THE EVOLUTION OF THE GORDON SURNAME: NEW INSIGHT FROM Y-DNA CORRELATIONS AND GENEALOGICAL PEDIGREES

Author(s): Tei A. Gordon and William H. Howard

The Evolution of the Gordon Surname: New Insight From Y-DNA Correlations and Genealogical Pedigrees

Tei A. Gordon and William E. Howard III

Abstract

Surnames can be grouped into families by two methods: (1) matching Y-DNA marker haplotypes assisted by pedigree information, and (2) using correlation techniques. Both methods, applied independently, yield remarkably similar results, with the correlation technique having a slight advantage in determining the members of family groups and clusters. Traditional and correlation techniques produce similar results, with similar uncertainties, when estimating the time at which the most recent ancestor of a pair of testees lived if they are within the genealogically interesting period within about 1000 years. The correlation technique has a decided advantage when times to the most recent common ancestor of a family group, the most recent common ancestor of separate family groups, and the most recent common ancestor of families within separate haplogroups are estimated. Correlation techniques and genealogical pedigrees, working together, are used to explore the history and evolution of surname groups as well as the haplogroups of which they are a part. Totally new information that shows remarkable relationships among pedigrees, cluster and subcluster membership, geographical location, and their evolution has become apparent through this study. The appearance of surname subclusters within a surname cluster can indicate a strong, confirmable tie to pedigrees when they are available for members of a subcluster. We have uncovered correlations between recent historical activity and the formation of subclusters. The times when Gordon family clusters first appeared, and when the most recent common ancestors of Gordon interclusters lived, are compared with the chronology of the Gordon surname and events in European and Scottish history to provide additional insight into the history of the Gordon surname and possible origins from 2500 BCE to the present. The earliest most recent common Gordon ancestors who were located in pairs of different haplogroups date to about 17,500 BCE, just after the European glaciers in the most recent ice age began to recede.

In the R1b1 haplogroup, the ISOGG time estimates, the RCC time scale, the Y-DNA evidence and our results are consistent with an origin of the Gordon surname in areas near modern Turkey and Greece. Comparison of the ISOGG dates with those determined using the RCC time scale shows good agreement and no inconsistency between the RCC- and ISOGG-derived estimates.

The times derived from the RCC matrix for the early migrations of the I1 haplogroup into the British Isles from Scandinavia and from Western Europe agree well with the history of the area derived from archaeological excavations, genetics and anthropologic studies.

Introduction

An alternative method for grouping and analyzing Y-DNA results has been presented, with the suggestion that a correlation approach be used in conjunction

with traditional methods and genealogical pedigrees (Howard 2009a, Howard 2009b). Those papers discussed the advantages and disadvantages of the correlation and traditional analysis. They introduced a time scale that can be applied over tens of thousands of years to explore the evolution of surname clusters and different haplogroups[1]. Our goal in this paper is to explore the extent to which correlation techniques and genealogical pedigrees, working together, offer additional insight into the history and evolution of a surname group.

The power of the correlation approach will be demonstrated as well as the degree to which genealogical pedigrees contribute to the analysis. The limitations of the approach will be explored.

The Gordon DNA Project accepts results from various labs; Kit Numbers referenced in this paper are from Family Tree DNA, which represents the bulk of those testees in the project[2]. All identifiable information has been removed to protect their privacy.

The Gordon surname was chosen as the focus of the analysis for reasons that included:

- The availability of a large number of Y-DNA testees with 37-markers, the benchmark for calculating the RCC matrix[3].
- The availability of a sufficient number (~ 12) of different Gordon surname pedigree groups to make comparisons with the clusters we expect to find from DNA.
- The origin of the Gordon surname in Scotland, a relatively isolated part of the world in early days.
- An unprecedented library of third party pedigree research, which can add credible insight into cluster analysis (Bulloch 1903 & 1907; Gordon 1949; Skelton & Bulloch 1912).
- Comprehensive documentation on three major lines reaching back nearly 1000 years (Bulloch 1903,6,7; Gordon (4 references); Skelton & Bulloch 1912).
- Documented matching pedigrees within the same line, which allows a check on the mutation rates used to calibrate the RCC scale[3].
- The possibility of resolving conflicting theories of the origin of the Gordons through the application of RCC correlation techniques.
- Potential of location-based surname origins dating back 4500 years to the Gordion (Gordium) site settlements in Anatolia (near Ankara, modern-Turkey), then Grecian Macedonia, Gaul, France and Spain.
- An international sampling of Gordon testees exists which can be useful in isolating unique mutations for tying together a common timeframe to a common ancestor throughout various parts of the world.

We decided to do the groupings by pedigree and marker string comparisons separately from the cluster groupings using RCC matrix analysis. By keeping the processes scrupulously separate, we could better compare the two group/cluster determinations, lending more credibility to the final comparisons of the two approaches. In this paper we refer to Gordon families derived by pedigree comparisons and traditional haplotype matching as groups. We refer to Gordon families derived by RCC matching as clusters.

The favorable outcome of using the correlation approach in conjunction with good pedigrees, strongly suggests that the methodology can be applied with success to other surnames under similar selection criteria.

OMNASTICS OF THE GORDON SURNAME[4]

Though theories on the origin of surnames abound for most every Scottish family, a unique combination of tradition, legend and references by some of the most respected names of antiquity give unique opportunities for Y-DNA and RCC to help test these theories and unravel the onomastics of the modern Gordons in Scotland.

Early Romans were using up to four names during the expansion of their empire. The practice of passing one of their names on to their children no doubt left their mark in territories they conquered. Such naming traditions likely were brought back by crusaders returning from the Holy Lands.

In the 12th century the widespread practice of given and family names was already in use first by aristocracy, nobility and the wealthy, followed by the merchant classes. By the 18th century, hereditary surnames became the norm among the general population, which had been using a single given name. It was common Western European practice to adopt a surname based upon one's location, occupation, patronymic name or even physical characteristic.

According to early-19th century Scottish historian George Chalmers in his series Caledonia, "the founder of the Gordon family came from England in the reign of David the First (1124-53), and was granted the lands of Gordon (anciently Gordun, or Gordyn or the Gaelic Gordin, "on the hill")" from which the family may have derived its name (Chalmers (1742-1825)).

Yet other ancient historians and even Chalmers in his research indicate that the Gordons likely already had their surname prior to arriving in Scotland, and he surmises possibly from France, and Macedonia before that, giving their name to the town of Gordon in the borderlands. Thus, Chalmers acknowledges that prior to Scotland, the Gordons may have originated in Gordonia in Macedonia and migrated to ancient Gaul, with their seat in present-day city of Gent, Belgium (Chalmers (1742-1825)).

Chalmers indicates in his writings that generally the eleventh and twelfth century Normans and other foreigners who came to Scotland and England had no family name when they received lands from William the Conqueror or from other noble families as many Scottish families did.

Extending such universal generalization to all early Scottish families, however, is perhaps an over-generalization. Allowing for a little phonetic flexibility reflecting the diversity of the period (Gordon versus Gourdon, Gordoun, Gordun, de Gordoun, de Gourdon, Gurdon, etc.), changing geographic boundaries (medieval Normandy), as well as the origin of the base of arms of the Gordons (stars and boars' heads) and badges (Ivy), the Gordon researcher learns that there are plenty of clues to an older family history that RCC can help reveal. Throughout this paper, we follow the spelling variation given in the original source.

Among the Clans, a few larger Scottish families enticed fragmented, smaller clans to align interests during economic downturns by offering food in exchange for changing their surname. For military families, such as the Gordons, this was an especially useful approach during the tumultuous 15th to 18th centuries as a means to increasing the ranks and thus, many Gordons were sometimes referred to as the Bowl o' Meal Gordons (Dickens 1887, pp. 58).

Leveraging the internet to access voluminous libraries of books, and by continuously monitoring new Y-DNA results, while applying new and evolving tools such as the RCC technique of analysis, we find that we may be able to finally prove some of these ancient traditions, theories and early citations.

To date, no studies have focused upon the time relationships between the haplogroups of the Gordons (R1b1, I1 and I2 groupings), nor has any previous study revealed the relationships of the Gordons and their subclusters.

DETERMINING GORDON HAPLOTYPE GROUPINGS USING TRADITIONAL MARKER AND PEDIGREE COMPARISONS

As is likely the case in many surname projects, the current groupings for the Gordon surname DNA project consisted of the following steps:

1. Group the DNA haplotypes using traditional marker comparisons, such as Family Tree DNA's proprietary algorithm FTDNA TiP calculator. Kits within a haplotype that match within 40-generations (approx. 1,000 years) are grouped together.
2. Look for a pedigree match in each group.
3. When a pedigree is found that describes the DNA grouping, select the best-documented with longest history as the most representative pedigree(s) for each group and designate it as the benchmark.
4. Look at the other members of each DNA group. As long as their pedigrees are NOT inconsistent, firm them up into the group.
5. Look for unique mutations and organize into sub-groups if possible.

DESCRIPTIONS OF THE GORDON PEDIGREE BRANCHES

Lending credibility to the French and Norman origins of the Gordons, R1b and I1 branches respectively, historians note that prior to the Scottish Gordon families, the French families of de Gourdon were already using and continue to this day to use the ivy leaf as their badge and the same base of arms as their Scotland cousins.

In line with a descent from the French and/or Norman de Gourdons, it is recorded that the following tree outlines the generally accepted Gordon family history, highlighting its three major branches[5]:
1. Jock & Tam Gordon branch
2. Sir William Gordon branch
3. Seton-Gordon, the Ducal line branch

Gordon Skeleton Tree of Three Branches

Sir Adam de Gordoun — G1
"Laird of Gordon" – Berwickshire
"1ˢᵗ of that Ilk"
b. abt. 1035 – France, d. 1090 – Northumberland

Sir Adam de Gordoun II — G2
"Laird of Gordon" – Berwickshire
"2ⁿᵈ of that Ilk"
b. abt 1070, d. 1138

Sir Richard — G3
"3ʳᵈ of that Ilk"
d. 1170

Adam de Gordon — G3
of Huntly
b. abt. 1105

Sir Thomas — G4
"4ᵗʰ of that Ilk"
d. 1214

Alexander — G4

Sir Thomas — G5
"5ᵗʰ of that Ilk"
b. abt 1193

Adam — G5

Alicia — G6
b. abt 1230

Sir Adam — G6
of Huntly
b. abt 1220

Adam — G7
of Huntly
b. abt 1254, d. 1296

Sir Adam — G8
"Lord of Gordon"
(Strathbogie)
b. abt 1273, d. 1333

Sir Adam — G9
"Lord of Gordon"
b. abt 1292, d. 1346

Sir William — G9
"Laird" (Strathbogie/Schol/Vernon, Lochinvar)

John — G10
b. 1315

Sir John — G11
b. Abt 1340, d. 1394

Sir Adam — G11
"12ᵗʰ of that Ilk"
b. abt 1350

"Jock" of Scurdargue — G12
b. abt 1360, d. abt 1420

"Tam" of Ruthven — G12

Elizabeth GORDON — G12
(d. abt 1438)
& Alexander SETON
(b. abt 1395, d. abt 1441)

'Jock and Tam' Gordon Branch (I 1)

DUCAL LINE "Seton-Gordon" Branch (R 1b1)
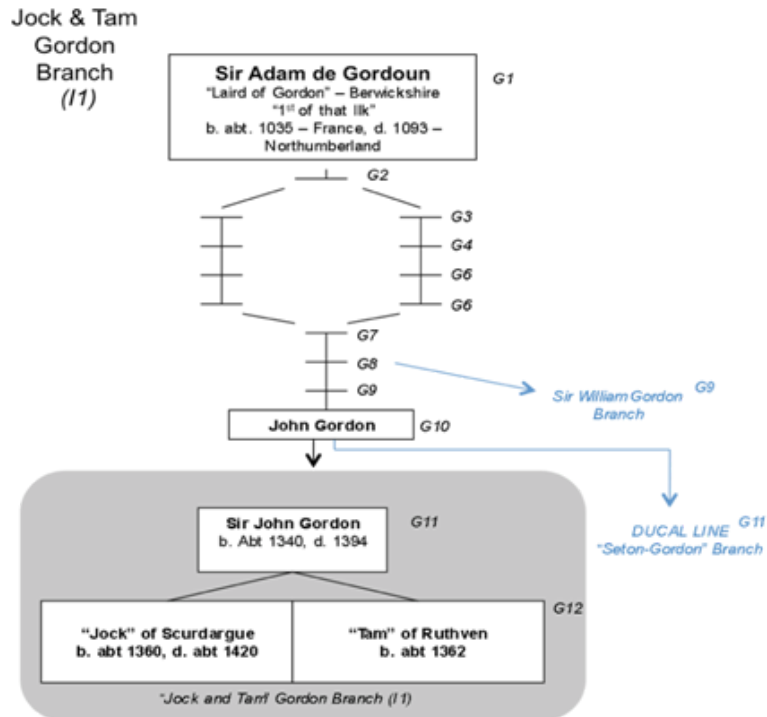
Sir William Gordon Branch (I 1)

Descriptions of these branches follow.

<u>The Jock and Tam Gordon Branch</u>

This is the largest, oldest documented branch of the Gordons, dating back to the 11th century and the first recorded and single male progenitor of modern Gordon lines in the Gordon surname DNA project – the Laird of Gordon – credited as the founder of the House of Gordon in Scotland. Given the limitations of FTDNA algorithms, this large group cannot be broken down into sub-groups. Several kits have verifiable documented pedigrees among those in the Jock and Tam branch.

The John "Jock" Gordon of Scurdargue and Thomas "Tam" Gordon of Ruthven (Jock and Tam) represent one of two unbroken lines of Gordon males back to the Laird of Gordon (died at Battle of Standard, 1138) and held its seat in the Highlands.
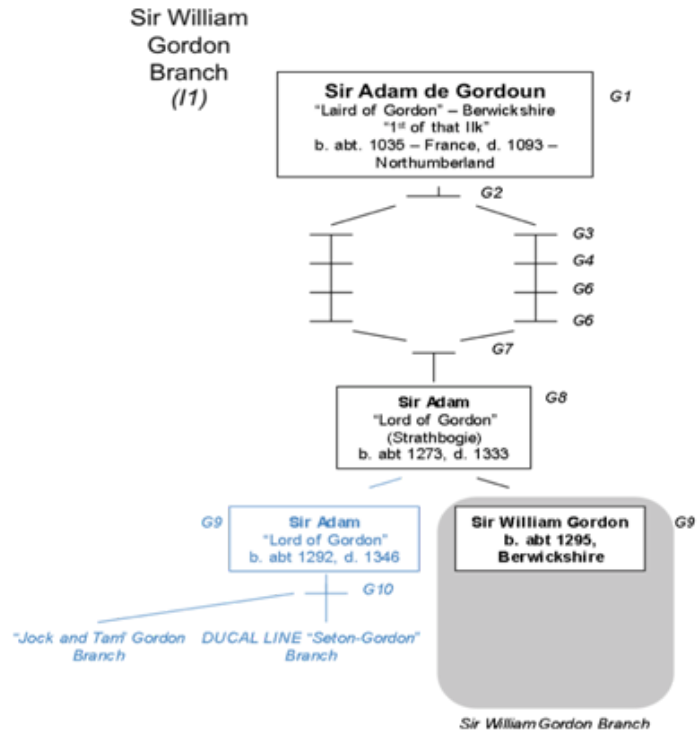
The Jock & Tam Gordons belong to the I1 Haplogroup, typically found in northern Scandinavian countries, which supports the theory that this branch of Gordons came from Normandy to Scotland.

Jock & Tam Gordon Branch (I1)
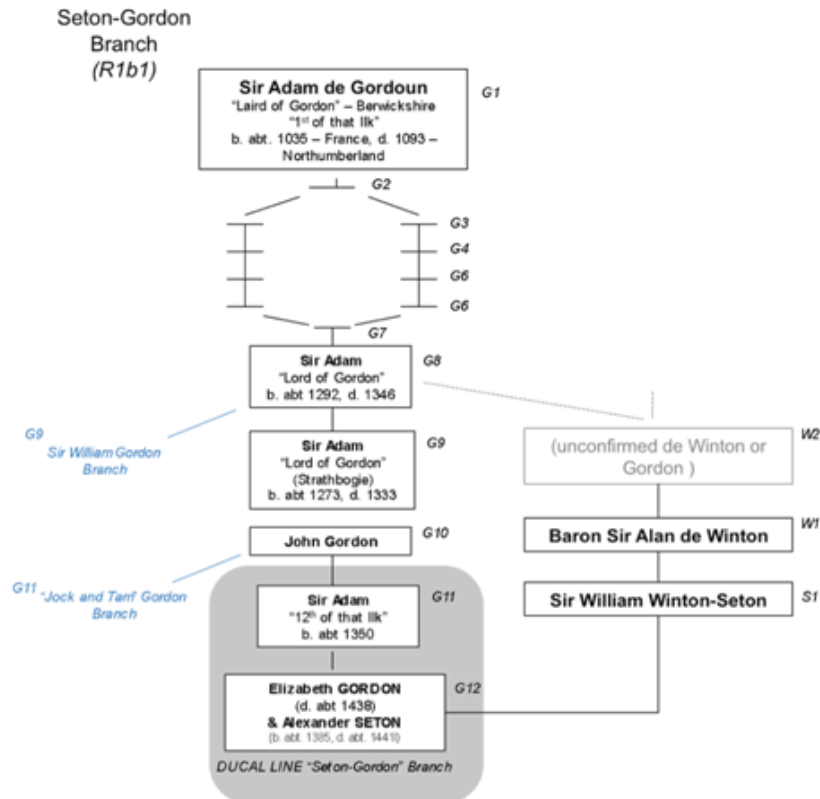

The Sir William Gordon Branch

This branch is documented to have branched from ancestors of the above Jock & Tam Gordons back in the 13th century (Bulloch, op cit). Prior to this analysis, it was thought more than likely that some of the kits currently listed under the above Jock & Tam Branch actually belong in this branch. Several other testees in this group and of the R1b1 Haplotype have family claims of descending from this line, but the I1 Haplotype would seem to support generally accepted Gordon family history. Therefore, the R1b1 sub-grouping is likely not in an unbroken line. One explanation could be that a Seton-Gordon married into the R1b1 branch of the Sir William Gordon family, thereby producing this anomaly. The benchmark kits for this branch are Kit Nos. 89515 and 93333. Both have well-documented pedigrees[6].

We have yet to determine a unique genetic sequence for the Sir William Gordon descendants, as the results are almost identical to those of the Jock & Tam Gordon descendants. However, as we see later in this paper, RCC analysis has revealed a branching of the Gordon Cluster AA cluster at around 1300 CE, about the time of the branching of this Sir William Gordon line. As further testees with documentation back to each of these lines are found, we should eventually be able to determine unique sequences for each branch.

Sir William Gordon Branch (I1)

Sir Adam de Gordoun
"Laird of Gordon" – Berwickshire
"1ˢᵗ of that Ilk"
b. abt. 1035 – France, d. 1093 – Northumberland
G1

G2

G3
G4
G6
G6

G7

Sir Adam
"Lord of Gordon"
(Strathbogie)
b. abt 1273, d. 1333
G8

G9 **Sir Adam**
"Lord of Gordon"
b. abt 1292, d. 1346

Sir William Gordon
b. abt 1295,
Berwickshire
G9

G10

"Jock and Tam" Gordon Branch   DUCAL LINE "Seton-Gordon" Branch

Sir William Gordon Branch

The Seton-Gordon Branch

The single verifiable pedigree for the Seton-Gordon branch is Kit No. 35045, as will be seen in a later section, this kit has not been included in the RCC analysis because it most closely matched only one other test kit and thus did not qualify as a major Gordon RCC cluster member.

Seton-Gordon
Branch
(R1b1)

Sir Adam de Gordoun   G1
"Laird of Gordon" – Berwickshire
"1st of that Ilk"
b. abt. 1035 – France, d. 1093 –
Northumberland

G2

G3
G4
G6
G6

G7

Sir Adam   G8
"Lord of Gordon"
b. abt 1292, d. 1346

G9
Sir William Gordon
Branch

Sir Adam   G9
"Lord of Gordon"
(Strathbogie)
b. abt 1273, d. 1333

(unconfirmed de Winton or   W2
Gordon)

John Gordon   G10

Baron Sir Alan de Winton   W1

G11 "Jock and Tam' Gordon
Branch

Sir Adam   G11
"12th of that Ilk"
b. abt 1350

Sir William Winton-Seton   S1

Elizabeth GORDON   G12
(d. abt 1438)
& Alexander SETON
(b. abt. 1385, d. abt. 1441)

DUCAL LINE "Seton-Gordon" Branch

Small Groupings of Gordons

Groups of two or more testees with the Gordon surname that match each other but that do not match one of the three major Gordon branches within the last 1,000 years (modern origin of the Gordon surname) have been included in this section. These Gordons may yet ultimately be attached to one of the 150+ Gordon branches. However, members of this grouping should also consider the possibility of an undocumented non-paternal event in their family history.

Gordon Septs

Not all Gordons were Gordons by birth. Many were Gordons by bond, pledging allegiance to the House of Gordon, some even taking the Gordon surname.

There are 136 unique haplotypes and 45 different surnames/Septs in the House of Gordon. Definitions of a Sept are ambiguous at best and open to interpretation. Since no documented ties have been found between Gordons and their Septs, one of the goals of the DNA project is to determine whether there may sometimes be a genetic link between any of its Septs with one of the three major Gordon pedigree branches within a genealogically significant timeframe.

All Septs, even if matching any Gordons, are listed in the project only on the Sept section of the webpage and a separate notation is made when a match with any Gordon is determined.

Ungrouped Gordons

Gordon testees who do not genetically match any other Gordons for the last 1,000+ years are included in this grouping. Furthermore, documentation for these singular Gordon testees do not indicate any relation to other Scotland Gordon branches or Gordon Septs and often documentation does not extend more than a few hundred years or about 16-generations. In other words, these independent testees do not have a linear paternal heritage, genetic or documented, leading to the progenitors of the three major branches.

Since there are a total of 150+ documented branches of the Gordons, including the three major branches, it is reasonable to expect new genetic branches will be found. However, we must also consider the possibility of a non-paternal event, such as a male from a closely aligned family of a different surname, marrying into the Gordons and assuming the Gordon surname, similar to the Seton-Gordons. Likewise, adoption or other undocumented paternal event must also be considered.

It should be expected that ungrouped testees will ultimately match other testees and will be moved to the Small Gordon Groupings, which include two or more matching kits, or possibly a non-paternal event uncovered.

THE RCC CORRELATION MATRIX ANALYSIS: BACKGROUND AND APPLICATION TO A GORDON SURNAME RCC MATRIX

A. The Evolution of Surname Clusters Inferred from the Distribution of Intercluster Data

A1. Background
To understand the relationship between the RCC matrix and the evolution of surname clusters and how interclusters in the matrix add to that understanding, the following observations are important:
- The values of RCC that connect pairs of testees, pairs of common ancestors (CAs) and pairs within interclusters can be mapped directly into a time scale that has been calibrated by pedigrees. This time scale appears to be valid to within a factor of two for tens of thousands of years (Howard 2009 a and b).
- In the RCC matrix, a common ancestor (CA) of a cluster is the progenitor of all pairs of RCC values of testees that make up that cluster or intercluster. That progenitor will have a haplotype that slowly evolves to those of the testees.
- The mutations that take place in the lines from a progenitor to present pairs of testees cause the haplotypes to change, resulting in the non-zero RCC values of current testee pairs.
- The time to the most recent common ancestor (TMRCA) of a surname cluster can be estimated.
- When all members of one cluster are paired with all members of another cluster, they appear in the matrix as an intercluster region, which contains its own unique set of RCC values. Each cluster will have a cluster progenitor (its CA) whose TMRCA can be estimated[1]. The intercluster will also have its own CA whose TMRCA can be estimated from the values of RCC in the intercluster. The intercluster CA will be the most recent common ancestor (MRCA) of the two cluster CAs. Thus the TMRCA of an

intercluster is an estimate of the time when the common ancestor of all the pairs of both clusters lived.
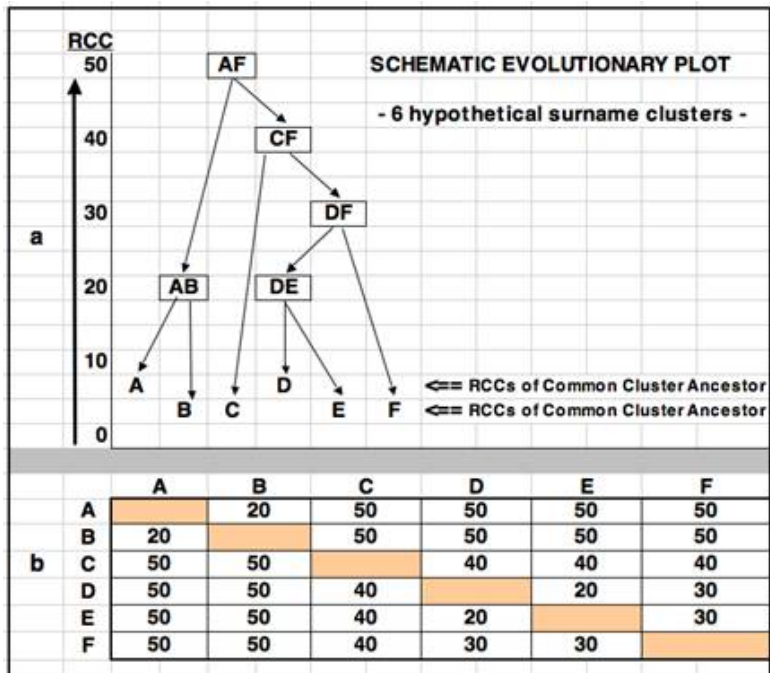
- The relationships between the CAs of clusters and interclusters allow us to find the time relationships among the CAs of all the interclusters, a process by which we gain valuable insight into how and when all surname lines probably evolved.
- During this evolution, different clusters will form, first as subclusters with paired member RCCs < 5, and then as clusters with paired member RCCs generally < 25.
- A common ancestor of a surname cluster or intercluster will have a haplotype whose marker values are unknown.
- When a common ancestor of an intercluster is the progenitor of a new line that leads to a different cluster or haplogroup, his marker values will be those of the progenitor of both lines.
- The haplotype of a progenitor will change through random mutations as the line progresses from the progenitor to, and through, all common descendants of the line, to the present testee. When only one pair of haplotypes are compared, the expected error (standard deviation) can be of the order of 30%, but when ten testees are compared in a cluster or intercluster, the error is expected to decrease by the square root of the number, or about 10% (SD).

Appendix A outlines a useful method for identifying clusters in an RCC matrix.

A2. How Surname Evolution Produces the RCC Matrix

A schematic matrix of a surname cluster and the intercluster regions of a pair of surname clusters were presented in Figures 1A and 1B of Howard (2009b). A method was developed to estimate the time to the common ancestor (TCA) of a cluster from the values of RCC of pairs of cluster members[1]. In addition, a method to estimate the time to the (earlier) common ancestor of a pair of surname clusters (i.e., the TCA of the two cluster CAs) was developed using the averages of the RCC of the cluster members in the intercluster region. This section extends the rationale of Figure 5 of Howard (2009b). It starts with a schematic evolutionary diagram of a hypothetical group of surname clusters, A, B, …. F shown in the following figure.

Figure 1: RCC Values of the Intersections of Six Hypothetical Surname Clusters (Upper Graph a) and How the RCC Values Appear in the RCC Intercluster Matrix (Lower Matrix b)

The figure contains a schematic evolutionary plot (a) labeled "SCHEMATIC EVOLUTIONARY PLOT - 6 hypothetical surname clusters -" with RCC values from 0 to 50, and nodes AF, CF, DF, AB, DE, and clusters A, B, C, D, E, F. Below it (b) is a distance matrix table:

		A	B	C	D	E	F
	A		20	50	50	50	50
	B	20		50	50	50	50
b	C	50	50		40	40	40
	D	50	50	40		20	30
	E	50	50	40	20		30
	F	50	50	40	30	30	

Consider the schematic evolutionary upper plot (a). The RCCs of the TMRCA of surname Clusters A, B, …. F are plotted at the bottom of the diagram in the interval between RCC 0 and 10. The earliest CA pair is AF at RCC=50 where the MRCA of Cluster A shares a MRCA with Cluster F. The common ancestor at AF has the "starting" haplotype, which then experiences marker changes as its lines evolve down the diagram on various evolutionary paths from RCC=50 to the present.

The downward connecting arrows in the diagram show the evolution of the clusters. Although AF is the oldest of the paired interclusters, and the lines of both Cluster A and Cluster F can be traced backward in time from the bottom of the diagram up to AF, we cannot tell which of the two clusters is the older since the schematic contains no information that would indicate their relative ages.

The CA of each pair of clusters is given on the graph as AB, DE, …. Clusters A and B will each have different most recent common ancestors. The common ancestor of those two common ancestors will appear at AB, etc. This is the MRCA of the Intercluster AB. Its haplotype will be the "starting" haplotype for the two lines that evolve to Clusters A and B. We have arbitrarily plotted the intercluster CAs of AB and DE at RCC = 20. The CA of Clusters D and F will appear at DF, and so on up the graph.

As time passes, the evolutionary lines (viz., the varying haplotypes) to Clusters A and F evolve down the timelines to the left and to the right, respectively. As this evolution takes place, the line to Cluster A develops a Subcluster B which eventually produces a progenitor at AB who is the common ancestor of the CA of both Cluster A and Cluster B. While that evolution takes place, the line to Cluster F spins off the Cluster C progenitor (at RCC=40), then Cluster D's progenitor (at RCC=30). The line to Cluster C then evolves directly to the present, while the line toward Cluster E spins off Cluster D (at RCC=20). From the figure, the rank order of surname cluster age, from oldest to youngest, is A&F, C, D, and finally B&E.

This schematic evolutionary sequence is mapped into the intercluster matrix in Figure 1 (b). The boxed intersections AF, CF, DF,….. AB and DE have the RCCs indicated at the row and column intersections of A and F, D and F, etc. in the graph. The entries for the remaining intersections are the RCC values where the two cluster lines intersect. Thus, for example, the entries for intersections AC and AD are both at the intersection AF (at RCC=50) and the entries for intersections CE and EF are both at the intersections CF and DF, respectively.

The challenge, described in the next section, is to turn the analysis around, and derive the upper evolutionary plot in Figure 1 from an RCC matrix that results from a correlation of pairs of Y-DNA testee results.

B. Derivation and Analysis of RCC Matrix Parameters for Gordon Clusters

In April 2010, the cutoff date of this analysis, 242 Y-DNA results were available in the Gordon surname project[2]. We selected only those results where testees had been tested at 37 or more markers and we use only the first 37 markers to form the correlation matrix and then the RCC matrix (Howard 2009a).

This process narrowed the analysis to 187 individuals from which we were able to group 119 testees (64%) into well-defined Gordon clusters and subclusters (viz., clusters within a cluster) in the RCC matrix. Of this RCC grouping, 104 (87%) were later matched with one of ten specific Gordon pedigree lines of which one category was "ungrouped". Pedigree lines for the remaining 68 testees were available, but they were not assigned to specific Gordon clusters using the RCC matrix.

B1. The Histogram of the RCC Matrix for All Gordon Testees

The first step in an RCC matrix analysis is to assess the distribution of surname pairs. Figure 2 presents a histogram of the entire RCC Gordon matrix before they were grouped into clusters.

Figure 2: Gordon Surname Histogram (187 individuals, each with 37 markers tested)

Histogram of All Gordon Surname Pairs
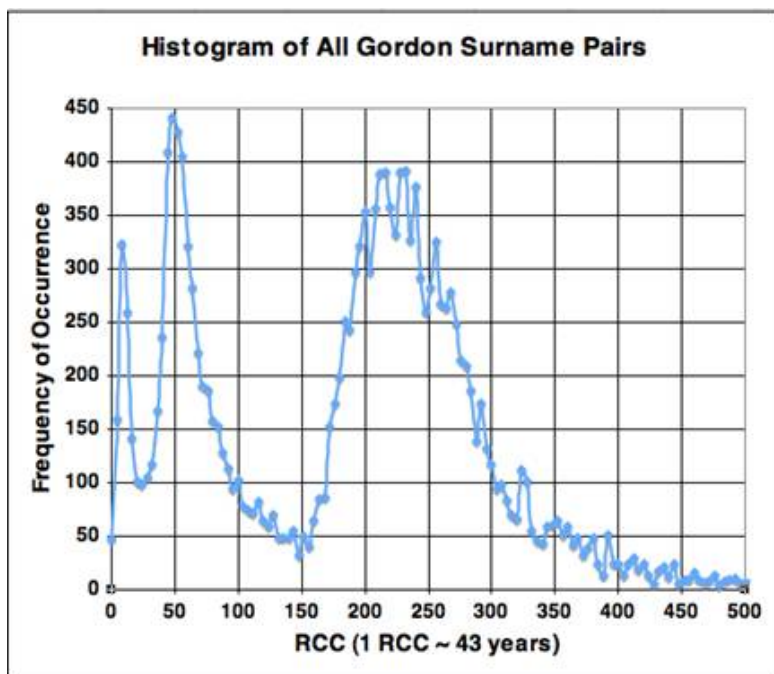
This histogram shows three prominent peaks. The first peak results from pairs of cluster members in subclusters and clusters; each pair is in the same haplogroup, but different haplogroups are present. The second peak is composed of pairs of testees who are either in different clusters or who are not in a cluster, but who are in the same haplogroup. The third peak is composed of pairs of testees who belong to different haplogroups.

B2. Identification of Gordon Clusters from the Gordon RCC Matrix

The clusters we found ran the gauntlet from sparse to well-populated. Since we wanted to choose a number of well-populated examples to afford us good statistical samples, we chose to study only major clusters, which we defined as an RCC grouping that must contain at least four different testees so that at least 6 pairs ((4 x 3)/2 = 6) would be available for comparison. This process led to the identification of a reasonable sample --10 major Gordon Clusters, A, C, D, and E (in Haplogroup I1), H, K, L, Q and T (in Haplogroup R1b1b2) and Cluster G (in Haplogroup I2b1). The members of each major Gordon cluster are given in Appendix B.

After averaging the RCCs of the individual members of each cluster and each intercluster region, and after determining the standard deviations (SD) of their means, we get the results given in Table 1. Entries along the diagonal show the average RCC of each Gordon cluster and the SD of that average. The averages of intercluster pairs are listed in their intercluster intersections above the diagonal and the SDs of their means are listed at the appropriate intersection below the diagonal.

TABLE 1: Average Values of RCC for Gordon Clusters and Interclusters, Identified by Haplogroup, and their Standard Deviations

	I1	I1	I1	I1	R1b1b2	R1b1b2	R1b1b2	R1b1b2	R1b1b2	I21b
	A	E	D	C	L	Q	H	T	K	G
A	11.3(0.25)	48.6	52.3	47.1	191.2	214.9	190.7	248.5	277.4	267.0
E	0.54	16.8(2.83)	46.2	53.1	208.9	215.4	208.5	257.7	292.9	226.4
D	0.43	1.51	7.5(0.78)	65.3	187.4	175.7	180.3	223.8	274.7	243.8
C	0.62	1.33	1.16	6.6(1.16)	190.2	192.3	197.9	239.1	249.9	219.6
L	1.04	3.48	2.58	2.68	18.4(2.65)	44.7	39.4	50.4	89.3	351.9
Q	0.76	1.77	1.90	1.34	1.72	5.3(0.63)	50.3	54.9	91.9	291.5
H	0.98	2.89	2.94	3.02	2.60	0.96	14.6(3.12)	51.6	123.0	403.5
T	1.39	4.14	4.02	3.69	3.17	0.97	1.84	20.5(2.04)	115.9	406.0
K	1.33	4.44	2.64	2.07	3.36	1.12	2.14	3.56	8.1(0.51)	344.1
G	1.01	2.58	2.33	1.07	4.44	0.97	2.68	1.93	2.45	5.4(1.27)

Gordon Cluster A contains 47 testees, resulting in (47*46)/2= 1081 testee pairs. This cluster is a special case since it contains subclusters, Aa, Ab, ….. Af. Most of these subclusters are sparsely populated. Subclusters are important because the RCC values of members paired with other members of the subcluster indicate that their TMRCAs probably fall within the time range of available pedigrees.

While this study concentrates on the major clusters, we kept the subclusters because the RCC values of their intercluster intersections with the major clusters might give us additional insight into cluster evolution. Further information on these Gordon subclusters, clusters, and interclusters can be found in Appendix B.

C. Locating the Points in the RCC Matrix that Share Identical Common Ancestors

In the schematic RCC matrix (Figure 1b) there are points that share identical common ancestors (e.g., the common ancestor at CF and RCC=40 is the same for CD and CE). By inspection, we recognize in Table 1 that many interclusters have average values of RCC that are nearly identical. Those points are the leading candidates where the common ancestors of Gordon clusters are shared. Scarcity of data causes uncertainties in those averages, which often vary greatly, producing problems of interpretation. To meet the challenge of mapping the results of Table 1 into an evolutionary diagram, we must identify those junction points – the times when the progenitor of a new cluster line was formed from an existing cluster line.

There are three haplogroups to which the main Gordon clusters belong. In Table 1 they are identified by different colors. The two inner boxes in the table contain the clusters in Haplogroup I1 and R1b1b2. Only one major cluster, Gordon G, was identified in Haplogroup I2b1.

To simplify forming an evolutionary diagram, we treat each of the three haplogroups separately. Figure 3 shows plots of the average value of RCC for each of the intercluster CAs for Haplogroups I1 and R1b1b2, taken from Table 1. The SD of the mean of each point is given by the error bars.

Figure 3: The Average Value of RCC for Gordon Intercluster Pairs in Haplogroups I1 and R1b1b2.
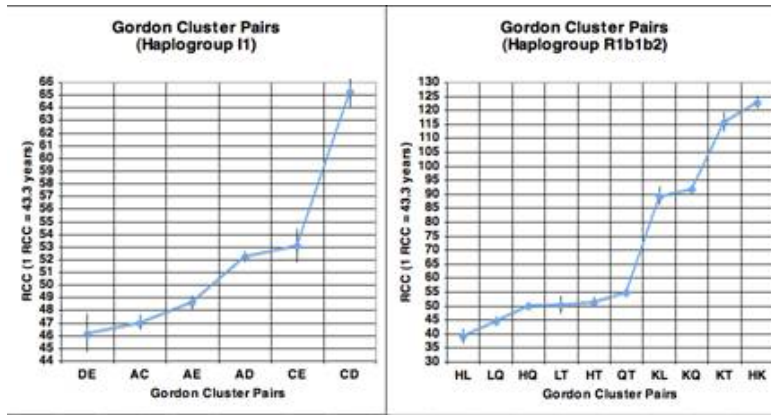
These two plots strongly suggest that the common ancestors of several interclusters lived at the same time[7]. Table 2 gives the details.

Table 2: Common Ancestor Locations of Gordon Intercluster Pairs[1]

Haplogroup	Intercluster Pair	RCC	SD	Years Ago	DATE (BCE)
R1b1b2	HK & KT	121.1	1.8	5250	3300
R1b1b2	KL & KQ	91.7	1.1	4000	2000
R1b1b2	HQ, LT, HT & QT	52.3	0.6	2250	300
I1	AD & CE	52.4	0.4	2250	300
I1	AC, AE, & DE	47.8	0.4	2100	130

D. Gordon Evolutionary Diagrams

Within each haplogroup we start with the oldest pairs of clusters that appear in the intercluster regions of Table 1. We plot the pairs from the oldest to the youngest in time, taking into account the locations of the Gordon intercluster pairs in Table 2. Each time a new cluster appears in the evolutionary track, we interpret it to mean that one of the original pairs has split off, producing the progenitor of a new line, which then proceeds to evolve, through mutations, down the evolutionary diagram to the present. The results are presented in Figures 4A, 4B and 4C. There are two RCC breaks in Figure 4A and one in Figure 4B. The dates below and above the lower breaks use the factors 43.3 and 52.7 in computing the dates for the common ancestors of the clusters and interclusters, respectively (Howard 2009a). The upper break in Figure 4A reduces the space needed to present the figure.

Figure 4A and 4B: The Evolutionary Diagrams of the Major Gordon Surname Clusters and Interclusters in Haplogroups I1 and R1b1b2
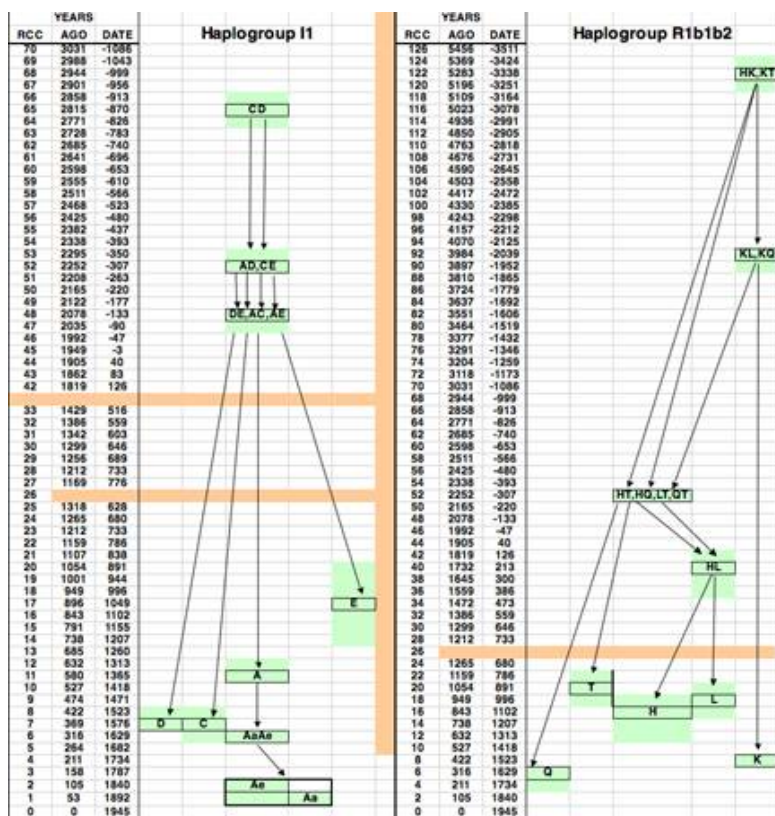
Haplogroup I1:

The intercluster value of Gordon Clusters C and D appear at RCC ~ 65, indicating that they shared a common ancestor (CA) about 2800 years ago. The SDs of these age estimates are indicated by the green zones around each plotted point. We do not know which cluster is older; we know only the location of their CA. Clusters C and D evolved into Clusters A and E at approximately the same time, 2300 years ago. From that time, the Gordon Clusters E, A, D and C evolved separately, with the TMRCAs of those clusters appearing in about 1050, 1360, and 1580 CE, respectively. Cluster A has formed interclusters among which are Aa and Ae. Their TMRCA as an intercluster lived about 1630, and the individual subclusters have CAs that are much more recent, in the 19th century.

Haplogroup R1b1b2:

There are three Gordon interclusters H, K and T that have a CA who lived about 5000 years ago (RCC ~ 118). Again, we do not know which cluster is the oldest. Gordon Clusters L and Q evolve along Cluster K's evolutionary line and they appear at RCC ~ 92, 4000 years ago. Cluster K then evolves directly to its CA at about 1500 CE. The earlier evolutionary lines of H, K, L and Q evolve to the shared intercluster locations of HT, HQ, LT and QT about 2250 years ago, or about 300 BCE. From there, the lines H and L evolve to RCC ~ 40, 1700 years ago, when the CA of Intercluster HL lived. From there the two lines evolve to their own CAs who lived near 1000 CE. The lines to Clusters Q and T evolve from their common intercluster ancestors at RCC~52 to their own CAs in the 17th and 8th centuries, respectively.

Major Branching within Haplogroups I1 and R1b:

Figures 4A and 4B show that major branching occurred in both Gordon haplogroups at RCC = 52, about 2250 years ago. This date, near 333 BCE, was a tumultuous time when Alexander first invaded Gordion (near Ankara, Turkey) in Anatolia, then Gaul from Thrace, followed by the Romans.
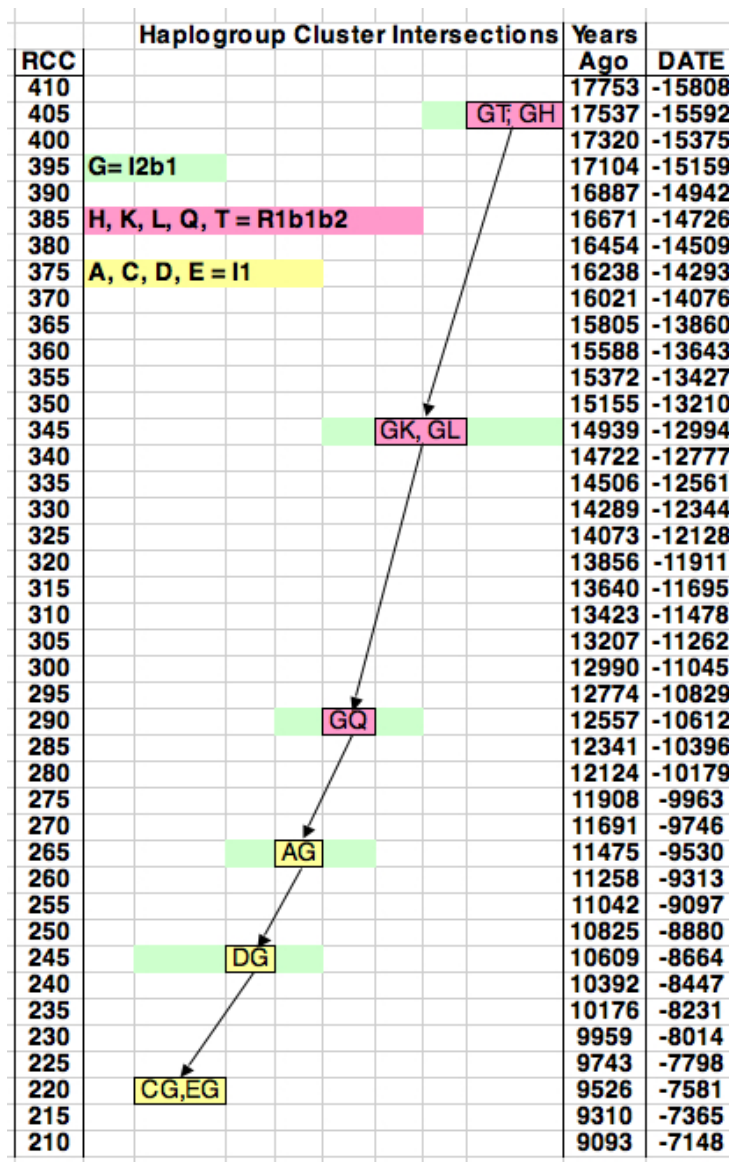
Between 800 and 1000 CE R1b1 seems to encounter considerable branching, between the times of Charlemagne and William the Conqueror, corresponding to tumultuous times in French, English, and Scottish history.

Given the relative small number (~1%) of I1 in Anatolia, the probability increases that origins of the paternal Gordon ancestors in Macedonia and present-day Turkey are R1b1 (Cinnioğlu et al, 2004).

Haplogroup I2b1 (Gordon Cluster G):
The intercluster regions between Haplogroups R1b1b2 and I2b1 occur much earlier in time. In Figure 4C we plot all the positions of the three haplogroup intersections with Gordon Cluster G.

Figure 4C: The Evolutionary Diagram of the Major Gordon Surname Cluster G (Haplogroup I2b1) and Its Intersections with Other Gordon Cluster Haplogroups

RCC	Haplogroup Cluster Intersections	Years Ago	DATE
410		17753	-15808
405	GT, GH	17537	-15592
400		17320	-15375
395	G= I2b1	17104	-15159
390		16887	-14942
385	H, K, L, Q, T = R1b1b2	16671	-14726
380		16454	-14509
375	A, C, D, E = I1	16238	-14293
370		16021	-14076
365		15805	-13860
360		15588	-13643
355		15372	-13427
350		15155	-13210
345	GK, GL	14939	-12994
340		14722	-12777
335		14506	-12561
330		14289	-12344
325		14073	-12128
320		13856	-11911
315		13640	-11695
310		13423	-11478
305		13207	-11262
300		12990	-11045
295		12774	-10829
290	GQ	12557	-10612
285		12341	-10396
280		12124	-10179
275		11908	-9963
270		11691	-9746
265	AG	11475	-9530
260		11258	-9313
255		11042	-9097
250		10825	-8880
245	DG	10609	-8664
240		10392	-8447
235		10176	-8231
230		9959	-8014
225		9743	-7798
220	CG,EG	9526	-7581
215		9310	-7365
210		9093	-7148

The common ancestor of the interclusters of Gordon G, H and T is located at RCC ~ 405, or about 17,500 years ago. Since Clusters H and T (red in the figure) are in Haplogroup R1b1b2, these two haplogroups had a common ancestor at least this far in the distant past. As evolution took place, the CAs of the interclusters of Gordon Clusters G, L and K appeared at RCCs ~ 345, or 15,000 years ago. Later, the CAs of the interclusters of Gordon Clusters G and Q appeared at RCCs ~ 290, or 12,500 years ago. As the line to Cluster G evolved, it intersected with Haplogroup I1 (yellow in the figure) where it shared a common ancestor with Cluster A at RCC ~265 (9500 BCE), with Cluster D at RCC ~ 245 (8600 BCE), and with shared CAs of Clusters C and E at RCC ~220 (7600 BCE)[1].

E. Time Relationships Between Gordon Haplogroups I1 and R1b1b2

Figure 5: RCC-Derived TMRCAs Among Gordon Clusters of Haplogroups I1 and R1b1b2
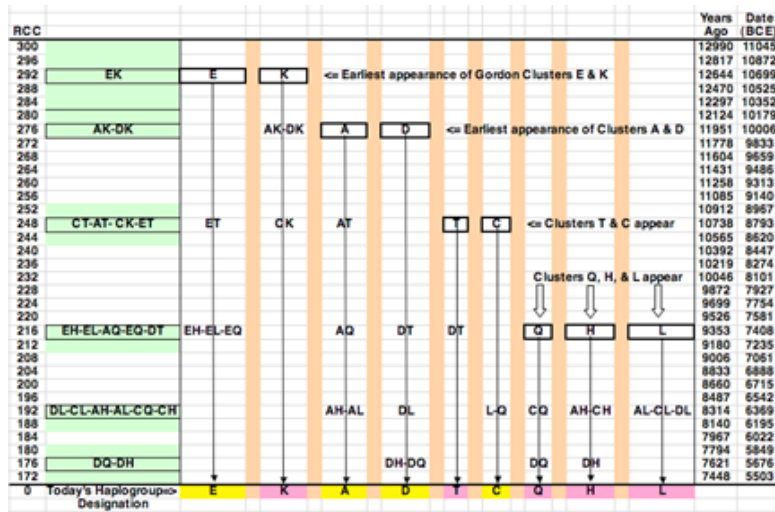
Time Relationships Derived from the Gordon RCC Matrix Results

Insight into the evolutionary relationships among the Gordon clusters of Haplotypes I1 and R1b1b2 can be gained from a study of Figure 5. In the figure, the RCC time scale is given at the far left of the diagram. The next column lists pairs of Gordon clusters that belong to different haplogroups. The bottom row in the Figure gives today's haplogroup designation of the Gordon clusters E, K, A…..L. Haplogroup I1 clusters are colored yellow; Haplogroup R1b1b2 clusters are colored red.

At the top left of the figure, the intersection EK at RCC~292 is where the present Gordon Clusters E and K shared a most recent common ancestor, 12,600 years ago. This is the earliest common ancestor resulting from a pairing of a cluster in Haplogroup I1 with a cluster in Haplogroup R1b1b2. Their common ancestor's date is determined by finding the average RCC of the intercluster region between Gordon Clusters E and K.

At RCC~292 that common ancestor has the progenitor haplotype of what will be Clusters E and K, and an assignment of a haplogroup to them at that time would be meaningless. Only as their haplotypes begin to evolve downward in the diagram to the present time do their haplogroup assignments become meaningful.

Evolution has separated the cluster pairings into six distinct RCC intervals – at RCC~ 292, 276, 248, 216, 192 and 176. The green zones surrounding each set of cluster pairs represent two standard deviation error bars and the figure shows that most of the clustered pairings are in distinct groups. For example, the common ancestor of the cluster pairs of Gordon CT, AT, CK and ET all share the same haplotype at RCC~248, about 10700 years ago, or 8800 BCE on the corresponding date scale at the far right of the figure. The four sigma range of uncertainty in RCC for these pairs goes from 244-252, or (252-244) x 43.3= 350 years[1].

The intersections of each Gordon cluster with a cluster in a different haplogroup are given in the center part of the figure. For example, Gordon Clusters E and K share a common ancestor at RCC~292 having the same 37-marker haplotype. As Cluster E evolves, its haplotype mutates in such a way that at RCC~248, it shares a common haplotype with Cluster T and at RCC~ 216 its haplotype is the same as Clusters H, L and Q. The other vertical columns represent the evolution of Clusters K, A, D, etc.

Another example of the evolutionary sequence is shown the vertical column beginning with Cluster K. It starts with a common ancestor with Cluster E, but evolves so that at RCC~276 it shares a common ancestor with Clusters A and D. Cluster K then evolves so that it shares a common haplotype with Cluster C.

Figure 5 suggests that when Cluster K evolves to RCC~276, Clusters A and D form, since the most recent common ancestor of AK and DK are the same and the Gordon A and D clusters are not evident in the data earlier in time. Then, as Clusters K and A evolve, they spin off Cluster C and Cluster T, respectively. This activity occurs at RCC~248, at which point Clusters T and C share a common ancestor and haplotype and evolve.

Further insight into the evolution of these Gordon clusters come from the earliest dates that they appear paired with another cluster in Figure 5 where they appear in a box. Table 3 shows the cluster pairing. Clusters that are underlined and in a larger font appear at the earliest dates of the cluster pairs in the Gordon clusters found in this work.

Table 3: Earliest Dates Found for the Common Ancestors of Pairs of Gordon Clusters

Cluster Pair	Earliest Paired RCC	Corresponding Years Ago	Corresponding Date (BCE)
E-K	292.9	12700	10700
AK-DK	276.8	12000	10000
CT-AT-CK-ET	248.7	10800	8800
EH-EL-AQ-EQ-DT	214.7	9300	7350
DL-CL-AH-AL-CQ-CH	191.2	8300	6300
DQ-DH	177.0	7650	5700

F. The Subclusters and Intersubclusters in Gordon Cluster A

Just as clusters and interclusters in the RCC matrix give us insight into the evolution of a surname, so do subclusters and their intersections, the intersubclusters. The RCCs of subclusters are very small, so the probability of linking them to testee pedigrees is high. Dictating against success, however, is that unknown mutations at low RCCs cause major uncertainties in discovering such links. Nevertheless, in this Section we will investigate what information might result from the results of the six subclusters a, b, …..f, within the major Gordon A surname cluster.

Subclusters a and e contained 55 and 15 pairs of testees, but Subclusters b, c, d and f contained only one pair. However, intersubcluster pairs for the latter four subclusters give an indication of how the lines evolved, so we retained them in the study. The approach for subclusters will be the same as that applied in Section C to clusters. Table 4 summarizes these data for the subclusters and their intersections within the main Gordon A cluster.

TABLE 4: Average Values of RCC for the Gordon A Subclusters and Intersubclusters, and their Standard Deviations

	a	b	c	d	e	f
a	1.1(0.1)	3.1	3.7	4.1	5.9	6.3
b	0.1	0(0.6)	5.6	6.1	9.1	8.7
c	0.2	0.6	0(0.6)	7.4	8.3	11.2
d	0.2	0.6	0.6	0(0.6)	9.9	6.2
e	0.3	0.3	0.3	1.0	1.6(0.4)	14.1
f	0.6	0.6	0.3	3.1	1.1	2.4(0.6)

The RCC of the most recent common ancestor of each intersubcluster appears in the diagonal of Table 4 along with the SD of the mean in parentheses. The averages of intersubcluster pairs are listed in their intersections above the diagonal and the SDs of their means are listed at the appropriate intersection below the diagonal. Yellow entries for the SDs (0.6) are estimates based on the average SD of other entries.

To simplify forming the evolutionary diagram, Figure 6, derived from Table 4, can be used to identify, by both inspection of the error bars and analysis, the pairs of intersubclusters that overlapped in time. They are (1) ab, ac and ad; (2) bc, ae, bd, df and af; and (3) ce, bf, be and de. The SD of the mean of each point is given by the error bars in Figure 6.

Figure 6: The Average Value of RCC for the Major Gordon A Subcluster and Intersubcluster Pairs
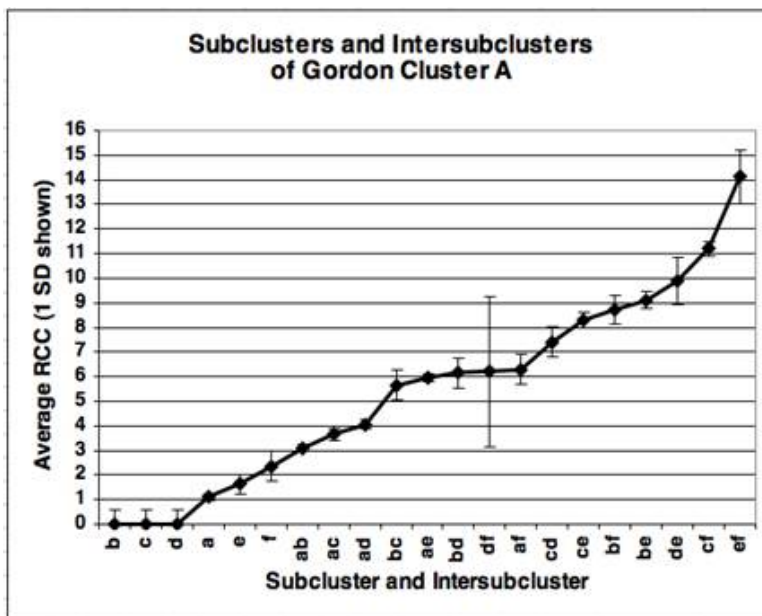
Within each intersubcluster we start with the oldest pairs that appear in the intersubcluster regions of Table 4. We plot the pairs from the oldest to the youngest in time, taking into account the locations of the Gordon intersubcluster pairs in Table 4. Each time a new subcluster appears in the evolutionary track, we interpret it to mean that one of the original pairs has split off the progenitor of a new line, which

then proceeds to evolve, through mutations, down the evolutionary diagram to the present. The results are presented in Figure 7.

Figure 7: The Evolutionary Diagram of the Subclusters and Intersubclusters of the Major Gordon ClusterA (Haplogroup I1)
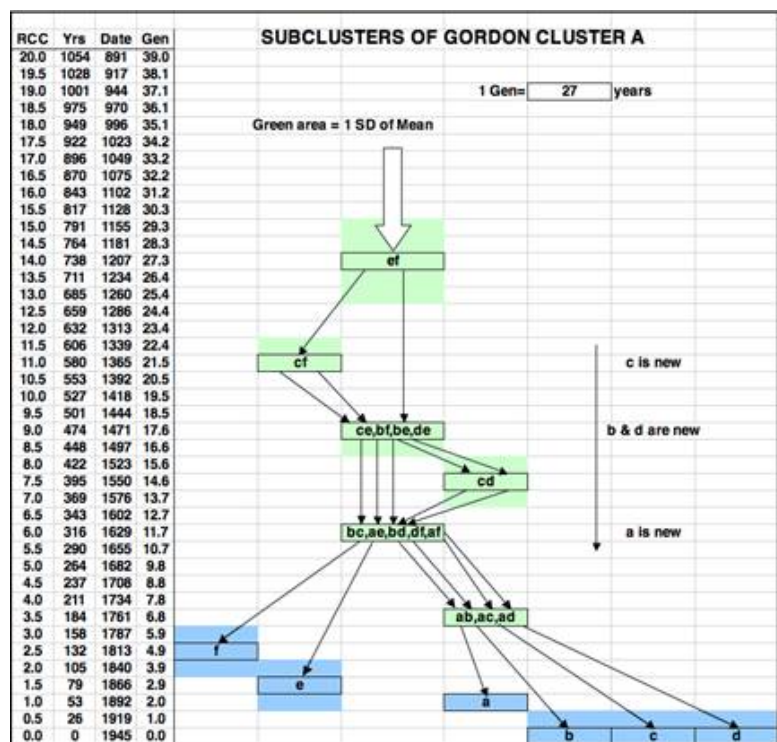
The columns to the left in Figure 7 indicate the RCC value of the common ancestor subcluster and intersubcluster intersections, followed by estimates of the number of years ago (from 1945, the assumed average birth year of the testees), the corresponding date, and the number of generations (assuming 27 years per generation) when the events to the right occurred[1,3].

- The common ancestors of each of the six subclusters (RCC<~3) probably lived within the time period of interest to genealogists. The evolution of those subclusters apparently starts with the lines leading to intersubcluster ef (at RCC~14) almost 30 generations ago. From the common ancestor of subclusters e and f, two Gordon surname lines originate (line e and line f). Line f first evolves to the intersubcluster cf where the earliest common ancestor of line c originated, at RCC~11, about 600 years ago. From the intersubcluster ef, line e evolves, along with lines c and f, to the intersubcluster at RCC~9, where not only line a originates, but where the lines to bc, bd and df have a shared common ancestor at RCC~6. From that point, lines e and f evolve directly to the CA of subclusters e and f. From the shared ancestor at RCC~6, lines a, b, c, and d evolve to RCC~3.5 where ab, ac and ad have a shared common ancestor. From that point, these four lines evolve independently to the individual common ancestors of subclusters a, b, c and d.

- Intersubcluster ef is a nice fit with generally-accepted Gordon tradition. See later chronology.
- Intersubclusters ab, ac, ad appear during the first generation after the last Jacobite Rebellions in 1745.
- Kit 7805 is the descendant of John George Gordon (b. ca. 1700 in Ireland) who first came to America before about 1725. He is in intersubcluster af (~1629), and it is interesting to note the considerable confluence of clusters a generation or two after this date. The previous major intercluster occurs around 1470, a couple of generations after the Jock & Tom Gordon groups branched.
- Intersubcluster cf could represent the Sir William Gordon branch. He was born after 1320. Intersubcluster cf is well documented in the lowland borderlands.
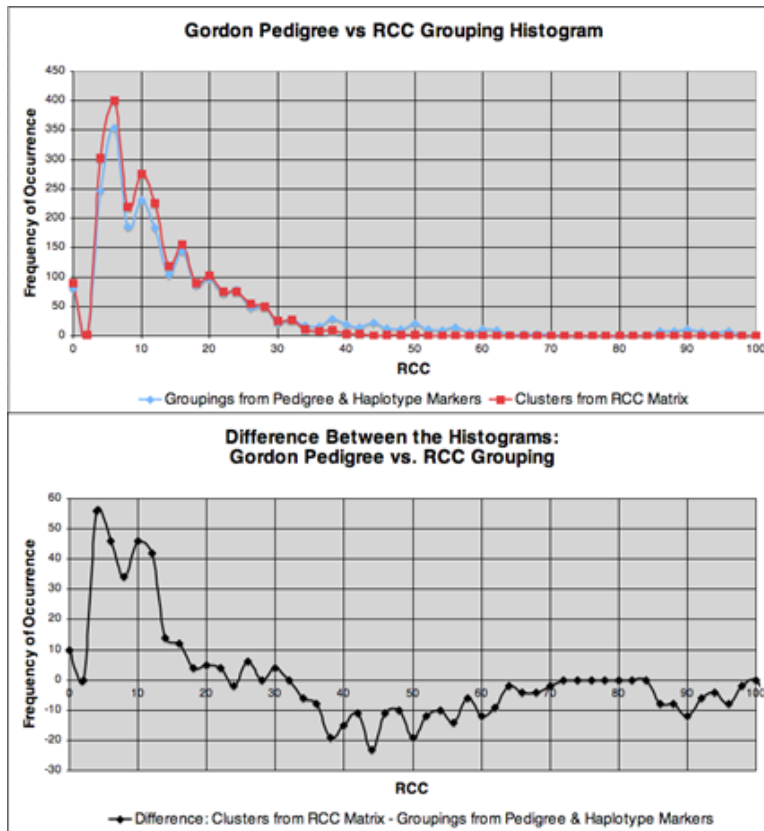
Figure 7 indicates that the suggested evolutions of these subclusters took place after the invasion of England by William the Conqueror in 1066. Since pedigrees sometimes extend back that far, the presence of subclusters may offer valuable insights into the formation of individual surname lines within periods of time covered by pedigrees. We must still stress that the presence of non-average mutations introduce uncertainties that might amount to several hundreds of years, so these lines must be approached with caution[1]. As more Gordons are tested, these subclusters, hence their evolutionary relationships, will become better defined. At this time, we can probably trust only the results that separate the oldest from the youngest subclusters.

The members of these major Gordon Cluster A subclusters (Aa through Af) are among those listed in Appendix B. Other members of Gordon A do not appear to be members of a subcluster.

A COMPARISON OF INDEPENDENT GORDON SURNAME GROUPINGS DONE BY PEDIGREES AND Y-DNA MARKER VALUES AND DONE BY RCC MATRIX ANALYSIS

Gordon surname groupings were made independently by one of us (Gordon) in the traditional way, using Y-DNA results and comparing them with pedigrees, when known, and by one of us (Howard) by sorting the RCC matrix so that small values of RCC appeared in different Gordon clusters. The following two histograms in Figure 8 show a comparison of the results from both methods (top), and the difference between the results of the two methods (bottom). Only pairs of testees are included who were matched by each method; no unmatched or ungrouped testees appear in the histograms. No changes in matchings or groupings were made by either author prior to deriving Figure 8.

Figure 8: Histograms of the Frequency of Occurrence of RCC values for Gordon Groupings from Pedigree and Haplotype Markers and the Difference Between the Two Methods

Gordon Pedigree vs RCC Grouping Histogram

Groupings from Pedigree & Haplotype Markers — Clusters from RCC Matrix

Difference Between the Histograms:
Gordon Pedigree vs. RCC Grouping

Difference: Clusters from RCC Matrix - Groupings from Pedigree & Haplotype Markers

The comparison histogram of the two methods is remarkably similar. The total number of pairs of testees in the pedigree and RCC matrix groups are 2352 and 2336 pairs, respectively, showing that both methods make approximately the same groupings using the same sample of testees. It is evident from the difference histogram that more groupings fall into RCC-derived Gordon clusters between RCC 0-20 than in the pedigree groupings. Moreover, the traditional approach includes some testees in groupings that the RCC method does not include in a cluster, and those inclusions, when paired, result in values of RCC between 30 and 70 and between 85 and 100. RCC values between 30-70 are more typical of intercluster relationships between those pairs of testees. Fewer pairs fall between 85-100.

Figure 9 shows the difference in the cumulative distribution of the two approaches. It shows that 95 per cent of testees have been included in matrix-derived clusters when their RCC ~ 28 whereas that same percentage of pedigree groupings is not achieved until their RCC ~ 60. These results are consistent with the contention (Howard 2009a) that the use of both methods will yield better results than using the traditional method alone.

Figure 9: The Cumulative Percentage of Gordon Testees Who Are Grouped by Haplotype Matching and Pedigrees and by RCC Matrix Matching as a Function of RCC

After comparing the groupings based on haplotypes and pedigrees with the clusters derived from the RCC matrix, it became evident that some of the pedigrees designations were not correct or could be refined. In Table 5 we show the original pedigree designation as well as the final name of the pedigree-cluster association and we will refer to the final designations in the rest of this paper.

Appendix B identifies, by Kit Number, the Gordon clusters derived from the Gordon RCC Matrix and the Gordon lines derived from pedigrees and traditional marker analysis. The statistics resulting from Appendix B are shown in Table 5.

Table 5: The Distribution of Gordon Testees Among Gordon Pedigree Lines and Gordon Clusters Derived from the Gordon RCC Matrix

Original Name	Final Name	←———— I1 ————→										I2b1	←	R1b1b2	→					SUM
		A	Aa	Ab	Ac	Ad	Ae	Af	C	D	E	G	H	K	L	N	O	Q	T	SUM
GORDON SEPTS																	2		1	3
Jock & Tam Gordons		21	10		2	2	6	2												43
SETON-GORDON BRANCH													4	4		1		7		16
Sir William Gordon Branch 1	Sir William Gordon Branch		1	3																4
Sir William Gordon Branch 2	Progenitor Branch									7	6									13
Sir William Gordon Branch 3	SMALL GROUPING-GORDONS – Subgroup 10																		4	4
SMALL GROUPING-GORDONS – Stewart-Gordons												7								7
SMALL GROUPING-GORDONS – Subgroup 0															4		1			5
SMALL GROUPING-GORDONS – Subgroup 2									5											5
SMALL GROUPING-GORDONS – Subgroup 8																2				2
UNGROUPED – GORDONS													1			1				2
	SUM:	21	11	3	2	2	6	2	5	7	6	7	5	4	4	4	3	7	5	104

OBSERVATIONS BASED ON <u>CLUSTER</u> GROUPINGS (From Table 5)
- The individual Gordon Clusters belonging to Haplogroups I1, I2b1 or R1b1b2 have no overlap in membership.
- All 43 testees in the Jock and Tam Gordon pedigree lines belong only to Gordon Cluster A and its subclusters.
- The 47 members of Gordon Cluster A are distributed only in the Jock and Tam Gordon and the Sir William Gordon Branch. Gordon Cluster A also contains all 4 testees in the Sir William Gordon Branch. The subcluster Gordon Aa partially overlaps these two lines but its one member may be a transition between subclusters Aa and Ab, or may actually belong to Ab.
- Gordon subclusters Ac, Ad, Ae and Af only appear in the pedigree line of Jock and Tam Gordons.

- These observations indicate that when more Y-DNA test results become available, the RCC matrix will show more subclusters for the major Gordon clusters, which will, in turn, contain more members, just like Cluster A.
- The close association of the Sir William Gordon Branch with members of Cluster A indicates a shared recent common ancestor.
- The 16 testees in the Seton-Gordon pedigree line are distributed among four separate Gordon Clusters (H, K, O and Q) in Haplogroup R1b1b2.
- The 13 testees in the Progenitor Branch are evenly distributed in Gordon Clusters D and E in Haplogroup I1.
- The 4 testees in the pedigree line Small Grouping-Gordons - Subgroup 10 are the only members of, and hence define, Gordon Cluster T in Haplogroup R1b1b2.
- The 7 testees in the Stewart-Gordon line are all located in Gordon Cluster G in Haplogroup I2b1.
- The same 5 members of Gordon Cluster C belong only to Subgroup 2 of the Small Gordon Groupings.
- Gordon Cluster O and Subgroup 8 of the Small Gordon Groupings line each contain only the same 2 members.

One of the goals of this analysis is to see how closely members of particular pedigree lines could be placed in a Gordon cluster derived from the values of paired testees in the RCC matrix.

Table 5 shows the Gordon clusters derived from Y-DNA results that have pedigrees. There were 68 testees who had pedigrees who were not assigned to RCC clusters. Of those 68, only one had a value of RCC less than 20, 28% had an RCC under 25, 44% had an RCC under 30, and 56% had an RCC under 34. We can therefore conclude that an RCC ~20 represents a practical limit for the identification of a Gordon cluster from the RCC matrix and that Table 5 indicates that they can be matched with available pedigrees. An RCC ~ 20 corresponds to ~ 900-1100 CE, a date consistent with the first use of surnames (Howard 2009a,b).

Of the 68 testees who had pedigrees but who were not assigned to a Gordon cluster, 121 pairs (2.6%) had values of RCC between 20 and 30, just over the practical cluster definition. Eight of these had a total of 76 RCC values between 20 and 30.

OBSERVATIONS BASED ON PEDIGREE GROUPINGS (From Table 5)
When relationships between pedigrees and subcluster membership are compared, they reveal totally new information that could have been derived only from this approach. The presence of subclusters was discovered in Howard 2009b, but the remarkable relationships among pedigrees, cluster and subcluster membership, geographical location, and their evolution has become apparent only through this study.

The Sir William Gordon Branch 2 has been renamed the "Progenitor Branch" of Haplogroup I1 because it is the earliest pair among the Haplogroup I1 major Gordon Clusters (D and E) (See Figure 5). Gordon Clusters D and E have a common ancestor at approximately RCC = 46.2 (about 2000 years ago) and their TMRCA with Gordon Subcluster Ab at RCC = 41 and 34, respectively. The Gordon subcluster Ab has an earlier TMRCA with Clusters D and E, approximately 1500 and 1800 years ago, respectively.

Second, the Sir William Branch 3 has been renamed "Small Grouping-Gordon – Subgroup 10." This group comprises Gordon Cluster T but does not fit either

documented history, or the I1 Haplogroup. Subgroup 10 may actually be more closely related to Seton-Gordons at an RCC of approximately 54 (2300 years ago).

It is noteworthy that subclusters Aa, Ab, Ad, Af and possibly Ae, all appear to have documentation connecting them to the Lowland Gordons. Moreover, subclusters Ab, Ad, and Af possibly have subcluster connections to the former larger southwestern region of Galloway. [Note: Tradition states that the original Lowland Gordon stronghold in the 11th century was in Berwickshire on the eastern seaboard]

It should also be noted that we do not have enough family information on Ae to determine where in Scotland they originated. Due to insufficient documentation, we were also unable to draw any Lowland connections for subcluster Ac.

The intersection of subcluster ef occurs during the origins of the Gordon family in Scotland.

The intersubcluster cf occurs about the time of Sir William Gordon's death at about 1370.

Subclusters ce, bf, be, de occur during a time of much upheaval in the Gordon family, when rival Gordon family factions were fighting for titles.

The intersection of subclusters bc, ae, bd, df, af at about 1629 occurs at the time when the two Gordon houses of Troquhain and Crogo in the South of Scotland came together through marriage of James Gordon of Troquhain and his cousin Janet Gordon of Crogo and Dalquharm, as well as the beginning of the house of Kenmure and continuation of Lochinvar. See:
http://www.thegordondnaproject.com/93333.html

CLUSTER A

No analysis is provided for Cluster A as a whole, due to complexities inherent in analyzing such a large group. The fact that we are unable to break it down into further subclusters may have one or more interpretations.

1.   Our sample is not large enough to identify further subclusters.
        With time and more testees, more subclusters will appear.
2.   Minimal mutations have yet to occur in the first 37-markers since the origin of Cluster A to permit identification of more subclusters.
        However, expanding RCC analysis to RCC 67-marker-based analysis may reveal further mutations, and thus, subclusters and insight.
3.   Subclusters may reflect a bias towards the higher number of testees from the outside the UK. This may be attributed the difficulties in recruiting testees from Western Europe, where DNA testing has yet to gain popularity as an extension of genealogical research.
4.   We found a very strong correlation between membership in a Y-DNA subcluster and membership in a pedigree group, indicating that if two testees share an RCC value of the order of 10 or less, then it is highly probable that they can be found in a pedigree group. Thus, using the RCC correlation technique, we have linked near-term genetics to a genealogical pedigree.

MERGING THE RESULTS OF GORDON CLUSTER GROUPINGS AND THEIR
ASSOCIATED DATES WITH THE RESULTS OF PEDIGREE/HAPLOTYPE GROUPINGS.
DATING THE GORDON PEDIGREE AND CLUSTER LINES.

Table 5 shows the distribution of Gordon testees among Gordon pedigree lines and
Gordon clusters derived from the Gordon RCC matrix. The evolutionary diagrams in
Figures 4, 5, and 7 show how these Gordon clusters evolved and give estimates of
when their most recent common cluster ancestor lived. When the results of Table 5
are convolved with the evolutionary diagrams of the same clusters, we can show the
various times when the TMRCAs of each pedigree line of testees lived. That
convolution is shown in Table 6.

Table 6: Date Groups Within Which the TMRCAs of Gordon Clusters Having Identified
Pedigrees Lived.

* See Footnote 1
The TMRCA calculation is found by multiplying the RCC of the TMRCA by 57.2, the
factor appropriate for the derivation of the TMRCA for a cluster (Howard 2009a).
Clusters in red belong to Haplogroup R1b1b2; clusters in yellow belong to
Haplogroup I1; the cluster in green belongs to Haplogroup I2b1.

OBSERVATIONS BASED ON CLUSTER DATES AND PEDIGREE LINES (From Table 6)

In this section, the Gordon Septs and the subclusters of Cluster A are not discussed
further because their association as Gordons (the Septs) are not proven and
because the dating of TMRCAs of subclusters have uncertainties that dominate the
dating process.

Of the 104 testees, 69 percent are located in clusters that have TMRCAs who lived
in the late 13th and early 14th centuries. Their RCC values cluster tightly between
RCC 10.2 and 12.6. We identify an "intermediate date group" that dominates the
data set. We note that these dates are estimates of the TMRCA of the Gordons
tested. They probably point to the times when their earliest identified ancestor
adopted the name Gordon. Each of these clusters has earlier ancestors, of course,
but the convergence in 1280-1410 is in agreement with other hypotheses about
when most of our early ancestors adopted their name. Unlike many other surnames
(e.g., Cook(e) or Cooper which were adopted from occupations that occurred
throughout Europe), names like Gordon, based on titles or places, are more tightly
grouped in location.

We identify an "older date group" containing 16 percent of the testees whose
TMRCAs lived before about 700 CE. They are all in Haplogroup R1b1b2 and have
RCC values that extend beyond the usual cluster boundary of ~20-25. It is unusual
for pedigree/cluster identifications to be made for testees in clusters whose TMRCAs
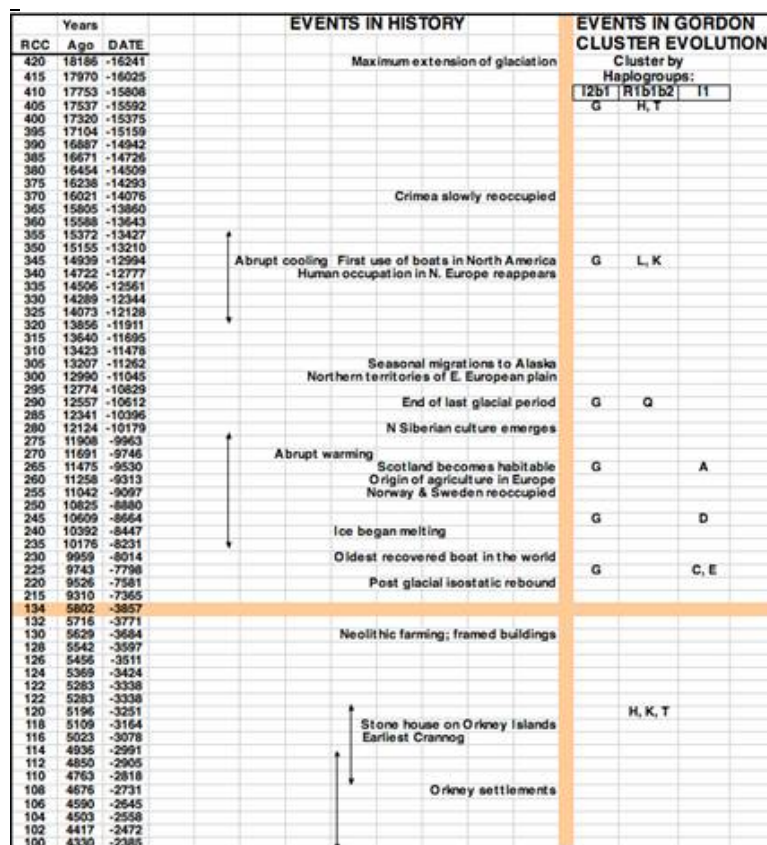are located so far back in time.

We identify a "younger date group" containing approximately the same percentage as members in the older date group. This group lies in a time interval where pedigrees are useful, but due to uncertainties in random mutations, the RCC time scale is not as useful when applied to this group. In fact, the presence of subclusters indicates only a close relationship among the testees in the subcluster, many of whom may know each other.

RELATIONSHIPS BETWEEN HISTORICAL EVENTS AND THE CHRONOLOGICAL EVOLUTION OF THE GORDON SURNAME

Events in Scottish and European History Compared with Events in the Evolution of Gordon RCC Clusters

The approximate ages of Gordon clusters and interclusters were derived in the previous sections. Figure 10A and 10B presents the events in history over the time intervals derived for the major Gordon clusters and interclusters.

Figure 10A: A Comparison of Events in the Evolution of Gordon Clusters and Events in European and Scottish History from the Maximum of the Last Glaciation to 2500 BCE

| Years | | | EVENTS IN HISTORY | EVENTS IN GORDON CLUSTER EVOLUTION |
RCC	Ago	DATE		
420	18186	-16241	Maximum extension of glaciation	Cluster by Haplogroups: I2b1 R1b1b2 I1
415	17970	-16025		G H, T
410	17753	-15808		
405	17537	-15592		
400	17320	-15375		
395	17104	-15159		
390	16887	-14942		
385	16671	-14726		
380	16454	-14509		
375	16238	-14293		
370	16021	-14076	Crimea slowly reoccupied	
365	15805	-13860		
360	15588	-13643		
355	15372	-13427		
350	15155	-13210		
345	14939	-12994	Abrupt cooling First use of boats in North America	G L, K
340	14722	-12777	Human occupation in N. Europe reappears	
335	14506	-12561		
330	14289	-12344		
325	14073	-12128		
320	13856	-11911		
315	13640	-11695		
310	13423	-11478		
305	13207	-11262	Seasonal migrations to Alaska	
300	12990	-11045	Northern territories of E. European plain	
295	12774	-10829		
290	12557	-10612	End of last glacial period	G Q
285	12341	-10396		
280	12124	-10179	N Siberian culture emerges	
275	11908	-9963		
270	11691	-9746	Abrupt warming	
265	11475	-9530	Scotland becomes habitable	G A
260	11258	-9313	Origin of agriculture in Europe	
255	11042	-9097	Norway & Sweden reoccupied	
250	10825	-8880		
245	10609	-8664		G D
240	10392	-8447	Ice began melting	
235	10176	-8231		
230	9959	-8014	Oldest recovered boat in the world	
225	9743	-7798		G C, E
220	9526	-7581	Post glacial isostatic rebound	
215	9310	-7365		
134	5802	-3857		
132	5716	-3771		
130	5629	-3684	Neolithic farming; framed buildings	
128	5542	-3597		
126	5456	-3511		
124	5369	-3424		
122	5283	-3338		
122	5283	-3338		
120	5196	-3251		H, K, T
118	5109	-3164	Stone house on Orkney Islands	
116	5023	-3078	Earliest Crannog	
114	4936	-2991		
112	4850	-2905		
110	4763	-2818		
108	4676	-2731	Orkney settlements	
106	4590	-2645		
104	4503	-2558		
102	4417	-2472		
100	4330	-2385		

The right hand side of Figure 10A lists the points in time when the ancestors of various pairs of Gordon interclusters lived. During this period, following Figures 4A, B, and C, there were only intersecting haplogroups. For example, Gordon G in Haplogroup I2b1 with Gordons H and T in Haplogroup R1b1b2 at RCC ~ 405 or

17,500 years ago. In other words, the shared common ancestor in those haplogroups had a 'beginning' haplotype that mutated down the lines of the Gordon G, H and T clusters.

The earliest Gordon haplogroup pairs appeared just after the time of the last glacial maximum in Europe. Then the pairings of Gordon Haplogroups G, L and K (L and K splitting from G) occurred when humans began to populate Europe as the glaciers melted. The common ancestor of Gordon G and Q lived at the end of the glacial period. When Scotland became habitable in about 9500 BCE, the common ancestor of Gordon A and Gordon G lived. This was the first appearance of paired Gordon clusters in Haplogroup I1. Gordon G and D had a joint CA at about 8700 BCE. The common ancestor of Gordon C and E had the beginning haplotype of their lines at about 7800 BCE when there was a post glacial isostatic rebound after the glaciers melted.

About 3700 BCE farming and framed buildings appeared. Gordon Clusters H, K and T shared a common ancestor haplotype when stone houses appeared in the Orkney Islands. By that time common Gordon ancestors had appeared within a single haplogroup, R1b1b2.

Figure 10B: A Comparison of Events in the Evolution of Gordon Clusters and Events in Scottish History Between 2000 BCE and the Present
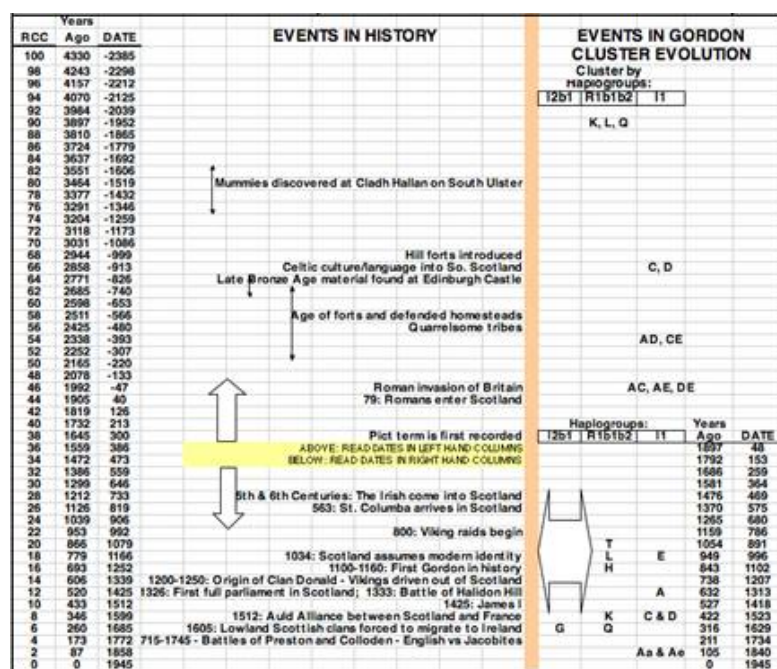
The first instance of a joint ancestor of major Gordon surname clusters (C and D) in Haplogroup I1 occurs about the year 1000 BCE when hill forts were first built, when the Celtic culture and language was introduced into Southern Scotland and when late bronze age material was being used at Edinburgh Castle.

Gordon Clusters A and E first appeared when their common ancestor-progenitors were paired as AD and AE near the year 400 BCE at about the time that tribes in Scotland became quarrelsome. The first intercluster progenitors that involved a

pairing of Clusters AC, AE, and DE lived at about the time of the Roman invasion of Britain and their entry into Scotland.

The Viking raids began in about 800 CE. At that time, or shortly afterwards, when Scotland was assuming its modern identity, the common ancestors of the currently defined Gordon Clusters lived, first those in Clusters T, L, and H in Haplogroup R1b1b2 and Cluster E in Haplogroup I1 in about 1000 CE, and next those in Cluster A about 600 years ago, followed by K, C and D, G and Q. The common ancestors of the subclusters of Gordon Cluster A lived more recently.

TIME ESTIMATE COMPARISONS:

Comparison of the Gordon RCC Time Estimates and ISOGG Time Estimates

Figure 4C indicates that at RCC~405 (17,500 years ago) the earliest pairs of Gordon clusters (GT and GH) in different haplogroups I1b1 and K had a most recent ancestor. Figure 5 indicates that at RCC~292 (12,600 years ago) the earliest pair of Gordon clusters (E and K) in different haplogroups I1 and R1b1b2 had a most recent ancestor, with Gordon Clusters A and D paired with Cluster K at about 12,000 years.

Table 7a: Summary of Times when MRCAs of Three Gordon Haplogroups Lived (Kyrs ago)

	G(I2b1)	H(R1b1b2)	T(R1b1b2)	K(R1b1b2)	A(I1)	D(I1)	E(I1)
G (I2b1)		17.5	17.5	14.9	11.5	10.6	9.5
H(R1b1b2)	17.5		2.25	5.3	8.3	7.6	9.35
T(R1b1b2)	17.5	2.25		5.3	10.7	9.35	10.7
K(R1b1b2)	14.9	5.3	5.3		11.95	11.95	12.65
A (I1)	11.5	8.3	10.7	11.95		2.25	2.1
D(I1)	10.6	7.6	9.35	11.95	2.25		2.1
E (I1)	9.5	9.35	10.7	12.65	2.1	2.1	

Table 7b: Estimated ISOGG Dates for the Origins or Splits of Haplogroups I and R (Kyrs ago)

ISOGG Event	Time of Event (Kyrs ago)	Comparison with Results in Table 7a
ISOGG I	Before 18-20	cf. I2b1 at 17.5
ISOGG I1-I2 split	28	cf. I2b1 at 17.5
ISOGG R	27	
ISOGG R1	~ 18-22	cf. R1b1b2 at 17.5
ISOGG R1b1b2	4-8	cf. internals at 2.1-5.3

Estimated dates for the origin of haplogroups are given in the International Society of Genetic Genealogy's Y-DNA Haplogroup Tree[8]. In its 2010 version, it is suggested that Haplogroup I likely divided into Haplogroups I1 and I2 approximately 28,000 years ago. Additionally, Haplogroup R is believed to have arisen about 27,000 years ago in Asia, but its subgroups, R1 and R2 arose more recently. R1 is estimated to have arisen during the height of the last glacial maximum, with R1b arising in southwest Asia. Haplogroup R1b1b2 also originated in southwest Asia and is observed most frequently now in Europe, especially western Europe. This branch of R holds the Gordon Clusters in Haplogroup R1b1b2 and the ISOGG estimates that

it originated approximately 4000-8000 years ago. These estimates are summarized in Table 7b.

With a larger sample of Gordon testees, an earlier date might be found for the cluster intersections between haplogroups. Nevertheless, a comparison of the ISOGG dates with those determined using the RCC time scale shows good agreement and no inconsistency between the RCC- and ISOGG-derived estimates. The ISOGG estimates that R1b1b2 arose approximately 4-8 Kyears ago in southwest Asia and that it spread into Europe from there. The TMRCAs of Haplogroup R1b1b2 Clusters H, T and K, when paired, show a date range of 2.25 to 5.3 Kyrs. Since these are lower limits to the date, there is good agreement and no inconsistency between the two date estimates. In fact, it may suggest that the progenitors of the Gordons within R1b1b2 formed when the cluster progenitors had already reached Western Europe or even to the Scottish Highlands.

The ISOGG time estimates, the RCC time scale, the Y-DNA evidence and our results are consistent with an origin of the Gordon surname in areas near modern Turkey and Greece. However, given the relative small number (~1%) of I1 in Anatolia, the probability increases that origins of the Gordon ancestors who carry Y-DNA haplotypes in Macedonia and present-day Turkey are R1b1 (Cinnioğlu et al, 2004).

Comparison of the Gordon RCC Time Estimates and Historical and Pedigree Records

The Gordon surname was probably chosen because of location and its association with a famous contemporary and not by occupation or physical characteristic. The origin of the Gordon surname may tie to early BCE settlements below 42 degrees North latitude in modern day Turkey, Greece and Crete and/or to individuals with names like Gordian, Gordias, Gortys, Gordus, Gordinis. The historical and chronological record may trace the evolution of the name from these areas into France during the first CE millennium and from there to the British Isles at the time of William the Conqueror[9].

Gordon clusters in Haplogroups I1 and I2b1 shared common ancestors as recently as 12.6 Kyears ago, and this places a lower limit on the epoch of their pairing, well within the ISOGG estimate that the Haplogroups I1 and I2 split about 28 Kyears ago. The origin of the Gordons at latitudes below 42 degrees North, was comfortably below the southernmost extension of the last glaciation. The earliest common ancestors of all Gordons, again a lower limit, lived about 17.5 Kyears ago, as the glaciers began receding. Members of the Gordon Haplogroups I1 and I2 then probably migrated northward, following the glacier melt. These dates are consistent with the contention that "Human site occupation density was most prevalent in the Crimea region and increased as early as ca. 16,000 years before the present. However, reoccupation of northern territories of the East European Plain did not occur until 13,000 years before the present"[10]. The earliest common ancestor found between Haplogroups I1 and R1b1b2 (Gordon Clusters E and K), lived about 12.6 Kyears ago according to the RCC Time Scale, in good agreement with the reoccupation of those northern territories after the glacier receded. It is then consistent with the DNA record that the Gordon Clusters in Haplogroup I migrated to the northern regions of Scandinavia while the Gordon Clusters in Haplogroup R migrated into France and other regions of Western Europe. This activity occurred at times well before they could be compared with pedigrees.

The first inhabitants of Britain probably came from France, across a much shallower English Channel or by boat from the seacoasts of Western Europe. Archeologists have found and dated artifacts near lakes and seashores used by hunters who first visited only during the warm season of the Mesolithic in 3500-8000 BCE, dates that are consistent with the TMRCAs of the Gordon clusters in R1b1b2 haplogroups when H, K, L, T, and Q members shared a common ancestor. Nomadic animal herders arrived some time after 5000 BCE and became Britain's first farmers. By the Neolithic period (3500-2500 BCE) Britain had become an island. At its end, starting in the Bronze Age (2500-500 BCE), farmers were settling, clearing forests and beginning to use stone tools that transitioned to bronze after about 2300 BCE. Weapons developed from bronze became more effective in the Iron Age (500 BCE-70 CE), a period when population pressures and the growth of the ruling class prompted the need for defensive structures. Powerful chiefs formed the nucleus of what was to develop into the Scottish clans, with the farmers transitioning to vassal status, serving the chiefs in exchange for protection. This time period was also before individuals would appear in pedigrees, but the ties of the vassals to the chiefs probably resulted in the choice of the chiefs' surnames when it was time to choose surnames.

While it is consistent with modern history and the DNA record that Gordons within Haplogroup R came to Britain and Scotland across the English Channel, Gordons within Haplogroup I probably came to Britain and Scotland as Viking raiders from Normandy, married, stayed, and were assimilated into Britain in the epoch between 500 and 1000 CE.

The first Gordon on record and in a pedigree, Richard of the Barony of Gordon, lived in the mid-12th century. Crude pedigrees and the formation of House of Gordon go back only to the 1300s when Sir Adam Gordon led the family in the Battle of Halidon Hill in 1333. More trustworthy pedigrees date only from the 14th century when the House of Gordon first appeared.

RELATIONSHIPS BETWEEN THE GORDON PEDIGREE LINES AND THE GORDON CLUSTERS (see Table 5).

A.   The Gordon Septs:
One of the goals of The Gordon DNA Project has been to determine whether there might be any genetic links between the Gordon Septs and other Gordons. Indeed, RCC reveals that there are indeed some genetic ties between three Gordon Septs belonging to the Lawrie, Todd, Atkinson and Craig families are represented by Clusters N (2 testees), and T (1) in Table 5. These clusters are in Haplotype R1b1b2, as are the Seton-Gordons.

Cluster N is amorphous and contains only four members, two of which have pedigrees of a Gordon Sept, one is classified in Subgroup 0 of the Small Grouping of Gordons, and one is ungrouped. Four out of the five members of Cluster T appear to belong the Small Grouping-Gordons - Subgroup 10; the other belongs to the Gordon Septs.

One testee, Kit No. 127855, was assigned to a specific Gordon group but was not included in a Gordon cluster. His haplotype and pedigree indicated that he was a member of the Gordon Septs and he is in Haplogroup I. But in Table 5, all three Gordon Septs belong to Haplogroup R1b1, an inconsistency that warranted closer

inspection since such "outliers" may offer valuable information on differences between the traditional/pedigree and the RCC matrix approaches. He definitely belongs to the Gordons since his RCC associations are under 50, near the cluster edges of A, C, D, and E; he is in Haplotype I1. The inconsistency is that he has been assigned to the Gordon Septs, and Haplogroup R1b1 instead of Haplogroup I. Since the Spring of 2010 when the composition of Gordons we studied was 'frozen', there have been more testees assigned to the Gordon Septs. Since September 2010 the number of Gordon Septs has grown from 4 to 41. There are 9 more Gordon Septs in Haplogroup I, although 75 percent of all Septs are in Haplogroup R1b1. The fact that the Gordon Septs appear to have membership in both Haplogroups I and R means that two Gordon Sept lines in two different haplogroups will share no common ancestor within 7,000 years, while two Gordon Septs within the same haplogroup may share a common ancestor more recent than about 5,000 years. The lesson to be learned here is that small sample statistics can be misleading and that care must be exercised when broad, significant conclusions are drawn from insufficient data.

B.   The Jock and Tam Gordons and the Sir William Gordon Branch
Of the 43 testees assigned to the Jock and Tam Gordon group, all, without exception, were assigned to Gordon Cluster A. About half of the members in Cluster A are in closer associations within subclusters. Four testees assigned to the Sir William Gordon Branch also appear to belong to Gordon Cluster A. This overlap between the Jock and Tam Gordons and the members of the Sir William Gordon Branch suggests the sharing of a common ancestor within the genealogical time frame.

C.   Small Grouping-Gordons

Clusters in this group of Small Gordon Groupings are fragmented with no documented or otherwise identifiable non-genetic connection between clusters; however, it is worth noting that most clustering occurs in the R1b1 groupings, possibly attributable to the high occurrence of the haplotype in Western Europe.

Through FTDNA, and ySearch.org, the project has identified several groupings that do have high-resolution 67-marker matches, such as the Stewart-Gordons. Thus, the project has hyphenated the name Stewart-Gordon to reflect the likely Stewart ties and hyphenates other small groupings if a high-resolution genetic link is found with other surnames.

D.   Small Grouping-Gordons – The Stewart-Gordons
This Gordon group had a documented common Gordon ancestor in the mid-1700s. Albeit, matching no other Gordons, this group matched 65+ markers at the 67-marker level with the Stewart families.
All seven testees studied who belonged to Haplogroup I2b1 were also assigned to Gordon Cluster G, the oldest Gordon cluster studied here. The Gordons in Haplogroup I2b1 share a MRCA who is the best candidate yet to be the progenitor of the Gordons – at least of the Gordons included in our study. Intersections of Gordon Cluster G with Gordon members of other clusters and haplogroups are estimated to have occurred about 17,500 years ago, at the end of the last glacial maximum.

E.   Subgroups 0, 2, and 8 of the Small Grouping-Gordons
All members of subgroup 2 are members of Gordon Cluster C in Haplogroup I1, and all members of Cluster C are in subgroup 2. The five members of subgroup 0 are in

Clusters L and N (Haplogroup R1b1b2). Cluster L contains only subgroup 0 members, while Cluster N contains membership from two Gordon Septs, one in subgroup 0 and one ungrouped Gordon. There are only two members of subgroup 8 and both are in Cluster O, which also contains a member assigned to the Seton-Gordon branch.

F.   The Seton-Gordon Branch – All are in Haplogroup R1b1b2
The 16 members of this pedigree branch appear in four different Gordon Clusters, H, K, O and Q. The four members of Gordon Cluster K and the seven members of Gordon Cluster Q are the only members of their clusters that contain a member of the Seton-Gordon Branch; other members of the branch are members of Clusters H and O. Figure 4B indicates that the haplotypes of Clusters K and Q have similar haplotypes so each cluster shares a common ancestor 300-400 years ago, making them leading cluster candidates for the Seton-Gordon relation.
One might expect this branch to be confined to only one Gordon cluster as the Jock and Tam Gordons are confined to Gordon Cluster A but this is not the case. It shows the difficulty in making a pedigree assignment to an RCC cluster (or the converse). More testees are needed to resolve this difficulty.

G.   The Ungrouped Gordons
Only two testees, not grouped by the traditional approach, appeared in a Gordon cluster and both were in different clusters, H and N. Table 5 shows their association with other members of those clusters. Clearly more data are needed to see if the ungrouped Gordons might be in other, as yet undiscovered, Gordon clusters.

In the foregoing discussion and in assessing the results shown in Table 5, one must be careful not to over-interpret situations where only a small number of testees and their pedigree groupings have been assigned to a Gordon Cluster. When the entries in Table 5 are low, perhaps below five, they should be viewed as suggestive. Thus, the most valid conclusions that can be drawn from Table 5 are:
  • The Jock and Tam Gordons belong only to Gordon Cluster A. Half are in subcluster Aa, indicating that there should be a MRCA for that subcluster within the past 600 years, and possibly earlier.
  • The Progenitor Branch appears only in Gordon Clusters D and E.
  • The Small Grouping-Gordons – Stewart-Gordons belong only to the oldest Gordon Cluster G of Haplogroup I21b.
  • Subgroup 2 of the Small Grouping Gordons belongs to Gordon Cluster C and should have a MRCA about 350 years ago.
  • From the data in Table 5 and the results in Figure 4, there is a higher probability of finding pedigree matches among the I haplogroup than among the R haplogroup. However, this situation may change with the addition of new testee data.

CONCLUSIONS:

A. General Conclusions:

Totally new information that shows remarkable relationships among pedigrees, cluster and subcluster membership has become apparent through this study. Our results yield insight into the evolution and time sequences of haplotypes. The analysis suggests how a surname may be traced to a geographical location.

The presence of subclusters within large surname clusters in the RCC matrix was noted in Howard 2009b. Detailed study of subclusters (e.g., in Gordon Cluster A) shows that available pedigrees correlate highly with membership in a subcluster. The close association of subcluster haplotypes within the RCC matrix, combined with the RCC time scale, indicates how subclusters and pedigree lines may be tied together.

For many testees who do not yet know how they connect to others with a shared surname around the world, these correlations offer a significant new clue for focusing their research.

A comparison of the ISOGG dates with those determined using the RCC time scale shows good agreement and no inconsistency between the RCC- and ISOGG-derived estimates.

B. Conclusions Applicable to the Gordon Surname

Our study has uncovered correlations between recent historical activity and the formation of subclusters. For example, there is activity in the Gordon A subclusters around the time of the Jacobite Rebellions and internal family feuds over titles. It may be significant that we see subclusters develop after such events, when families are torn apart.

The ISOGG time estimates, the RCC time scale, the Y-DNA evidence and our results are consistent with an origin of the Gordon surname in areas near modern Turkey and Greece, with one major branch (the I Haplogroup) migrating to areas near Scandinavia and then into modern day UK and the other major branch (the R Haplogroup) migrating into western Europe and then into Britain. The times derived from the RCC matrix for the early migrations into the British Isles from Scandinavia and from Western Europe agree well with the history of the area derived from archaeological excavations, genetics and anthropologic studies like the Genographic Project[11].

The theory of Gordons originating from Normandy and with Malcolm Canmore coming to Scotland fits the time scale of the I1 profile. The Gordons became Scots and lived together having different I and R haplogroups. In modern times many Gordons have populated other areas around the earth, but their haplogroups give good clues as to the origins of many of their individual Gordon branches.

When surnames were adopted, those choosing the Gordon surname probably had placename roots or adopted the name of a Clan chieftan.

The assignment of a testee to a grouping based on traditional haplotype matching and existing pedigree information correlates highly with the assignment of that testee to a Gordon cluster through a comparison of his RCC value with those of other testees in the sample. When about five or more testees are assigned to a cluster and when the RCC values of the cluster testees are less than about 20, there is a remarkable agreement between the cluster identity and its pedigree assignment.

The "value-added" feature of the RCC approach is to add a time dimension to the analysis of the Gordon clusters. Application of that time dimension to the pedigree line should not only suggest the proper pedigree line to which a testee belongs, but also the time frame where his most recent common ancestor with other cluster members may have lived.

Further exploration of French, Spanish and Latin documents should be made for first-hand accounts on Gordons, prior to their arrival in Scotland. Male Gordon testees with well-documented pedigrees should be recruited for each of the Gordon branches, especially the Seton-Gordons. The testee group should be broadened to include Gordons and surname variations from Anatolia, Macedonian, Ghent, French and Spanish regions. Additional studies are suggested on the House of Gordon USA website.

REFERENCES:

Bulloch, John Malcolm (1903), *The House of Gordon, Volume 1,* New Spalding Club, Aberdeen.

Bulloch, John Malcolm (1906), *Scottish Notes and Queries.* D. Wyllie and Son.

Bulloch, John Malcolm (1907), *The House of Gordon, Volume 2*, New Spalding Club, Aberdeen.

Bulletin de la Société scientifique, historique et archéologique de la Corrèze, Volume 32. M. Roche, impr., 1910

Chalmers, George [b.1742-d.1825], (1887), *Caledonia, or an Account, Historical and Topographic, of North Britain, from the Most Ancient to the Present Times:* with a dictionary of places, chorographical and philological, Volumes I, II, III, IV.; Downloadable from University of California Libraries.

Charles, 11th Marquess of Huntly, ed. (1894), *The Records of Aboyne, 1230-1681*, New Spalding Club. Aberdeen.

Cinnioğlu, C; King, R; Kivisild, T; Kalfoğlu, E; Atasoy, S; Cavalleri, GL; Lillie, AS; Roseman, CC et al., (2004) *Excavating Y-Chromosome Haplotype Strata in Anatolia*, Human genetics 114 (2): 127–48.

Dickens, Charles (1887), *All the year round,* Volume 60, published by Charles Dickens.

Eutropias (370 CE), *Book IX of Abridgement of Roman History*.

Fordun, John of , (ca. 1384), *Chronica Gentis Scotorum* (The Historians of Scotland).

Howard, William E. III, (2009a), *The Use of Correlation Techniques for the Analysis of Pairs of Y-Chromosome DNA Haplotypes, Part I: Rationale, Methodology and Genealogy Time Scale*, J. Genet. Geneal., 5: 256-270.. *J Genet Geneol,* 4:104-124.

Howard (2009b): 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*, J. Genet. Geneal., 5: 271-288.

Howard, William E. III, and Schwab, Frederic R., (2012), *Dating Y-DNA Haplotypes on a Phylogenetic Tree: Tying the Genealogy of Pedigrees and Surname Clusters into Genetic Time Scales*, J. Genet. Geneal, this issue.

McBride, Nancy S., (1973), *Gordon Kinship*. McClure Printing Company.

Rymer, Thomas, (1704-1735),*Foedera*, Vol. 20, A. & J. Churchill

Seaton, Oren Andrew (1906), *The Seaton Family, with Genealogy and Biographies*, Crane & Company, pp. 55-56

Seton, Robert, Monsignor, (1899), *An Old Family, History of the Setons of Scotland and America*, Brentanos, New York, pp. 44-49

Skelton, Constance Oliver and Bulloch, John Malcolm (1912), *The House of Gordon, Volume 3: Gordons Under Arms, a Biographical Muster Roll of Officers named Gordon in the Navies and Armies of Britain, Europe, America and in the Jacobite Risings*, New Spalding Club, Aberdeen.

Wyntoun, Androw (c. 1350 – c. 1423) (1872), Edited by David Laing, *The Orygynale Cronykil of Scotland*, Syllabus 1-9 (Vol. 1-3), Edmonston and Douglas.

READING LIST:

Chalmers, George, Letters (1784-1816) to George Chalmers regarding the genealogy of the families of Gordon and Gregory. Personal and estate papers, Heritage Division of Aberdeen University.

Ferrerius, John, (1545),MS. *Historiae compendium de origine et incremento Gordoniae familiae*, Joanne Ferrerio Pedemontano authore, apud Kinlos, fideliter collectum.

Fordun, John (Skene, William F. ed.), (1877), *The Historians of Scotland*, Edinburgh.

Gordon, Charles, (11th Marquess of Huntley), (1894), *The Records of Aboyne*, MCCXXX-MDCLXXI, New Spalding Club.

Lythe, S.G.E., and J. Butt, (1975), *An Economic History of Scotland, 1100-1939*. Blackie, Glasgow.

Maitland, Richard, Sir, of Lethington, Knight, *The History of the House of Seytoun to the Year MDLIX, with the Continuation. Alexander Viscount Kingston, to MDCLXXXVII.* Printed at Glasgow. MDCCCXXIX.

Seton, George, Advocate, M.A., Oxon, (1896), *A History of the Family of Seton during Eight Centuries. George Seton, Advocate, M.A. Oxon*., etc. Vol. I & 2. Edinburgh.⌞SEP⌟

Smith, William, ed., (1870), *Dictionary of Greek and Roman Geography*, Little Brown & Co., Boston.

Most above Referenced and Further Reading books have electronic, downloadable copies available at:
www.books.google.com
www.archive.org

WEB PAGES:

The Gordon DNA Project
www.TheGordonDNAproject.com

House of Gordon research
http://www.houseofgordonusa.org/

Seton and Winton family research
http://www2.thesetonfamily.com:8080/cadets/Winton_Family.htm

DNA processing lab
http://www.FamilyTreeDNA.com

APPENDIX A – IDENTIFYING CLUSTERS IN THE RCC MATRIX:

We have found the following method to be the most useful and straightforward for identifying the clusters in the RCC matrix :
- After the RCC matrix is formed, set up a time slice matrix (Appendix A of Howard 2009a) just below it.
- Color the diagonal and remove the zero values from the diagonal.
    - There are the intersections where the column and row designations refer to the same testee.
- Set the low RCC value equal to zero.
- Set the high value of RCC high enough to fill the entire matrix.
    - This time slice matrix will now look identical to the one above it, and it will contain RCC values, and their formulas.
- Copy and paste the full RCC matrix to a new spreadsheet that consists only of RCC values, not formulas.
    - If you are using Excel, copy and paste the full matrix using the 'values only' paste option.
    - You will now have two matrices – one, the time slice matrix that can have its high and low RCC values changed, and the other, a full, values-only matrix.
- Use the previous time slice matrix to assist in identifying the rows where sequences of low value RCCs are located.
    - Set the low RCC value to zero and vary the high value to identify those rows and columns that contain low RCC values.

- By inspection, cut and paste the rows so that low value RCC intersections are grouped together in adjacent rows.
  - In general, those values of RCC will be of the order of 20-30.
- Cut and paste the columns so that a colored diagonal again appears in the reformed RCC matrix.
  - The row and column sequences in the reordered RCC matrix will then be identical.
- Repeat the process if there are groups of low-value RCCs still in the matrix that are not oriented near the diagonal.
- Group the low RCC values together into surname clusters.
  - Look for subclusters of very low RCC values within each surname cluster.

APPENDIX B – Data on Kit Number associations with Gordon Subclusters, Clusters and Interclusters, with Haplogroup and Pedigree designation, and with associated statistics.

APPENDIX B1: List of 119 Subcluster and Cluster members of the Gordon RCC Matrix by Kit Number, Haplogroup and Gordon Cluster membership.

APPENDIX B2: For each subcluster, major cluster and intercluster, the average RCC, standard deviation (SD) of the distribution, number of testee pairs and the SD of the mean, with conversions to time in the past and year in the past based on 1 RCC = 43.3 years for intercluster TCAs and 1 RCC = 52.7 for the TCA of clusters.

APPENDIX B3: 104 Gordon Clusters (from RCC Surname Matrix) and Corresponding Gordon Lines (from Pedigrees and Traditional Marker Groupings)

Kit No	Gordon Cluster	Gordon Pedigree Designation	Kit No	Gordon Cluster	Gordon Pedigree Designation
3071	Q	SETON-GORDON BRANCH	77948	Ae	Jock & Tam Gordons
6608	Ac	Jock & Tam Gordons	80269	Ab	Sir William Gordon Branch
6662	C	SMALL GROUPING-GORDONS – Subgroup 2	82714	E	Progenitor Branch
6770	Q	SETON-GORDON BRANCH	84517	Q	SETON-GORDON BRANCH
6920	A	Jock & Tam Gordons	84557	N	GORDON SEPTS
7805	Af	Jock & Tam Gordons	87584	Aa	Jock & Tam Gordons
7876	A	Jock & Tam Gordons	88518	E	Progenitor Branch
7912	Aa	Jock & Tam Gordons	89515	Ab	Sir William Gordon Branch
8058	H	SETON-GORDON BRANCH	90576	H	UNGROUPED – GORDONS
8882	E	Progenitor Branch	93250	N	SMALL GROUPING-GORDONS – Subgroup 0
9310	Q	SETON-GORDON BRANCH	93333	Aa	Sir William Gordon Branch
9564	Ad	Jock & Tam Gordons	95325	O	SMALL GROUPING-GORDONS – Subgroup 8
10326	N	UNGROUPED – GORDONS	95985	L	SMALL GROUPING-GORDONS – Subgroup 0
11284	D	Progenitor Branch	96009	D	Progenitor Branch
12577	T	SMALL GROUPING-GORDONS – Subgroup 10	96537	H	SETON-GORDON BRANCH
14070	Aa	Jock & Tam Gordons	96713	T	SMALL GROUPING-GORDONS – Subgroup 10
17187	C	SMALL GROUPING-GORDONS – Subgroup 2	97554	Aa	Jock & Tam Gordons
19162	A	Jock & Tam Gordons	97675	G	SMALL GROUPING-GORDONS – Stewart-Gordons
19380	L	SMALL GROUPING-GORDONS – Subgroup 0	102156	E	Progenitor Branch
20304	A	Jock & Tam Gordons	105253	L	SMALL GROUPING-GORDONS – Subgroup 0
20391	A	Jock & Tam Gordons	108388	Ac	Jock & Tam Gordons
21381	A	Jock & Tam Gordons	111517	K	SETON-GORDON BRANCH
23202	G	SMALL GROUPING-GORDONS – Stewart-Gordons	111972	A	Jock & Tam Gordons
23215	G	SMALL GROUPING-GORDONS – Stewart-Gordons	114003	Af	Jock & Tam Gordons
29243	Ad	Jock & Tam Gordons	114624	A	Jock & Tam Gordons
29811	A	Jock & Tam Gordons	116419	A	Jock & Tam Gordons
32378	C	SMALL GROUPING-GORDONS – Subgroup 2	120456	H	SETON-GORDON BRANCH
34523	D	Progenitor Branch	121077	T	SMALL GROUPING-GORDONS – Subgroup 10
35965	E	Progenitor Branch	124394	A	Jock & Tam Gordons
36388	Aa	Jock & Tam Gordons	125318	A	Jock & Tam Gordons
39896	G	SMALL GROUPING-GORDONS – Stewart-Gordons	126051	A	Jock & Tam Gordons
40174	Q	SETON-GORDON BRANCH	129211	Aa	Jock & Tam Gordons
41514	K	SETON-GORDON BRANCH	134805	A	Jock & Tam Gordons
42605	T	SMALL GROUPING-GORDONS – Subgroup 10	137716	Aa	Jock & Tam Gordons
45353	G	SMALL GROUPING-GORDONS – Stewart-Gordons	143101	Ae	Jock & Tam Gordons
45694	Aa	Jock & Tam Gordons	145232	Aa	Jock & Tam Gordons
46758	C	SMALL GROUPING-GORDONS – Subgroup 2	149140	A	Jock & Tam Gordons
47249	G	SMALL GROUPING-GORDONS – Stewart-Gordons	149866	O	SMALL GROUPING-GORDONS – Subgroup 8
47335	Aa	Jock & Tam Gordons	155174	H	SETON-GORDON BRANCH
48643	A	Jock & Tam Gordons	157262	T	GORDON SEPTS
49327	D	Progenitor Branch	158402	C	SMALL GROUPING-GORDONS – Subgroup 2
50914	Ae	Jock & Tam Gordons	160431	K	SETON-GORDON BRANCH
56420	Ae	Jock & Tam Gordons	163702	A	Jock & Tam Gordons
58199	K	SETON-GORDON BRANCH	165571	N	GORDON SEPTS
58440	Q	SETON-GORDON BRANCH	166456	Ab	Sir William Gordon Branch
62423	Q	SETON-GORDON BRANCH	167874	A	Jock & Tam Gordons
62548	L	SMALL GROUPING-GORDONS – Subgroup 0	171413	A	Jock & Tam Gordons
69173	Ae	Jock & Tam Gordons	172482	O	SETON-GORDON BRANCH
69847	A	Jock & Tam Gordons	173539	A	Jock & Tam Gordons
70303	E	Progenitor Branch	N12530	D	Progenitor Branch
73551	D	Progenitor Branch	N34227	G	SMALL GROUPING-GORDONS – Stewart-Gordons
73552	D	Progenitor Branch	N69338	Ae	Jock & Tam Gordons

[1] Despite the apparent precision of dates in this analysis, they are probably uncertain by of the order of 300 years (SD) because, for recent times, differences in mutations, which average out over long periods of time, will cause unpredictable uncertainties in the times when the common ancestor of recent clusters lived (see Howard 2009a). An uncertainty of 300 years (~10 generations) is 30 percent when we deal with genealogically interesting times of 1000 years. The RCC time scale has random, mutation-induced errors that are about the same as those assigned by the testing agencies. However, an uncertainty of 300 years that results from analyses of cluster intersections or haplogroup intersections that often range above RCC ~100 translates to smaller uncertainties when many testees are involved in the time determination.

[2] http://www.familytreedna.com/public/gordondna/default.aspx?section=yresults. Ties between Kit numbers, haplotypes, Gordon cluster designations and pedigree associations can be found in Appendix B.

[3] When two Y-DNA haplotype 37-marker strings are correlated, the resulting correlation coefficient is usually a number greater than 0.9. In order to simplify the analysis, we define the Revised Correlation Coefficient (RCC) as the reciprocal of the correlation coefficient minus one times 10,000. Thus RCC will typically be a number between 0 and 1200. It is proportional to the elapsed time between the TMRCA of the pairs of haplotypes. If TCA is the time when the common ancestor of a cluster of intercluster lived, we found (Howard 2009a):
TCA, cluster = Average RCC of all pairs of cluster members x 52.7 years.
TCA, intercluster = Average RCC of all pairs of intercluster members x 43.3 years.

[4] It must be emphasized that the theories of the pre-pedigree ancestry of the Gordons, as well as their familial origins cited throughout this paper, are often conjectural and most may never be convincingly proved. The veracity of ancient records is nearly impossible to verify. Because parts of these conjectures may have a factual basis, we present them here in one place because they may be useful to future researchers. New genetic tools such as the RCC method to be applied to the Gordons later in this paper may serve to support or cast more doubt in these theories, and we hope that applying the correlation approach will further our understanding of the Gordon ancestry and its origins. Our paper attempts to draw substantive conclusions only from the most creditable available information; its purpose is to stimulate dialogue among future generations of researchers as more extensive DNA testing and new analytical tools become available.

[5] After 75 years this type of information is in the Public Domain; Bulloch 1903, and 1907

[6] William Gordon of Crogo see: http://www.thegordondnaproject.com/93333.html
Kit No. 89515 has a well-documented pedigree, but its haplotype and RCC values are close to members of other subclusters and other members of Cluster A that have not been assigned to a subcluster. Similar observations show a similar anomaly with Kit No. 93333. We believe that these uncertainties indicate a very close association with the TMRCAs of subclusters Aa, Ab and Ae within major Gordon Cluster A, with these two testees having ancestors near the transition point into individual subclusters.

[7] It is evident that when an applications program (e.g., Mathematica) is used to form a phylogenetic tree from the same data, the common ancestor for interclusters that have values of RCC that differ by less than about 10% lived at the same time (see Howard and Schwab 2012, this issue)

[8] The ISOGG's Y-DNA Haplogroup Tree can be found at <http://www.isogg.org/tree/index.html>. It is being continually updated.

[9] See the genealogical section of the House of Gordon <http://www.houseofgordonusa.org>

[10] http://en.wikipedia.org/wiki/Late_Glacial_Maximum

[11] https://genographic.nationalgeographic.com/genographic/index.html