCA's at both RCC ~ 11 and 19, or about 1450 and 1100 AD. Milligan Kit No. 12068 descends directly from a Scottish family who settled in Co. Antrim early in the 1600s; he shares a MRCA with the Greens at RCC ~ 10 and with Milligan Kit No. 153408 at RCC ~ 11. In Ireland, mainly in the old province of Ulster, surnames like Milliken and Milligan, and in some cases Mulligan, originated from the southwest of Scotland, where it evolved separately and independently of Ireland. The usual interpretation founded in Irish Surname books points to Co. Donegal. This has led to much confusion over the years.

NIALL OF THE NINE HOSTAGES AND HIS UÍ NÉILL DESCENDANTS¹⁵

The extended database we used contains two types of entries. The first, and most trustworthy type, includes testees who have received an M222 Deep Clade Test that assures membership; the second, which appears to be fairly trustworthy, consists of testees whose haplotype strings contain many marker values that the originators of the database have selected as identifiers of M222. We adopted the M222 identification criteria used by the M222 Project administrators and conducted a detailed study to see if the database would be useful for investigating Niall and his Uí Néill descendants.

Niall *Noígiallach*, also known as Uí Néill or Niall of the Nine Hostages, son of Eochaid Mugmedon, was an Irish king who was first believed to have reigned between about 368 and 405 AD (RCC ~ 36). More recent studies of Uí Néill genealogies and the dates given for his supposed sons and grandsons have led modern historians to believe he is likely to have lived some 50 years later than the traditional dates, dying circa 450 AD (RCC ~ 34–35). He was an ancestor of the Uí Néill clan who had a strong presence in Ireland from the 6th century to the 10th century (RCC ~ 24 to 33). His paternal half-brothers also produced dynasties with known surnames¹⁶.

Since all testees carry the M222 SNP, they will have a single progenitor-ancestor who lived long ago and whose haplotype first carried the M222 mutation. It is evident from the dates in Table 1 that the M222 SNP originated before Niall reigned in the middle of the fourth and fifth centuries, i.e., around RCC ~ 35. If Niall of the Nine Hostages shared the M222 haplotype and one of his 37 markers happened to mutate from his father's, he would be represented as one of the tree branchings between RCC 30 and 40. (Only one of these branches is shown in Figure 2). This corresponds to the time when Niall lived.

We investigated surnames that have been identified as Uí Néill descendants¹⁷. Unfortunately there are not enough unique Uí Néill surnames to undertake a separate study of their haplotypes. In addition, the markers that might identify them are virtually identical to those that identify M222. Both groups, if they are later found to be separately identifiable, eventually trace back to the progenitor of M222 who lived before approximately 1200 BC¹⁸.

We looked for a 'Genghis Khan Effect' – one or a small group of very important males being responsible for a very large number of sons. Candidates in Ireland might be Niall of the Nine Hostages or the Uí Néills, who dominated the leadership in the eras 300–400 AD and 800–900 AD, respectively. The large phylogenetic tree shows no anomaly in the distribution at those times, so we conclude that there is no evidence for such an effect in our analysis.

TIME OF ORIGIN OF THE M222 SNP

Earlier studies of the origin of the M222 SNP have suggested dates ranging between 1360 and 1740 years ago (RCCs between 31 and 40)¹⁹. A Trinity College study in 2005 estimated the time of origin to be 1730 years ago, or about 275 AD with a large SD of 670 years²⁰. These determinations resulted from studies of surname pairs within a much smaller database than we now have. In 2005 John McEwan estimated a date of origin 3362 years ago, or about 1357 BC²¹. A more recent determination estimates the date as 3000 years ago, or about 1000 BC²². The links on our larger M222 trees provide valuable additional information that shows that the mutation occurred far before the life of the putative King Niall or his descendants.

From archaeological evidence centered on the spread of bronze workings, males whose haplotypes fall within the R1b haplogroup that contains the M222 SNP had already settled around the Alps by 2300 BC. They had reached Iberia by 2250 BC, Britain by 2100 BC and Ireland by 2000 BC. Although there may be some uncertainty in the dates, the east-to-west migration pattern seems well established, with both Britain and Ireland having at least a small, primitive population by about 2000 BC. These times were well before the arrival of seaborne explorers, pillagers and traders from the Mediterranean or Scandinavia.

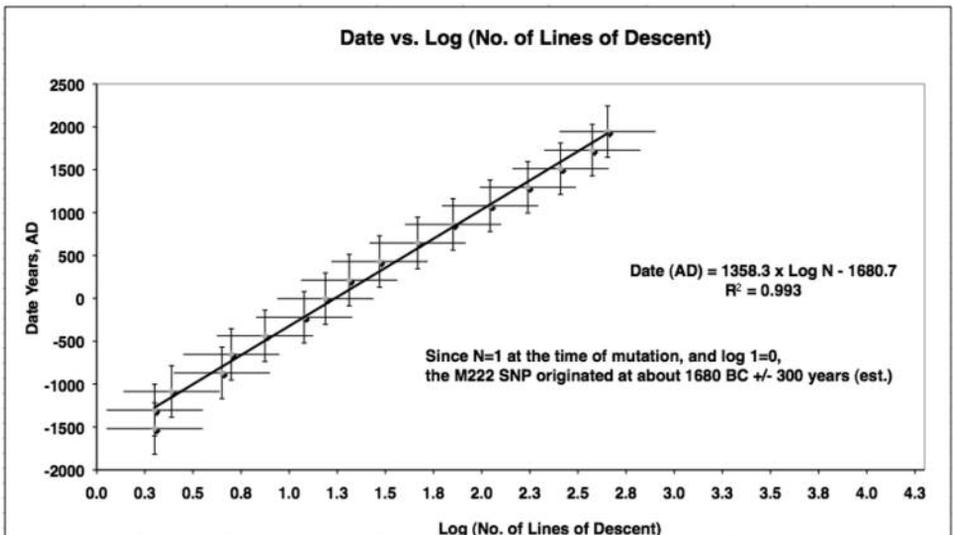
The fact that there appears to be a geographical concentration of the M222 SNP in Ireland and Lowland Scotland puts a time constraint on its origin. The time of origin must have been prior to the arrival of the early seaborne explorers because there is no evidence of a large M222 concentration outside of Northern Ireland and the Scottish Lowlands as there should have been if those seafarers had brought the SNP with them. This suggests that the origin of M222 occurred among the early native population of either Britain or Ireland.

We can estimate the M222 subclade's time of origin by an original, innovative approach based on counts of the numbers of lines of descent on a phylogenetic tree at various points in time. We can extrapolate these lines of descent back in time as their numbers converge to unity, yielding the date when the mutation that caused the SNP occurred. For example, in Figure 2, there are 98 testes at RCC=0. The lines

FAMILIA

linking the testees can then be counted at RCC intervals of 0, 5, 10, 15, ... When these results (98, 41, 24, 12, ...) are plotted against RCC, the plot is exponential. If we then plot the logarithm of the number of these junctions against the value of RCC (converted to a date), we get a straight line. Such a plot for Figure 2 indicates a date of 355 AD for the progenitor of all the testees represented in that figure. This date, corresponding to an RCC of 37, is the point at which the progenitor of the Greer surname joins to the rest of the surnames in the figure. So the extrapolation and the RCC of the progenitor in Figure 2 agree, proving the method. However, that extrapolation points only to the progenitor of the sample of 98 testees. It proves the technique, but it does not point to the original progenitor of the M222 SNP. To do that, we need a larger set of M222 testees.

At least two larger sets of testees exist – one for the 320 testees who have been SNP-tested and another for the testees in our 683 testee database that have a high percentage of the M222 markers.²³ A plot of the number of lines of descent vs. date indicates an origin for the M222 subclade at 1772 BC and 1484 BC for the 320 and 683 database samples, respectively. The two graphs of the data are nearly identical, especially when the error bars are considered. We averaged the two plots and present the result in Figure 3. The derived date of origin is 1680 BC (RCC ~ 84) with an estimated standard deviation (SD) of 300 years.



A DATED PHYLOGENETIC TREE

FIGURE 3: Date vs. Log of the Number of Lines of Descent in Two Well-Populated Phylogenetic Trees (The error bars (SD) on each point are the log (square root of the counted number of lines) and 300 years)

LIFE IN SCOTLAND AND IRELAND IN 1400–2000 BC

Archaeological evidence for the earliest human settlement in this area of the world dates from 8000 to 7000 BC when England was probably still connected to Europe by a land bridge opening up new lands for early hunter-gatherers whose DNA showed they were continuing to move westward. Passage cairns dating just before 3000 BC indicated places of early worship. By that time farming was underway and cattle, sheep and goats had been domesticated. Portal tombs dating to 3000–2000 BC have been found and wedge tombs were being built at about the time of the M222 progenitor. In 1680 BC Stonehenge would still be in use for another half millennium in Britain.

These original pockets of inhabitants in England, Scotland and Ireland probably did not intermingle and they may have been competitive. Their first settlements probably took place near the coast. Fishing probably began first; settlements broadened along the shorelines, and as the settlers moved inland, hunters became farmers. They probably preferred warmer climates where farming was easier, so they avoided the Scottish highlands. With time, hunters would have taken up farming, remaining in regions where crops would have sustained a very slowly growing population.

1680 BC is in the Early Bronze Age. Because of the time when the M222 progenitor lived, he probably spoke a pre-Indo-European language that was very simple, although alphabets were beginning to be used in more advanced civilizations east and south of the Mediterranean Sea. Today's inhabitants probably could not understand his language. The living conditions in both Lowland Scotland and Northern Ireland in 1400–2000 BC were primitive. The lower mobility of early settlers, combined with struggles to survive, necessitated the construction of structures like crannogs for protection. Their language and their isolation may have rendered communication with neighboring settlements very difficult. Their settlements consisted

FAMILIA

of small numbers of isolated people who were barely able to make a living and who needed protection from both the elements and from marauders. Crannogs were common by 1200 BC. They were often built as circular structures on artificial islands near bodies of water, large enough for families and reachable by boat or causeway. Crannogs provided shelter, protection from invaders and a base from which to fish and hunt. A harsher climate probably drove the construction of crannogs first in Scotland around 3000 BC. They were built in Ireland later where they are now most common.²⁴

THE PLACE OF ORIGIN OF THE M222 SNP

No hard evidence exists about the early origins of the inhabitants of Scotland or Ireland at this early date. We will transition our discussion to follow from relatively hard facts to conjecture and heuristic arguments.

First, here is what we believe we know for certain:

- 1 The M222 SNP falls within the R1b haplogroup, indicating that the ancestors of M222 migrated from east to west in Europe and finally arrived in the UK and Ireland from Europe. Their haplogroup strongly suggests an early origin in Europe rather than Scandinavia.
- 2 We have a map showing the distribution of the M222 SNP in Ireland published by Moore et. al., of Trinity College, Dublin, which we will refer to as the Trinity study²⁵. Figure 4 shows the current distribution that they found.
- 3 We have a moderately well-established date for the origin of M222, 1680 BC (SD 300 years), derived from the numbers of lines of descent resulting from mutations on a phylogenetic tree of M222 results, where the time scale is calibrated by pedigrees.

FIGURE 4: A Phylogeographical Contour Map of Ireland Showing the Frequency Distribution of the M222 SNP That Peaks in Northwestern Ireland, Reaching a Regional Maximum of 17 percent of the Male Population. (From Moore et al.²⁵)

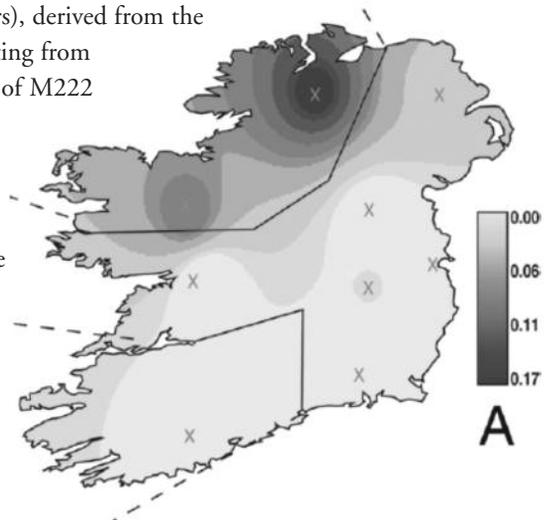


Figure 4 shows an extension of the M222 concentration toward the northeast, toward Scotland. In a recent communication, a Trinity author, Daniel Bradley, confirmed that they saw a concentration (using only six markers) in northwest and west-central Scotland when it was virtually absent in Britain, but Scotland was not the focus of their paper²⁶. We note that the high density near Londonderry does not indicate that the total numbers were higher, since there are large areas of low density along the west coast and throughout much of the north.

The following two contentious issues are still subject to active debate among researchers:

Did M222 originate in Scotland or Ireland?

It is important to note that this M222 distribution was found in 2005, nearly 3700 years (120–150 generations) after the origin of the M222 SNP. The 1680 BC date of origin is more than 2000 years before the arguable reign of Niall and his Uí Néill descendants, 2500 years before the adoption of surnames, and 3600 years from today. It was certainly far before the La Tène culture that signaled the arrival of the Celts. Significant population changes have occurred since the origin of M222 that confuse the analysis and contribute uncertainty. The Normans invaded Ireland in 1170. They completely altered the political landscape in the midlands, setting up their own kingdoms. Many of the native Irish lost their territories completely. Many were forced to migrate elsewhere; a large number went extinct. That may explain the relative dearth of M222 in the midlands when compared to Connacht and the northwest. These areas were largely untouched by the Norman invasion and all kingdoms in that area were retained by the native Irish.

One reason that may explain the high incidence of M222 in northwest Ireland found by the Trinity study relies on a tie to Irish culture and its system of kings and chieftains, all of whom were related by kingship as part of the same tribe. The concentration extends down into the heart of Connacht into Roscommon, Sligo and Mayo. The kings and chieftains held these same territories from the time of Niall or earlier, well into the 1500s until the English conquest. Offshoots of the same tribe in these territories also controlled most of the important church positions and lay positions within the church and society. They

had a selective breeding advantage. The Uí Néill and Connachta enjoyed this privileged position for centuries and they did not move. The one known exception to this rule was the push by the Uí Néill into other parts of Ulster around 1000 AD, and that is also probably why the density of M222 is higher in those areas.

Moffat and Wilson contend that the prevalence of M222 in Ireland proves it originated there and then migrated to Scotland, probably with the Dal Riada settlement in western Scotland²⁷.

A convincing reason why there are two M222 concentrations toward the west rather than being more uniformly concentrated along the northern coast still eludes us. It might be explained either by ancestral arrivals there by sea from the east or from early land migrations prompted by a recognition that life was easier there than in areas to the east.

England and southern Scotland were probably populated before Ireland because they were nearer the major migrations from Western Europe. In 1680 BC the UK and Ireland were still sparsely populated. This epoch in history was much before the beginnings of sea trade and exploration, so migration patterns, although rare, were probably over the land, along the coasts, or over short stretches of sea between land masses. Early population density, higher in the east, might have caused migrations westward toward Ireland, which was separated from lowland Scotland by less than 15 miles at its narrowest point. Such arguments would point toward Scotland as the origin of M222. But the point of origin of the M222 SNP remains uncertain.

Did Niall or his Uí Néill descendants originate the SNP, or carry it? Did they have any influence on the M222 distribution?

There is only one group among Irish pedigrees that is clearly pre-Niall. That group is the Connachta, said to descend from half brothers of Niall. Since the identifiable exemplars of the Connachta carry the M222 SNP (that includes McGoverns, O'Reillys, O'Rourkes and others), then the M222 SNP must have existed prior to Niall and simply flowed through him. This conclusion is consistent with our pre-Niall date of 1680 BC.

We really do not know where Niall lived around 400–450 AD. There

were no writings in Ireland at the time and most of the written stories of Niall originated much later, about 600–800 AD. It is said that he was High King of Ireland, which equates to being King of Tara in Meath, the royal center of Ireland for centuries. It is possible that the High King reigned and had a royal residence at Tara but he might actually have grown up somewhere else, such as Rath Croghan in Connacht, one of the two concentrations of M222.

Niall is never associated with northwest Ireland, nor were any of his predecessors. They were associated with the midlands of Ireland and/or Connacht. Current thinking among Irish historians is that they actually first established themselves in Connacht in the west and their initial capital was at Rath Croghan in northern Roscommon, four miles away from where Niall and his brothers, Brion and Fiachra allegedly lived. They contend that after the time of Niall some of his sons spread north for the first time into Donegal and Tyrone in Ulster, but the rest of Ulster was controlled by the Airghialla, the Ulaidh and the Dal Fiatach. The supposed Dal Riada invasion of Scotland, to which many try to link M222 with Scotland, occurred about 500 AD. But at that date the descendants of Niall were not in northeast Ireland (ie, Antrim). Niall's descendants were not in northwest Ireland originally. That concentration came later after known migration. Irish history also tells us the Uí Néill began expanding from their base in Donegal and Tyrone into other parts of Ulster in about 1000 AD. That area includes counties Londonderry, Armagh, and Antrim. But this expansion was long after the supposed Dal Riada invasion of Scotland.

Pointing out the analogy with the male-line descendants of Genghis Khan, the Trinity study suggested that the high incidence of the M222 signature may be the result of a similar linkage to an ancient, enduring dynasty in Ireland and further noted that the prestige and power of the Uí Néill descendants of Niall supports both the veracity of the semi-mythological early genealogies and provides a powerful illustration of the potential link between prolificacy and power and how the enhancement of the M222 SNP in northwest Ireland where the Uí Néills dominated can be influenced by social selection.

The Trinity study's argument links the Uí Néill descendants only through their presence and the M222 concentration. It is not based on

a population explosion over and above the normal increases in population that were occurring at that time. The lack of a bulge in the number of lines of descent on the M222 tree at these times gives no support to any contention that an enhanced reproductive activity of Niall or the Uí Néills led to such a population explosion nor to the concentration of M222 observed in NW Ireland. We believe that the Trinity study may have been too hasty when it linked Niall to the concentrations of DNA they found. While it may have happened, such a bulge in the number of lines of descent at the time that Niall or the Uí Néills lived is not apparent from our data. Concentrations of DNA do not necessarily indicate that one or more male Uí Néills were responsible for that concentration. It may also indicate a lack of mobility of a non-Niall progenitor and his descendants who carried the mutation. This issue is unresolved.

THE DISTRIBUTION OF THE M222 SNP APPEARS TO BE BI-MODAL

In the course of this investigation we took 320 haplotypes that had been confirmed to contain M222. For each marker in these data strings we computed a modal value. For each of the 320 haplotypes, we determined whether the marker matched the modal value. Since at least 73 percent (27 out of 37 markers) matched the modal for inclusion in the sample, we plotted a histogram of the frequency of occurrence at matches of 26, 27, 28, ... 37 markers. The resulting histogram is shown in Figure 5a. While the error bars in the distribution may not pass a rigid test of significance, there is a definite suggestion from the data that the distribution is bimodal. The same program Mathematica that was used to generate the phylogenetic trees was used to produce an optimum pair of Gaussians, which appear in Figure 5b.

The cause of the two well-separated Gaussians in Figure 5b is unclear. There may be a connection with the two M222 geographic concentrations shown in the Trinity study. There also may be a connection in Irish history where a pre-Niall concentration in Connachta in the south is the starting point of the Uí Néills later migration to Donegal.

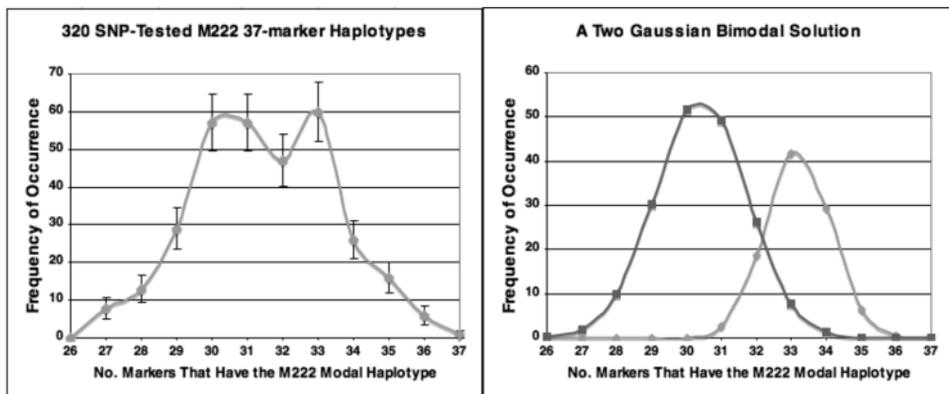


FIGURE 5: 5a. Histogram of 320 SNP-tested M222 37-marker Haplotypes. 5b. The Resolution of the Distribution in Figure 5a into two Gaussians.

The error bars in Figure 5a are estimated to be the square root of the counted number.

A phylogenetic tree that contains only the testees in Figure 5b shows the right hand Gaussian contains testees that all share a MRCA at about RCC ~ 25 (about 850 AD). It is younger than the testees in the left hand Gaussian that trace to before RCC ~ 75, much nearer our date of origin for M222. This suggests that the right hand Gaussian may be related to the Uí Neills and the Trinity study near Londonderry. While the Trinity study showed that it was denser, the left hand Gaussian is larger, and older, again showing a possible tie with the broader areal concentration in the Trinity study dating to the pre-Niall era that is centered on Connacht.

Some manageably small number of descendants of the M222 progenitor who were living 2500 years after the progenitor lived will still have descendants living today. Their haplotypes will have been different from one another and each will approximate the modal haplotype of their descendants. This may explain why the younger right Gaussian differs from the older left Gaussian and how we might associate the former with Uí Néills and the Londonderry concentration and the latter with the pre-Niall older and broader distribution.

FAMILIA

This discussion is speculative. We do not know of concentrations in Scotland and we are unsure that the bimodality exists. We are unsure of the relevant details of Irish history. We are sure that M222 predates Niall and his descendants, but our explanation appears to be consistent with their carrying the M222 SNP, and with the times and locations in northern Ireland. More work clearly needs to be done.

A MORE TENUOUS ARGUMENT ABOUT THE M222 PLACE OF ORIGIN

We can address the pre-Niall point of origin of M222 by using another approach relating to surnames carrying the M222 SNP. Each testee has been asked to identify the country where his earliest male ancestor lived and one of us (JDM) has done much research on the countries of origin of surnames. We have used this information to estimate the country of origin of the M222 subclade in the following way.

We noted 32 zones on a phylogenetic tree that consisted of 295 testees, identified by surname, who had the country of origin of their male line. Each zone generally had between 3 and 117 testee results within it and we noted the Kit Number on the tree where the zone began and ended. We next traced the lines on the tree to the RCC junction point where they all had a most recent common ancestor. These values of RCC varied between 11 and 47.5.

At this point we applied the following approach separately to the testee-named origin and to the country of origin estimated by one of us (JDM). We counted the number of testees who named Ireland or Northern Ireland as their ancestral country of origin, and the number who named Scotland. We ignored citations of the UK, England, Wales, other countries, and 'unknown'. We then plotted the ratio of Ireland and Scotland origins against the date corresponding to the RCC of the zonal progenitors, after normalizing them to the same number of entries. Figure 6 shows the result.

Points above the horizontal zero line in Figure 6 point to Ireland as an origin; those below the line point to Scotland as an origin. While there is considerable scatter in the points ($R^2 \sim 0.1$), there is a suggestive tendency of the points to lie below the line at earlier dates (indicating Scotland). Each of the two linear trend lines, based on the

A DATED PHYLOGENETIC TREE

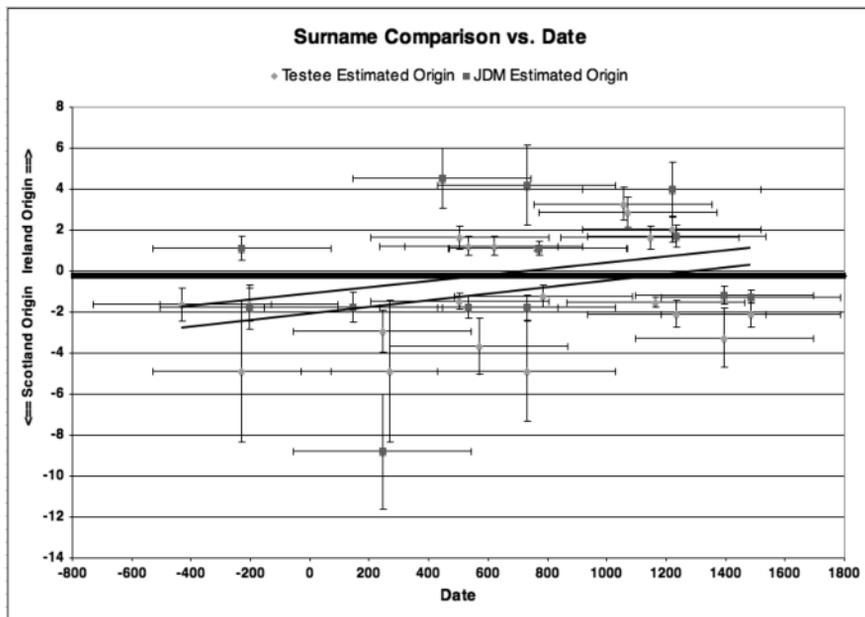


FIGURE 6: Estimated Origins of Surnames vs. Date for Testees Who Carry the M222 SNP

data of each of the two estimates of origin, shows that same trend. Since the progenitor of the M222 SNP lived far to the left, at about 1680 BC, this approach points very arguably to Scotland.

The Trinity study did not cover Scotland. The work of Alan Milliken²⁸ has shown that there are a number of surnames on our M222 phylogenetic trees in which at least one descendant carrying the M222 SNP has a paper trail that leads to an ancestor who lived before 1600 AD and can be traced to an area in Scotland southwest of a line between Ayr and Dumfries. Figure 7 shows the locations of those surnames. The presence of M222 in Scotland only shows that the SNP existed there prior to the exodus to Ulster during the reign of James I. It does not prove that its origin was there.

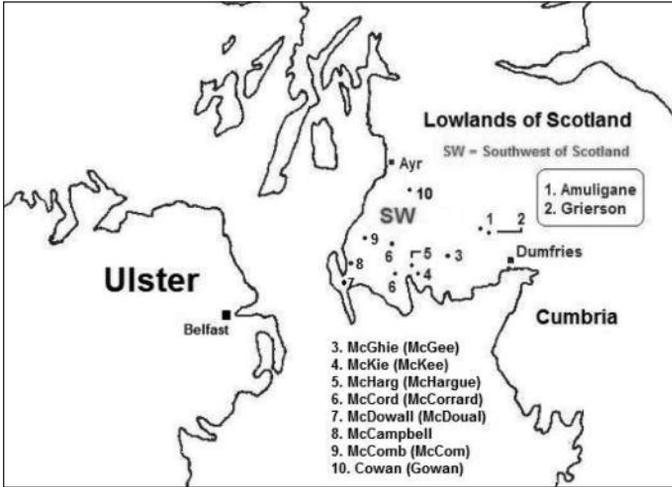


FIGURE 7: The Location of Surnames in the Scottish Lowlands Prior to 1600 in which One or More Members Carried the M222 SNP (Courtesy Alan Milliken).

CONCLUSIONS

We have developed a new approach to dating a SNP through counts of the number of lines of descent on a phylogenetic tree that appear at successive slices of time on the tree. We find the origin of the M222 SNP to be at 1680 BC with an estimated standard deviation of 300 years.

We have studied phylogenetic trees from large 37-marker databases that show evolutionary relationships among males who carry the M222 subclade. The trees contain a time scale and permit a detailed study of the evolution of surname groups within the subclade.

We have shown how ten major surname clusters in the M222 subclade are related to each other and how the times of origin of a M222 lines that later become a surname cluster can be derived from a progenitor whose descendants may have different surnames as a result of marker mutations. We have explored the association of ten multiple surname associations within clusters, finding possible genetic connections among those families that are shown in those clusters.

A major, smooth proliferation of the SNP occurred at RCC \sim 25 (850 AD), just before the adoption of surnames. There is no evidence of abrupt changes in this proliferation caused either by enhanced reproductive activity of Niall or the Uí Néills, or by population changes brought about by plagues that significantly decreased the population elsewhere. However, we have found a possible bimodal distribution for SNP-tested M222 testees that may explain an association among the high concentration of M222 found in Co. Donegal, the age of that concentration, and a tie to Irish history that links it with the Uí Néill descendants who moved to Donegal from Connacht, leaving behind an older and broader pre-Niall concentration of M222 in Connacht.

The location of the progenitor at the time the mutation occurred is an unsolved problem. Archeological studies have shown that Ireland was certainly inhabited, or re-inhabited shortly after the Younger Dryas. Concentrations of modern carriers of the M222 SNP occur in northwest Ireland, but there are no published statistics on M222 in Scotland or the British Isles that can be considered definitive. Working as we do in public databases from people who mostly do not know where their ancestors lived is also not definitive. Heuristic arguments related to broad migration patterns and a very tentative argument involving the origin of surnames point to Scotland. Little research has been done on a possible Scottish origin, but locations there, if found, may connect in time with concentrations found in Ireland. Future work may resolve this issue. The R1b haplogroup of the progenitor certainly indicates that his ancestors came from Western Europe, probably through England and arrived in Ireland by land from lowland Scotland or by sea to the area near Co. Donegal and Connacht.

Lack of sufficient testees, lack of attention to possible concentrations of the M222 SNP in Scotland and a lack of definition of how the Uí Néill descendants can be differentiated from carriers of the M222 SNP have hindered research into its place of origin as well as finding meaningful ties to Niall and his Uí Néill descendants as a separate entity from the descendants of the M222 progenitor.

FAMILIA

FUTURE WORK

More work is needed to achieve a deeper understanding of the place of origin of the M222 SNP. The Trinity study should be repeated in more depth and broadened, with more testees and attention paid to both Scotland and Ireland. Males whose surnames appear to link to the descendants of the Uí Néills should be urged to be DNA-tested. Our date for the origin of the M222 SNP needs confirmation although it is consistent with dates determined by McEwan and quoted by the Europa site. The Trinity date overlaps ours when the error bars of both studies are considered.

ACKNOWLEDGEMENTS

We would like to thank Linda MacLachlan for suggesting this study and for suggesting improvements in presentation at earlier stages of the analysis as well as the following people for discussions that have contributed significantly to this work: Paul Burns, Allene Goforth, J. David Grierson, David E. Hogg, James M. Irvine, Alan Milliken, Sidney Sachs, Fredric R. Schwab, and numerous other discussions with participants on the Rootsworld list <dna-r1b1c7@rootsworld.com> who have contributed to our understanding of the M222 subclade.

AUTHORS AND POINTS OF CONTACT

William E. Howard III: McLean, Virginia – wehoward@post.harvard.edu
John D. McLaughlin: St. Louis, Missouri – Lochlan@aol.com

A DATED PHYLOGENETIC TREE

NOTES

- 1 Wolfram (2010): Wolfram Research, Inc., 100 Trade Center Drive, Champaign, Illinois, 61820-7237; *Mathematica*, Version 8.0
- 2 Howard, William E. III, 'The Use of Correlation Techniques for the Analysis of Pairs of Y-Chromosome DNA Haplotypes, Part I: Rationale, Methodology and Genealogy Time Scale', *Journal of Genetic Genealogy*, 5, No. 2, Fall 2009, p. 256.
- 3 Schwab, Frederic R., and Howard, William E. III, 'Generating a Dated Phylogenetic Tree Directly from Y-DNA Haplotypes, a paper in progress'.
- 4 Howard, William E. III, 'The Use of Correlation Techniques for the Analysis of Pairs of Y-Chromosome DNA Haplotypes, Part II: Application to Surname and Other Haplotype Clusters', *Journal of Genetic Genealogy*, 5, No. 2, Fall 2009, p. 271.
- 5 A short tandem repeat (STR) in DNA occurs when a pattern of two or more nucleotides are repeated and the repeated sequences are directly adjacent to each other. They are referred to as marker values in this paper. A SNP is single base-pair change that is found at certain locations on a chromosome. The Y-DNA test yields 37 numbers, or markers, that we use here to differentiate among the ancestors of testees and when they lived. A SNP must be separately paid for and the presence of a SNP may correlate with particular medical conditions. The M222 SNP and the family of haplotypes carrying the M222 SNP are highly correlated. Recent studies have begun to differentiate between the results of testees who tested positively for the M222 SNP (M222+) and who have been shown not to carry the SNP (M222-). In this paper we are using a small data set of M222+ testees and a larger mixture of those testees and others. Both databases meet our criterion that at least each of 27 out of the 37 markers must have the modal values of M222+. Thus our two databases, having statistically identical databases, can be used in our analysis. The SNP test does not depend on STR values, but STR values are quite predictive of the outcome of the test.
- 6 This genetic grouping was originally announced by Laoise T. Moore, Brian McEvoy, Eleanor Cape, Katherine Simms, and Daniel G. Bradley, "A Y-Chromosome Signature in Gaelic Ireland," *American Journal of Human Genetics*, Vol. 78 (Feb. 2006). See the website of Family Tree DNA, Genealogy by Genetics, Ltd., 1445 North Loop West, Suite 820, Houston, TX 77008, USA. The database for this study can be found at <<http://www.familytreedna.com/public/R1b1c7/default.aspx>>,?section+yresults>_>
See also: the many online files of John McLaughlin at <<http://clanmaclochlainn.com>, (and especially at <<http://clanmaclocklainn.com/R1b1c7>>).
- 7 The designation now in use is: R1b1a2a1a1b4b.
- 8 The test results are the 569 haplotypes presented as a spreadsheet at

FAMILIA

<<http://www.familytreedna.com/public/R1b1c7/default.aspx>>, ?section+yresults > on 18 Mar. 2011 plus 23 additional test results posted on the FTDNA M222 Haplogroup Project web site on 23 Feb. 2011. Subsets of these results are used in this paper.

- 9 See http://genealogy.about.com/cs/surname/a/irish_surnames.htm, a web site that also contains remarks about common Irish surnames, meanings and places of origin.
- 10 Surnames originate from place-names, crafts or occupations, personal characteristics, geographical features, nicknames, and names traced to a patron or a patronymic origin. They were often chosen or assigned without any consideration of the male line that Y-DNA now traces. The DNA of a modern male testee usually follows his father's surname back to the time the surname was adopted, near RCC \sim 20–25, unless an NPE occurs in the line. Earlier than that, the junction point on the phylogenetic tree may trace to many modern surnames. See http://en.wikipedia.org/wiki/Family_name or http://www.powell-pressburger.org/Reviews/45_IKWIG/Mac.html.
- 11 When dates are calculated for the TMRCA of testees in a cluster, we use the estimated date of an average testee's birth (1945) minus (RCC times 52.7). When dates are estimated for the TMRCA shared by two clusters, the estimate of the TMRCA is derived from an estimated date of an average testee's birth (1945) minus (RCC times 43.3). (Howard 2009a)
- 12 Errors in the phylogenetic tree due to mutations can be expressed in units of time. These internal errors, expressed as one standard deviation, are of the order of 2.8–3.2 in RCC, or about 130 years, or about 4–5 generations. The presence of mutations causes large percentage uncertainties in time that amount to over 20 percent within a genealogically interesting time period of less than 1000 years. Since there is no evidence that the RCC time scale is non-linear with time, the percentage error decreases to only a few percent over genetically interesting time periods of tens of thousands of years.
- 13 The values of the number of years per RCC found during the time scale calibration from the 100+ pedigrees ranged from about 39 years to 52 years. In a many-sample Gaussian distribution the SD is of the order of the extreme interval divided by a factor of about 5 or 6. The extreme here is $\sim (52-39)=13$, so the SD is $13/5$ or $13/6$, or about 2.2 to 2.6. This uncertainty in RCC is an estimate of the internal error of the calibration. A change of one-mutation causes a change in RCC by 2.8 to 3.2, depending on which marker changes. We note the approximate coincidence these two RCC standard deviations although they refer to different causes.
- 14 There are two major sources for this surname, neither online. *The Lag Charters* 1400–1720 by Sir Philip J. Hamilton-Griersons published in 1958 by the Scottish Record Society and R.C. Reid's 3 volume collection of Grierson papers/documents relating to a number of families in Dumfriesshire and Galloway, including copies of the original Lag Charters, which are now lost. This important collection can only be accessed by visiting the Ewart Library in

A DATED PHYLOGENETIC TREE

- Dumfries. There are some online abstracts of the Lag Charters.
- 15 See http://www.dnaancestryproject.com/ydna_intro_famous.php?id=niall and < <http://www.familytreedna.com/landing/matching-niall.aspx>> for further information on Niall and his descendants, the Ui Neills.
 - 16 M222 Project co-administrator John McLaughlin assigns surnames to two of Uí Néill's sons and two of his brothers as follows:
 - Cenel Eoghain: Bradley, Donnelly, Campbell, Gormley, McLoughlin, McCaul, Devlin, O'Kane, Quinn.
 - Cenel Conaill: O'Gallagher, O'Donnell, O'Boyle, Cannon, O'Doherty, McMenamin.
 - Connachta: O'Connor, Hynes, O'Rourke, O'Reilly, McGovern, Flynn, McManus
 - Unknown: McKee, Egan, Mulloy
 — John McLaughlin, 'Uí Néill DNA,' <http://clanmaclochlain.com/dna.htm>.
 See also, transcriptions of Irish historical and genealogical manuscripts at same website.
 - 17 See Moore, Laoise T., McEvoy, Brian, Cape, Eleanor, Simms, Katherine, and Bradley, Daniel G., 'A Y-Chromosome Signature of Hegemony in Gaelic Ireland', *American Journal of Human Genetics*, 78, Feb. 2006 that lists the following surnames as possible descendants of Uí Néill: (O')Gallagher, (O')Boyle, (O')Doherty, O'Donnell, O'Connor, Cannon, Bradley, O'Reilly, Flynn, (Mc)Kee, Campbell, Devlin, Donnelly, Egan, Gormley, Hynes, McCaul, McGovern, McLoughlin, McManus, McMenamin, Molloy, O'Kane, O'Rourke, and Quinn. And, from Irish nobility: O'Neill, O'Donnell and Doherty.
 - 18 When we compared the marker values of each of the 37 markers in the haplotype of each testee, we found that the percentage of markers where each testee's results agreed with the modal value for the 592 testees in the database averaged 85 percent (31–32 markers out of 37) with a standard deviation of 6 percent. Thus, considering the statistics, we conclude that no testees should be excluded from the database. Twenty testees had 27 or fewer markers, of which eleven were closely connected to each other (inclusion error) and two had the highest values of RCC found on the extended phylogenetic tree. Had the latter two been excluded, this highest RCC would have dropped from 89 to 73. The formal upper limit would then date to earlier than 1200 BC.
 - 19 Nordtvedt K. (2008) Note to Rootsweb's Genealogy-DNAL@ rootsweb.com of 20 May 2008.
 - 20 Moore, Laoise T., McEvoy, Brian, Cape, Eleanor, Simms, Katherine, and Bradley, Daniel G., 'A Y-Chromosome Signature of Hegemony in Gaelic Ireland', *American Journal of Human Genetics*, 78, Feb. 2006, pp 334–338.
 - 21 McEwan, John, 'Estimating TMRCA and mutation rates for the phase 3 Y chromosome STR clusters via ASD estimates', 11 Mar. 2005
 <<http://mcewanjc.org/p3asd.htm>>

FAMILIA

- 22 A modern European site that gives many haplogroup ages can be found at http://www.eupedia.com/europe/origins_haplogroups_europe.shtml
- 23 The criterion for inclusion into the M222 database is that at least 27 of the 37 DYS markers (73%) must equal the M222 modal values.
- 24 See: <http://www.absoluteastronomy.com/topics/Crannog>
- 25 Moore (2006), op.cit. The M222 SNP was not tested for by the Trinity group. It was known at the time but not routinely tested for until several years later. The concentrations found in the Trinity study highly correlate with the presence of the M222 SNP markers so we consider the Trinity concentrations as indicators of the M222 SNP.
- 26 Private communication to W.E.H., July 19, 2011 from Daniel Bradley at Trinity College, Dublin.
- 27 Moffat, Alistair and Wilson, James F., *The Scots: A Genetic Journey*, Birlinn Ltd., April 2011
- 28 Milliken, Alan, private communication 2011. See also: <http://freepages.genealogy.rootsweb.ancestry.com/~alanmilliken/Research/DNAStudies/r-m222-map.jpg>