ARTICLE

Scotland's R1a1 Highland Clansmen, DNA Genealogy and the search for Somerled

Andrew MacEacharn[1], Anatole A. Klyosov[2*]

We have examined over 300 R1a1 Highland Clansmen yDNA haplotypes searching for a common male ancestor of any of the studied clans, and whose name was Somerled hero of the Celts, who lived in the twelfth century. For this search we employed the emerging science of DNA genealogy, which is a merge of DNA sequencing and chemical kinetics. We identified the Old Scandinavian (4100 ybp) and Young Scandinavian (1900 ybp), the Gaelic and perhaps the Danish "*Dark Stranger"* clansmen lineages all with a common ancestor from the Russian Plains about 5000 years ago. This analysis ended with us defining the ancestral ethnicity of the R1a1 clansmen who gave Scotland her name and her culture.

[1]25 Howard Road, Minto Heights, NSW 2566, Australia

amceache@unwired.com.au

[2]36 Walsh Road, Newton, MA 02459, USA

aklyosov@comcast.net

URL: http://aklyosov.home.comcast.net

*Correspondence: aklyosov@comcast.net

Introduction

We know so little of this man called *Somerled.* Who was he, where did he come from? For a man of such reputed fame we find very little of him in historical documents. Historians contemporary with him write very little or nothing of him, except perhaps of his death. He is not mentioned unambiguously in any of the various Annals of Ireland such as Tighernac, Ulster etc in the times we know him as living in being 1100 – 1164AD. We do find an entry in the Annals of Ulster for the year 1164AD:

U1164.4

Somharlidh Mac Gilla-Adhamhnain and his son were killed and slaughter of the Men of Airthir-Gaedhel and of Cenn-tire and of the men of Insi-Gall and of the Foreigners of Ath-cliath.

In the above passage we find Somerled the son of Gilladomnan, not Gillebride.

And also an entry in the Annals of Tigernach for 1164AD:

1164

Somerled son of Gilla Brigte king of the Hebrides and Kintyre, and his son Gilla Bright were killed by the men of Scotland, and along with them the Foreigners of Dublin were slaughtered.

In this entry we are told that Somerled has a son Gillebride.

Investigation of nearly every reference to the man *Somerled* in fact leads us nowhere. Many references to him are written retrospectively so can barely be relied upon as being historically factual.

Of his genealogy we understand his father was Gillebride son of Gilladomnan[1-3]. We believe his ethnicity to be Gaelic but we do not know. Recently it has been reported his paternal ancestry is Norse (clan-donald Web Resources, see the list at the end) and there seems to be some significant ground swell of attraction to this.

What of the Highland Clans? Where did they come from? Some such as Clan Donald, Clan MacAlister and Clan MacDougal (see Web Resources) claim direct paternal descent from Somerled. Others such as Clan Iness claim paternal ascendancy to Somerled (Web Resources). Clan Innes, whose ancestral lands are in Morvern, claim a confederation existed called Siol Gillebride (The Seed of the Servant of St Bride) between the MacMaster, MacGilvray, MacInnes clans and the MacEacharn family. Other clans and families make no claims at all. Some such as the MacIains of Ardnamurchan understand their relationship to Somerled to be through Clan Donald (Web Resources). Most Highland Clans, in reality do not definitively know their paternal ancestry. Paper trails hold no value as we have no way to cross reference and verify these.

To understand the ethnic makeup of a Scottish Highland Clan we must first look back in time to the migration of two cultures which began over 1500 years ago. The first culture clashed with the indigenous peoples of Scotland with the eventual unification of these cultures by Kenneth MacAlpin, who shared Gaelic and Pictish ancestry, in the ninth century AD. The second culture being the Scandinavians who started their raids on Scotland and Ireland around 795AD.

In the 5th Century we see the beginning of a nearly half millennium long movement of the Scoti or Scots (Web Resources) into Scotland from Ireland. Eventually Scotland would be named after this race of people. However, the roots of this race and its name could have had a much more ancient history. One of the oldest proto-Russian tribes had a name “Scotichi”, and

as such it was mentioned (and further described) in the very beginning of "The Book of Veles"[4], arguably the oldest proto-Slavic text. The first lines of the book say – "Vendi were brothers with the Scotichi in the steppes". Then the text describes that the Scotichi split and moved, apparently to the West. According to some scholars, it was thousands of years ago and certainly in the BC times.

If we turn to linguistics, Russian *scot* is *cattle* in English, the latter in turn has its roots in late Latin capitāle (wealth, possessions), while the Russian word is related to *cátus-pad* in Vedic Sanskrit texts, which means "four-footed", and "four" in Sanskrit is catvāras[5]. This points to ancient connections between proto-Slavs and proto-Celts. They lived close to each other and had the same lifestyle.

Sometime around 500AD Fergus Mor Mac Erc and his brothers, Angus and Lorne, moved the power base of the Scoti from County Antrim in Ireland to Argyll in Scotland. The Scoti brought the Gaelic language and culture with them as they settled mainland Scotland. Fergus Mor Mac Erc is traditionally considered the "father of Scotland" and is credited with bringing the Stone of Destiny, Lia Fáil, to Scotland. Around the same time period we see the beginning of a movement of ecclesiastical men into Scotland, the most well known being St Columba, who founded Iona in 563AD. From this time onward expansion of the Irish Scoti was assisted in part by the spread of Christianity. This movement appears to have initially crossed from Northern Ireland into the Argyll area of Scotland. St Columba's Church at Kintyre in Scotland is well known as one of the first mainland ecclesiastical sites in Scotland founded by an Irishman. Some time later St Columba founded his well documented monastery on Iona Island. His church in Kintyre is known as Cille Colum Cille or literally, the Cell of St Columba of the Cell. Another church situated above Loch Aline in Morvern, Somerled's apparent ancestral lands, is named the same and may be contemporary with St Columba. Ancestors of one of the authors of this study lie in this graveyard.

Throughout Kintyre and Islay we see a litany of "clerical" placenames such as Kilmartin (the Church of St Martin), Kilellan (the Church of St Fillan) and Kilchoman (the Church of St Comgan) to name a few. There is no shortage of archaeological, cultural and historical evidence for this migration[6] (also Web Resources at the end). St Comgan and his sister St Kentigerna are well known to have travelled to Scotland after leaving Leinster in the late 600's AD. With them came St Kentigerna's son Fillan later to be named a Saint also. Clan Cowan, a Highland Clan, are one clan of many, whose name in itself may provide some clue to their origins. The clans original ancient name is MacEichgillychomgan, however by the mid 1600's was already being written in English like M'Ilchoen when listed as some of the men murdered in Scotland's largest and virtually unacknowledged massacre of over three hundred Highlander men women and children, the Dunaverty Massacre at Kintyre in 1647.

Starting from around 795AD the first of the Norse raids starts on Ireland and Scotland. The monastery at Raithlin Island is attacked and pillaged and for the next thirty or forty years monasteries in Ireland are attacked nearly yearly. Iona was raided many times. There was some resistance to these attacks; however by 842AD the Norse had settled Dublin, Cork and Waterford, which was named Vadrefjord. Around 849AD we see a new type of northern invader, this being the Dane. These were known to the Irish as the "Dark Foreigners" and the Norse northern invader was known as the "Fair Foreigners". There are some reasons to believe now, that "Fair Foreigners" belonged to haplogroup I1, and "Dark Foreigners" carried haplogroup R1b1, likely R1b-U152. This subclade of R1b haplogroup is still among Danish people, many of whom are dark-haired and noticeably differ from neighbouring light-haired Norwegians and Swedes.

It is during the 9th century we see the Norse kings of Dublin in constant warfare with the Irish kings. This warfare chiefly comes to an end when Brian Boru reinforced from Argyll confronts the Norse army, reinforced from the Isle of Man and Orkney, at Clontarf in 1014

and after twelve hours of battle defeats them. This defeat put to an end any serious Norse threat to Ireland. The settlements of Orkney, Shetland, Isle of Man, and the Outer Isles and to some extent a portion of the Hebrides remained under Norse control. It is sometime after this defeat at Clontarf we see the rise of Godred Crovan, a Norse ruler of Dublin and King of Man and the Isles. It is after the death of Godred (1095AD) that the Kings of Man and the Isles become vassals to the King of Norway. Reginald King of Man and the Isles, son of Godred is contemporary with Reginald son of Somerled, who was also styled King of the Isles.

Now we find ourselves back to the time of Somerled. We see two different cultures in the area known generally as the Western Mainland and Isles of Scotland. By this time in history the Pictish culture had been absorbed by the Gaels. The Outer Isles, Isle of Man to the south, Orkney, Shetland, Skye and most of the northern mainland including Ardnamurchan appear to be under Norse control and influence whilst the mainland areas of Kintyre, Lorne, Morvern and the islands of Islay and Mull are under Gaelic control and influence. The mainland areas south of Ardnamurchan were constantly under attack from the Norse and anecdotally this is why Somerled drove them off. His father Gillebride was supposed to have been defeated by the Norse and retreated from Morvern back to Ireland.

The clans claiming a direct paternal relationship to Somerled are Clan Donald (the MacDonalds), Clan Dougal (the MacDougalls) and Clan Alister (the MacAlisters and Alexanders). Clan Innes or MacInnes claim that Somerleds father Gillebride may have been related to the progenitor of their clan, a man they call Oengus or Angus, one of the brothers of Fergus Mor Mac Erc (Web Resources). The MacIains of Ardnamurchan, an area just north of Morvern, claim their relationship to Somerled is through a younger son of Angus Mor of Isla, John Sprangach d 1340 (Web Resources). Clan Alister and the MacIans of Ardnamurchan consider themselves branch families of Clan Donald (Web Resources).

Clan Donald claims that their common ancestor with Somerled is John of Islay[7] (also Web Resources). Clan Donald gives their genealogy as Somerled d 1164, Reginald, Donald "de Isla" d 1289, Angus Mor MacDonald d 1300, Alexander MacDonald d 1303 succeeded by his brother Angus Og MacDonald d ca 1329, and John of Islay d 1386.

Clan Dougal claims a common ancestor with Somerled through another son of Somerled's called Dougal. Clan Dougal give their genealogy as Somerled, Dougall (1164), Duncan de Ergadia (1207), Ewen (1248), Sir Alexander (1266), Sir John of Lorn (1310), Duncan of Dunollie, Ewen (1344), Iain of Dunollie (1375).

Clan Alister claims that they too have a common ancestor with Somerled and also with Clan Donald. Clan Alister give their genealogy as Somerled, Reginald, Donald "de Isla" (same man as Clan Donald), Alexander (1253), younger brother of Angus Mor MacDonald. Some MacAlisters are now known by their Englished name Alexander.

Did Somerled have any brothers? Did he have any sons? The night before the Battle of Renfrew in 1164, Somerled and his eldest son Gillecallum were supposedly murdered by Maurice MacNeill, a nephew of Somerled. The MacDonald historian Hugh Macdonald calls this MacNeill a son of Somerleds sister Beatrice. At the same time Hugh MacDonald also makes mention of another Maurice MacNeill, foster brother to Olav, King of Man who was a close friend of Somerled.

Various "unofficial" sources (Web Resources) give the following for Somerleds male sons

1. Somerled
2. Gillecallum
3. Dugal
4. Angus

5. Reginald
6. Olave
7. Gilles
8. Gall

Whilst there must be some questions surrounding any definite historical proof for any of the sons of Somerled, the above list seems to have been compiled chiefly from the documents and claims of clans purporting a relationship to him.

Of the sons of Somerled and their sons, what do we know? Gillecallum apparently died with Somerled the night before the battle of Renfrew in 1164. Somerled MacSomerled died soon after 1220. Apparently he had two young sons, John and Maolmory. John was supposed to have died young without issue and Maolmory settled in Ireland, also without issue. Angus and his sons were killed on the Isle of Skye by the sons of Reginald, his nephews, in 1210. Dugal forms Clan Dugal and is considered the progenitor of the MacDougals of Lorne. Gilles is pursued in Kintyre by Reginald and killed. Reginald becomes the progenitor for Clan Donald. Little or nothing is known of Olave and Gall.

Clan Innes who potentially have claim to paternal ancestry from Somerled or his father appear to have no relations to their chief alive. In 1358 the Chief of the Clan and his sons were murdered by Clan MacLean on the orders of John of Islay, Lord of the Isles. The murders were affected at Ardtornish Castle, Morvern.

After researching the history and formation of Highland Clans, one fact stands out. None of them can for certainty know who their progenitor is. From about 1400AD many clans have excellent paper trails; however we find that we can not rely on these as a source of definitive information as to the Clans ancestors.

So we must turn to analysis of Y chromosomal DNA for our answers. In recent times this field of science has made very positive advances in terms of clarifying family associations and time periods to common ancestors. Given the limited amount of historical information for the time period we are concerned about and the litany of claims in clan histories that are more fallacy than fact, we must rely upon science to help us unravel the mystery that is the R1a1 Highland Clansmen.

Reviewing the Y chromosomal DNA makeup of the highland clans we see some clans with a chiefly line which is R1b such as Clan MacPherson. Clan MacNeill and Clan Campbell. These clans have a very limited amount of R1a1 members and appear to be predominately ancestrally Pictish. Some of the R1a1 MacPherson Clan members are ancestors of a man who was kidnapped from Scotland in the 1690's, so in themselves are actually not of MacPherson lineage. Other Clans such as MacDonald and Matheson have a chiefly line which is R1a1 with many R1a1 and R1b1 clan members.

Why R1a1? R1a1 is the haplogroup to which many Highland Clansmen and their Chiefs belong to. It is a predominantly an Eastern European haplogroup, though it originated allegedly in South Siberia around 20,000 years before present[8], it is somewhat frequent there nowadays[9,10], and it is excavated there and in adjacent regions from ancient cemeteries with an ever increasing pace[11-13]. It is strongly represented in Russia (up to 62% of the population), Poland (up to 57%), in Central Asia (50% in Kyrgyzstan), and in some other Asian regions (e.g., in India, Pakistan, Iran, among the Altaians in South Siberia, all between 39% and 15% of the population, etc.)[10].

Below are some questions addressed in the hope for clarity as to our ancestors and cultural ethnicity:

Where did the ancestors of R1a1 Highland Clansmen come from? Do we see two distinctly different cultures such as Norse and Gaelic appear in the framework of DNA genealogy and within the same R1a1 haplogroup?

Is there any proof of the existence of their "patriarch" or "patriarchs" in the time of the man we know as Somerled?

Can we tell if Clan Donald and Clan Alister share a common ancestor who died around 1289, that is about 720 years before present?

Do we see that Clan Donald have a common progenitor who died around 1386AD?

If we analyse only the markers of the living Clan Donald Chiefs and Chieftains, when did their common ancestor live? Does this agree with their clan genealogical documents?

Do we see any indication that any other clans may have Somerled as their ancestor?

Do we see any evidence for the MacEacharn Campbells of Craignish from Lorne?

What is our (R1a1 Highland Clansmen) deep ancestry 1300-1500 years back, in the middle of the 1st millennium AD? Where did we originate from?

If the haplotype tree based on haplotypes of Highland Clansmen, produce branches (lineages), how long ago did these branches happen? From where? From which upstream lineages did this happen?

What clans belong to which branch on the haplotype tree?

Do we see any evidence of a common ancestor around 400 – 500AD?

What is the date of the nearest common ancestor for the kidnapped MacPhersons and one of the authors of this study?

Do we find a MacInnes that could be a close ancestor to Somerled, based on the haplotype tree?

If we find evidence of his existence, was Somerled Norse? Or, in other words, was he Scandinavian?

Materials and Methods

Haplotypes

Haplotypes were collected for analysis from Family Tree DNA and Clan Project Webpages (Web Resources). Collection was done with knowledge of Highland Clan names in mind. However another non Highland name was included, these being Perkins due to them being seemingly much like the Highland R1a1 markers. Some Irish R1a1 markers were included.

Markers were not collected from public DNA sites such as YSearch for the primary haplotype dataset in this study. However, as it will be shown below, YSearch haplotypes match results in this study very well.

Haplotypes were grouped into sets of 67, 37, and 25 markers.

So as to not "overload" the analysis only a subset of available Clan Donald haplotypes were added for the total dataset. All available Clan Donald haplotypes were also analysed separately. Essentially, as is shown below, both approaches give the same results.

Clan Dougal markers are not available. The clan chooses to keep their markers private.

Analysis of mutations and their rates

Principal methodology was described earlter[8,14-16]. Haplotype trees were constructed using PHYLIP, the Phylogeny Inference Package program[17], as was explained in detail in[14-16]. A "comb" around the wheel (a "trunk"), in haplotype trees identifies "base" haplotypes, identical to each other and carrying no mutations compared to their ancestral haplotypes (e.g., in Figs 2, 5). The farther the haplotypes lay from the wheel, the more mutations they carry compared to the base haplotype and the older the respective branch.

The "base" haplotypes are the ancestral haplotypes in an ideal case. However, since those haplotypes often are deduced ones, it would be inaccurate to call them "ancestral". Hence we call them "base" haplotypes.

The timespans to a common ancestor were calculated using both "logarithmic" and "linear" approaches. The logarithmic method is based on the assumption (which largely or practically always holds true) that a transition of the base haplotypes into mutated ones is described by the first-order kinetics:

$N = Ae^{kt,}$

that is

$\ln(N/A) = kt$

where:

N = a total number of haplotypes in a set,

A = a number of unchanged (identical, not mutated) base haplotypes in the set,

k = an average mutation rate,

t = a number of generations to a common ancestor.

One can see the logarithmic method does not consider mutations in haplotypes; only mutated and non-mutated (base) haplotypes are considered.

Mutation-counting methods are all based on accumulation of mutations in haplotypes over time. They include the "linear", "quadratic" (ASD) and "permutational" methods[8,14]. In this paper only the "linear" method of those is employed. In this method a total number of mutations in a set of haplotypes is counted, an average number of mutations per marker is calculated, a correction for back mutations is introduced, either numerically[14], or using a handy table[14], and a time span to a common ancestor is calculated, either using the Table, or applying the respective mutation rates.

Average mutation rates employed in this paper, are[14]:

For 12 marker haplotypes – 0.022 mutations per haplotype, 0.00183 mutations per marker,

for 25 marker haplotypes – 0.046 mutations per haplotype, 0.00183 mutations per marker,

for 37 marker haplotypes – 0.090 mutations per haplotype, 0.00243 mutations per marker,

for 67 marker haplotypes – 0.145 mutations per haplotype, 0.00216 mutations per marker.

These mutation rates were calibrated employing 25 years per generation. This is a fixed mathematical figure, not an actual length of generations, which is a "floating" value, depending on many factors, including cultural, demographical, economical, and largely varied between ancient times and today. Time spans to a common ancestor were calculated

typically employing the first 25, 37 and 67 markers in haplotypes, and the data was compared to each other. Margins of error for time spans to common ancestors are calculated as described in[14], and correspond to the 95% confidence interval (two sigma).

Corrections for back mutations were made according to the table, published in earlier[14].

Results

A haplotype tree for the available 116 haplotypes in 67 marker format is presented in Fig. 1. These extended haplotypes provide us with the most resolved, refined branches (lineages) of the tree. On the other hand, many more 37 and 25 marker sets were available, namely, 229 and 253 of them, respectively. However, a 37 marker tree is less resolved, and a 25 marker one is even less resolved, since they lack as many as 30 and 42 markers on each haplotypes compared to 67 marker haplotypes. Those 30 extra markers (compared to a 37 marker tree) per haplotype add a total of 3510 markers in the pattern in Fig. 1, which might introduce some important information in terms of branches and their mutual positions. This is a common problem of the choice between quantity (of haplotypes) and quality (of branches resolution). To overcome this problem as much as we can, we will consider both the 67 and 37 marker trees, and for a comparison the 25 marker tree as well.

The right hand side which is compact and therefore a relatively "young" branch in Figure 1 contains 49 haplotypes, 30 of them belong to the Donald Clan, and the residual 19 haplotypes belong to the Perkins family (2 haplotypes out of 6 in their family, available in 67 marker haplotypes), MacCain (7 haplotypes out of 8 in their family), three McDaniels, two McAllisters (out of four in their family), McInnes, McNeill, Douglas, Matheson. This branch has a base haplotype that is approximately equidistant with respect to all 49 haplotypes in the

branch. In other words, all 49 haplotypes collectively have a minimal number of mutations from that base haplotype. Therefore, the base haplotype has a minimal genetic distance from all 49 haplotypes in the branch. It can be presented as follows (the four panels are separated; they contain markers 1-12, 13-25, 26-37, and 38-67):

13 25 15 11 11 14 12 12 10 14 11 31 -- 16 8 10 11 11 23 14 20 31 12 15 15 16 – 11 12 **19 21** 17 16 17 18 34 39 12 11 – 11 8 17 17 8 12 10 8 11 10 12 22 22 15 11 12 12 13 8 14 23 21 12 12 11 13 11 11 12 12

A characteristic feature of this base haplotype is a pair YCAII = 19, 21 (in bold above). All 49 haplotypes in the branch have this very pair. Most of the European R1a1 haplotypes as well as the Russian Plain and Asian (Central Asia, India, Iran) haplotypes have 19, 23 at this location.

Attentive consideration of the right-hand side branch, which we will call the Donald branch, since the Donald Clan haplotypes are prevalent there and includes the living chiefs of Clan Donald, shows that it contains three sub-branches. They can be combined and treated as one, or they can be considered individually. For demonstrational purposes we will analyse them both ways, since to date no one has performed such a comprehensive analysis of branches of a haplotype tree and published the results.

Combined, all 49 haplotypes of the Donald branch contain 45 mutations in the first 25 markers, 104 mutations in the first 37 markers, and 135 mutations in the 67 marker set, compared with the base haplotype shown above. Since the average mutation rate constants for these panels of markers equal to 0.046, 0.090 and 0.145 mutation per haplotype per

generation of 25 years[14] (the mutation rate constants were calibrated using 25 years per generation, and for, say, 30 years per generation the mutations rate constants should be simply proportionally adjusted, and will be equal to 0.055, 0.108 and 0.174 mutation per haplotype per generation of 30 years, respectively), we obtain 45/49/0.046 = 20 generations, that is **500±90** years to a common ancestor when we analyse the first 25 markers of the haplotype dataset, 104/49/0.090 = 24 generations, that is **600±80** years (when 37 markers are considered), and 135/49/0.145 = 19 generations, that is **475±60** years to a common ancestor (when all 67 markers are analysed).

Considering the random pattern of mutations of all 67 markers in 49 haplotypes, this data shows a fair consistency. However, since we combined three sub-branches of the branch into one pool of 3,283 alleles, one can expect that this could underestimate or overestimate the mutation count (it depends on the specifics of the sub-branches). Therefore, we analysed the sub-braches more accurately, as shown below.

The three sub-branches clockwise along the Donald branch contain 9, 22, and 18 haplotypes, respectively. None of the sub-branches contain just one Highland Clansmen family. The first one contains haplotypes of four MacDonalds, along with two MacAlisters, and three McKaine/McCains; the second contains 13 Donald haplotypes, along with four McCain's, two Perkin's, McDaniel, and some others. The third sub-branch (the most distinct on the branch in Fig. 1) contains 13 Donald haplotypes, two McDaniel's, Douglas, McNeil, and a McDonald from an allegedly different family.

All three sub-branches have essentially the same base haplotype as shown above, except slight variations in only three alleles out of 67 (shown in bold):

13 25 15 11 11 14 12 12 10 14 11 31 -- **16** 8 10 11 11 23 14 20 31 12 15 15 16 – 11 12 19 21 17 16 17 **18** 34 **39** 12 11 – 11 8 17 17 8 12 10 8 11 10 12 22 22 15 11 12 12 13 8 14 23 21 12 12 11 13 11 11 12 12

The first 9 haplotype branch has 15.00, 17.89 and 38.89 there, as average across all 9 haplotypes. The second, 22-haplotype branch has 15.73, 18.00, and 38.23 there. The third, 18-haplotype branch has there 16.00, 18.61, and 38.78, respectively. Those relatively minor variations caused separations of the sub-branches, since a haplotype tree is very sensitive even to small variations in alleles. The first sub-branch contains 28 mutations in all 9 haplotypes, that is 28/9/0.145 = 21 generations, or **525±110** years to a common ancestor of the branch. The second sub-branch contains 35 and 60 mutations in 37 and 67 markers in all 22 haplotypes, which results in **450±90** and **475±80** years to a common ancestor of the sub-branch. The third sub-branch contained 30 and 36 mutations in all 18 of 37- and 67-marker haplotypes, which gave **475±100** and **350±70** years to their common ancestor, respectively.

In order to calculate a timespan to a common ancestor of all these sub-branches (and all three are "sitting" on a single "stem" of the haplotype three in Fig. 1, which signifies the common ancestor for all the three sub-branches), we heed those "genetic distances" between the base haplotypes, listed above. The "genetic distance" between the base haplotypes for the three sub-branches and the base haplotype for the whole Donald branch is equal to 3.09. Therefore, 3.09/3/0.145 = 7 generations from the average "age" of all the three branches, which is 18 generations, results in 25 generations from the common ancestor of all the tree branches, that is 625 years before present. With a margin of error it will give **625±90** years to a common ancestor.

Similar calculations can be done pairwise. For example, the genetic distance between the first two sub-branches equals to 0.73+0.11+0.66 = 1.50. This corresponds to 1.50/0.145 = 10 generations between their common ancestors. Since their common ancestors lived 21 and 19 generations before present (see above), a common ancestor of the two sub-branches lived (10+21+19)/2 = 25 generations ago, that is the same 625 years calculated above, or about 1385AD.

This nicely corresponds to the records describing years of the life and death of John Lord of the Isles (d. 1386, which is 624 years ago).

In order to further examine the Donald branch we have composed a haplotype tree for this branch only, but with 25-marker haplotypes (Fig. 2). The purpose of this was to verify that the branch technically has only one common ancestor (technically, because in reality it might have been two or more individuals with the same ancestral haplotype, such as a father and his son, or two brothers, etc.). In other words, we intended to verify that the accumulation of mutations in haplotypes and disappearance of base haplotypes in a timespan between the common ancestor and present time followed the first-order kinetics[14]. In that case a number of base haplotypes in the branch and a number of mutations in the same branch should match each other via the first-order kinetic equations (see the Materials and Methods section), that is give the same number of generations to the common ancestor:

$M/N/k = t,$

and

$[\ln(N/A)]/k = t$

where:

N = a total number of haplotypes in the dataset,

A = a number of unchanged (identical, not mutated), that is base haplotypes in the set,

k = an average mutation rate,

t = a number of generations to a common ancestor.

Fig. 2 shows that the Donald branch contains 20 base haplotypes out of 46 (they look like identical bars on the very top of the haplotype tree).

As was indicated above, all 49 of 25 marker haplotypes contain 45 mutations, which gives 45/49/0.046 = 20 generations to a common ancestor. Since there are 20 base haplotypes in the branch (Fig. 2), we have [ln(49/20)]/0.046 = 19.5 generations to a common ancestor. These are practically identical figures. This means that technically (or truly) all 49 haplotypes in the branch descended from one common ancestor. He lived approximately **500±90** years ago using calculations based on 25 marker haplotypes. However, with more informative 67 marker haplotypes the data shifts to **625±90** years before present. At any rate, the common ancestor lived relatively recently, compared with other common ancestors for other major branches of the tree in Fig. 1. We will show this later.

In order to further examine haplotypes of the Donald Clan, we have considered 103 R1a1 haplotypes of the Clan listed in the Clan's public website (Web Resources, 65 of them were listed in the 67-marker format. Most of them had a pair YCAII = 19, 21 (only two haplotypes have 21, 21 in those loci). In order to better see a position of the Donald Clan haplotypes along with other R1a1 67-marker Scandinavian haplotypes, a combined haplotype tree was

composed (Fig. 3). In this tree 103 of the Donald Clan haplotypes were intentionally mixed with 40 Scandinavian haplotypes (taken from the public YSearch database) also having YCAII = 19, 21.

The tree obviously splits into two branches. Again, the Donald Clan haplotypes form a relatively young, compact branch of 69 haplotypes. It is on the right-hand side and continues in the upper left side. Four non-Donald Scandinavian haplotypes joined them. The obviously older left-hand side branch of 36 haplotypes does not contain Donald Clan haplotypes at all.

The base haplotype of the older, left-hand side branch is as follows:

13 25 15 11 11 14 12 12 10 13 11 30 -- 15 9 10 11 11 23 14 20 32 12 15 15 16 --11 12 **19 21** 16 16 17 17 34 38 12 11 – 11 8 17 17 8 12 10 8 11 10 12 22 22 15 11 12 12 13 8 14 23 21 12 12 11 13 11 11 12 12

All 36 haplotypes contain 152 mutations from the first 25 markers of the base haplotype, which gives 152/36/0.046 = 92 generations without a correction for back (reverse) mutations, and 102 generations with the correction, that is **2550±330** years to a common ancestor. This base haplotype was first described in[15], and was coined "the Scandinavian base haplotype" to distinguish it from the "Old Scandinavian base haplotype" which has YCAII = 19, 23, and the respective common ancestor lived **4100±700** years before present. It is remarkable that this "Old Scandinavian base haplotype' is an ancestral one to both "Young Scandinavian base haplotype" (see below) and to a series of Central Asian (Kyrgyz) haplotypes, published in[10]. (NOTE: There was no comparison of the Kyrgyz haplotypes and Scandinavian haplotypes in[10]). Furthermore, the Young Scandinavian and the Central Asian haplotype branches (lineages) appear to be directly related.

The Donald Clan branch of 69 haplotypes in Fig. 3 contains 89 mutations in the first 25 markers from the following base haplotype:

13 25 15 11 11 14 12 12 10 14 11 31 -- 16 8 10 11 11 23 14 20 31 12 15 15 16 – 11 12 **19 21** 17 16 17 18 34 39 12 11 – 11 8 17 17 8 12 10 8 11 10 12 22 22 15 11 12 12 13 8 14 23 21 12 12 11 13 11 11 12 12

This is exactly the same base haplotype as found from the 49-haplotype right-hand side branch in Fig. 1. It is always impressive how stable base haplotypes are, even 67 marker ones, when determined accurately.

A common ancestor of the branch lived 89/69/0.046 = 28 generations (without correction), that is 29 generations with a correction, or **725±105** years to a common ancestor. This is within the margin of error with **625±90** years before present, obtained from data shown in Fig. 1. Since John Lord of the Isles died 624 years before present, these dates do not conflict with each other.

The base Donald Clan haplotype differs from the Scandinavian Base haplotype (see above) by 9 mutations in all 67 markers. It places their common ancestor at 2500 years before present. It means that the common ancestor of the Scandinavian haplotype, with the determined timespan **2550±330**, is the "upstream" ancestor of the Clan Donald haplotype. If John Lord of the Isles was the "father" of the Clan Donald haplotypes, as claimed, and in fact fits the DNA genealogy, then the ancestor of the Scandinavian series of haplotypes, who lived in the middle of the 1st millennium BC, was the "grandfather" of the Donald Clan.

The data obtained above, can be verified with the 37 marker haplotype tree of a larger dataset of 229 haplotypes (Fig. 4).

The 37-marker tree is almost double in size compared to the 116 haplotype tree in the 67 marker format. The Donald Clan branch in the 37 marker tree contains 83 haplotypes, also almost double compared to 49 haplotypes for the branch in Fig. 1. However, that increase in the number of haplotypes did not bring any significant changes in the principal results and conclusions. Among these 83 haplotypes, 54 belonged to the Donald Clan members (65%), and the rest is occupied by other Highland Clansmen as described above, among them MacLeods, McAlisters, McNeils, McDaniels, Frasers, Douglases, Mathesons/Mathiesons, McCains/McKaines, Johnstons, Perkins. While in the 67 marker haplotype tree, three sub-branches were found in the Donald branch; and in the less resolved 37 marker tree only two of them could be clearly identified. Both branches had a common ancestor with the following, base haplotype

13 25 15 11 11 14 12 12 10 14 11 31 -- **16** 8 10 11 11 23 14 20 31 12 15 15 16 – 11 12 19 21 17 16 17 **18** 34 **39** 12 11

which is exactly as that shown above for the Donald Clan branch in the first 37 markers of the 67 marker series, and slight mutational differences were observed in the same markers (marked above) between the two sub-branches. A smaller sub-branch of 13 haplotypes (in the 37 marker tree) had 15.00, 18.00, and 37.92 in the loci indicated above, and its 13 and 31 mutations from the respective 25 and 37 marker base haplotype gave a timespan to a common ancestor of **550±160** and **700±140** ybp, respectively. A larger sub-branch of 70 haplotypes (in the 37 marker tree) had 15.86, 18.24, and 38.49 in the same loci, and its 93 and 236 mutations from the respective 25 and 37 marker base haplotype gave a timespan to a common ancestor of **750±110** and **975±120** ybp, respectively. The mutational difference of 1.67 mutations on their 37 marker base haplotypes corresponds to 450 years difference between the two common ancestors, and identifies a time span to their common ancestor as 875-1060 ybp

plus-minus 170 years. Obviously, the larger branch is an upstream one (paternal) with respect to the smaller branch.

Moving clockwise from the Donald Clan branch, we see in Fig. 1 a quite separate and distinct branch of 8 haplotypes. It contains mainly haplotypes of Mathesons of Lochalsh (Scotland) and Mathesons from elsewhere, all descended from a common ancestor who apparently had the following haplotype:

13 25 16 10 11 14 12 12 10 14 11 31 -- 15 8 10 11 11 24 14 20 31 12 14 15 16 – 11 12 **19 23** 16 16 19 19 33 38 12 11 – 12 8 17 17 8 11 10 8 11 10 12 22 22 16 11 12 12 13 8 12 23 21 12 12 11 13 11 11 12 12

These 8 haplotypes contain 25 mutations in 37 marker haplotypes, and 37 mutations in 67 marker haplotypes, which gives **900±200** and **825±160** years, respectively, to their common ancestor, who lived indeed in the Somerled times, give or take a couple of centuries. However, their haplotypes are completely different from all other haplotypes of Highland Clansmen. They are typical Eastern European haplotypes. If we consider genetic distances between their base haplotype and the Donald Clan base haplotype, they differ by 18 mutations in the 67 markers. The difference from the "Scandinavian" base haplotype is close to it being 19 mutations. Furthermore, the Matheson haplotypes differ more significantly from the Central European R1a1 base haplotype, with as many as 27 mutations:

13 25 16 10 11 14 12 12 11 **13** 11 29 **-- 16 9** 10 11 11 **23** 14 20 **32** 12 **15** 15 **15** – 11 **11** 19 23 **17** 16 **18** 19 **35 40 14** 11 – **11** 8 17 17 8 11 10 8 **12** 10 12 **21** 22 **15 10** 12 12 13 8 **14 25** 21 **13** 12 11 13 11 11 12 **13**

The closest to the Mathesons is the Russian Plain haplotype (Rozhanskii and Klyosov, 2009), with only 16 mutations between them:

13 25 16 **11** 11 14 12 12 10 **13** 11 30 -- 15 **9** 10 11 11 24 14 20 **32** 12 **15** 15 16 – 11 11 19 23 16 16 **18** 19 **34 39 13** 11 – **11** 8 17 17 8 **12** 10 8 11 10 12 22 22 **15 10** 12 12 13 8 **14** 23 21 12 12 11 13 11 11 12 **13**

This places an "upstream" ancestor of the Mathesons haplotypes to the Russian Plain, at about 3600 years ago, in the middle of the 2nd millennium BC. It would be tempting to associate the Mathesons common ancestor to Vikings who originated in the Russian steppes; however, we do not have enough convincing evidence for this.

The same "Mathesons" branch in the 37 marker tree (Fig. 4) contains as many as 38 haplotypes, and now includes eight Campbells (out of 12), Perkins, seven Frasers (out of 14), four McLeods (out of 16), one McNeil (out of 12), Lamont, two members of the Donald Clan, and some others. Close to the tip of the branch on the very right hand side there is a very distinct branch with DYS388=10. This branch split from other R1a1 haplotypes (having predominantly DYS388=12) more than 5000 years before present, and has the following base haplotype in the 25 marker format[8]

13 25 16 10 11 14 12 10 10 13 11 30 -- 15 9 10 11 11 24 14 19 32 12 14 15 16

30 European haplotypes of this branch, described earlier[8], had a common ancestor who lived **3575±450** years before present. Let us see where the six Highland Clansmen with DYS388=10 and their related haplotypes fit there. Those six haplotypes in Fig. 4 have the following base haplotype in the first 25 markers:

13 25 16 10 11 14 12 10 10 13 11 29/30 -- 15 9 10 11 11 24 14 19 32 12 14 15 16

It is practically identical to the above base haplotype for DYS388=10, published earlier[8]. All six haplotypes have 28 and 59 mutations in the first 25 and 37 markers, which gave

2850±610 and **3050±500** years to their common ancestor. Clearly, it cannot be Somerled. This is an ancient lineage.

The residual 32 haplotypes in the branch have the following base haplotype

13 25 15 10 11 14 12 12 10 14 11 31 -- 15 9 10 11 11 24 14 20 31 12 15 15 16 – 11 12 **19 23** 16 16 18 19 34 38 12 11

All 32 haplotypes contain 162 and 369 mutations in the respective 25 and 37 markers, which give **3100±390** and **3675±410** years to a common ancestor of the branch. Apparently, the less discriminatory effect of 37 marker haplotypes, having a lower resolution compared with that of 67 marker haplotypes, combines several branches into one. The branch contains YCAII = 19, 23, which is a feature of old European haplotypes, with DYS19 = 15, which is more likely a Scandinavian characteristic, as well as DYS389 = 14, 31. Overall, the branch is a rather old one, and it is not related to the main subject of this study.

There are a few more small branches of haplotypes in the 67 marker tree (Fig. 1). For example the 8 haplotype branch consisting of almost exclusively Alexander and MacAlister families (AL and MA on the tree), in the very upper left "corner". All eight of them form a very tight branch, with a base haplotype

13 25 15 11 11 14 12 12 10 14 11 31 -- 15 8 10 11 11 23 14 20 31 12 15 15 16 – 11 12 **19 21** 17 16 17 18 34 38 11 12 – 11 8 17 17 8 12 10 8 11 10 12 22 22 15 11 12 12 13 8 14 23 21 13 12 11 13 11 11 12 12

and have only 14 mutations in all eight of 67-marker haplotypes. This gives 14/8/0.145 = 12 generations, or only **300±90** years to a common ancestor of both the Alexander and MacAlister haplotypes. Since their base haplotype differs by only five mutations from the

Donald Clan base haplotype, it places their common – with the Donald Clan – ancestor at about 960 years bp.

An interesting branch albeit a small one is formed by five haplotypes, among them MacEacharn. Their base haplotype was coined as the "Young Scandinavian haplotype" earlier[15], with a common ancestor who lived **1900±400** years before present, and for these five individuals it is as follows:

13 25 **15** 11 11 14 12 12 10 **14** 11 **31** -- 15 8 10 11 11 24 14 20 31 12 15 15 16 – 11 12 **19 21** 15 16 16 18 34 37 12 11 – 11 8 17 17 8 12 10 8 11 10 12 22 23 15 11 12 12 13 8 13 23 22 12 12 11 13 11 11 12 12

It carries characteristic features of the Young Scandinavian haplotype (marked), and all five haplotypes have 31 and 41 mutations in their 37 and 67 marker haplotypes. This gives **1850±380** and **1525±280** years to their common ancestor, who, hence, lived in the first half of the 1st millennium AD. As it will be shown below, this is the ancestral branch with respect to the Donald branch.

This branch contains 15 haplotypes in the 37 marker haplotype tree (Fig. 4), and it is located on the upper left-hand side. The branch contains haplotypes of MacEacharn and both MacPhersons in one sub-branch, along with one member of the Donald Clan (a common ancestor of the four lived **1250±330** ybp), another sub-branch includes haplotypes of five Alexanders (out of a total of six) along with two Douglas family haplotypes (a common ancestor of all six lived only **325±120** ybp), and yet another sub-branch of four MacAlisters (out of a total of eight), with a common ancestor of **275±140** years bp.

Base haplotypes of these three sub branches were as follows (their mutations from each other are shown in bold):

13 25 **16** 11 11 14 12 12 10 14 11 31 -- 15 8 10 11 11 23 14 20 **29** 12 15 15 16 – **10** 12 19 21 **16** 16 17 **19 35** 38 12 **11**

13 25 15 11 11 14 12 12 10 14 11 31 -- 15 8 10 11 11 23 14 20 31 12 15 15 16 – 11 12 19 21 17 16 17 18 34 38 **11** 12

13 25 15 11 11 14 12 12 10 14 11 31 -- 15 8 10 11 11 23 14 20 31 12 15 15 16 – 11 12 19 21 17 16 17 18 34 **39** 12 12

They coalescence to the following base haplotype which is almost indistinguishable from the Young Scandinavian haplotype[15]:

13 25 15 11 11 14 12 12 10 14 11 31 -- 15 8 10 11 11 23 14 20 31 12 15 15 16 – 11 12 19 21 17 16 17 18 34 38 12 12

All three base haplotypes (sub-branches) above contain collectively 11.25 mutations from their ancestral haplotype (Young Scandinavian, shown above with rounded up alleles), which brings their common ancestor to **1720±300** years before present (11.25/3/0.09 = 42 generations without a correction for back mutations, or 44 generations corrected, that is 1100 ybp down from 620 ybp, which is the averaged time spans to the common ancestors for all the three sub-branches). This **1720±300** ybp is within the margin of error with **1900±400** ybp obtained earlier for the Young Scandinavian ancestral haplotype[15].

The rest of the tree in Fig. 1 is occupied by a large and "fluffy" branch, indicating that it has a rather ancient common ancestor. The branch contains 43 haplotypes, with the base haplotype

13 25 16 11 11 14 12 12 10 13 11 30 -- 15 9 10 11 11 24 14 20 32 12 15 15 16 – 11 12 **19 23** 16 16 18 18 34 38 12 11 – 11 8 17 17 8 12 10 8 11 11 12 22 22 15 11 12 12 13 8 14 23 21 12 12 11 13 11 11 12 12

25, 37 and 67 marker haplotypes of the branch contained 253, 555 and 801 mutations, respectively, which gave **3700±440**, **4200±460**, and **3700±390** years to the common ancestor of the branch, on average **3870±460** years to a common ancestor. These are all MacLeods, all Cowans, all Frasers, most of the Campbells, most of the Douglas family, half of the Perkins (another half shares the branch with the Donald Clan), most of the Ulster Heritage Ireland group, one MacNeill (another one is in the Donald branch), Lamont, McCain, etc. All of them descended from an ancient common ancestor, who lived in the 2nd millennium BC. Most of them have the common European pair YCAII = 19, 23, but some – within the same family – carry YCAII = 19, 21. Clearly, that mutation has appeared between about 3900 and 2600 years before present. The Mathesons likely descended from this branch. Their base haplotype of 825 years "old" (in the 67 marker format) has as many as 14 mutations from the above base haplotype, which places a common ancestor of the Mathesons and the above listed clans at about 3700 years before present.

The last haplotype is close to the "Old Scandinavian Haplotype"[15]:

13 25 15 11 11 14 12 12 10 13 11 30 -- 15 9 10 11 11 24 14 20 32 12 15 15 16 – 11 12 **19 23** 16 16 18 19 35 38 13 11 – 11 8 17 17 8 12 10 8 11 10 12 22 22 15 10 12 12 13 8 13 23 21 12 12 11 13 11 11 12 13

and has only one mutation from it in the first 25 markers, 4 mutations in the first 37 markers, and 8 mutations in all 67 markers. Their common ancestor lived about 4700 years before present. However, this haplotype is even closer to the Russian Plain base haplotype

13 25 16 11 11 14 12 12 10 13 11 30 -- 15 9 10 11 11 24 14 20 32 12 15 15 16 – 11 11 19 23 16 16 18 19 34 39 13 11 – 11 8 17 17 8 12 10 8 11 10 12 22 22 15 10 12 12 13 8 14 23 21 12 12 11 13 11 11 12 13

from which it does not have mutations in the first 25 markers, has only 3 mutations in the first 37 markers, and 6 mutations in all 67 markers. It places a common ancestor of that 43-haplotype branch to about 4100 years before present. These dates, between 4000 and 4700 ybp, corresponds to a common ancestor of the major part of the population of R1a1 in Europe[15].

One can see that the Old Scandinavian base haplotype differs from the Russian Plain base haplotypes by four mutations in all 67 markers. This corresponds to 4/67/0.00216 = 28 generations, that is 700 years between their common ancestors; with a slight correction for reverse mutations it will make 725 years. This fits well to 4100 and 4800 years for those common ancestors, plus-minus margins of error, as was listed above. Hence, the Old Scandinavian ancestral haplotype is indeed a downstream one from the Russian Plain ancestral haplotype. A more detailed consideration has shown that the Old Scandinavian branch might be a superposition of two "sister" R1a1 branches of similar "ages", however, this is not important to the main goals of this particular study.

In the 37 marker tree (Fig. 4) this "fluffy" 89-haplotype branch on the left-hand side did not bring anything really new compared to the same 43-haplotype branch in the 67-marker tree (Fig. 1). The 89-haplotype branch has the following base haplotype:

13 25 16 11 11 14 12 12 10 14 11 31 -- 15 9 10 11 11 24 14 20 31 12 15 15 16 – 11 11 19 23 15 16 18 18 34 39 12 11

All 89 haplotypes had 480 and 1015 mutations in their 37 and 67 markers, which gave **3350±370** and **3650±380** years to their common ancestor. Again, this haplotype had some features of both the Eastern European base haplotypes, such as DYS19 = 16, YCAII = 19, 23, and the Scandinavian base haplotypes, such DYS389 = 14, 31. Apparently, 37 marker haplotypes do not possess as good resolving capability as that of the 67-marker haplotypes.

A 25-marker haplotype tree has even lower resolving capability. The 253-haplotype tree is shown in Fig. 5 just as an illustration, not for calculations. As one can see, haplotypes in a shorter, 25-marker format, are often reduced to the same, identical haplotypes. Differences between them start to appear only in the subsequent panels of markers, in the 26-37 and 38-67 marker panels.

However, even in that case a "short haplotype" tree can be useful for approximate calculations. For example, let us consider two sets of "base" haplotypes, one of which has 22 of them, and located on the very top of the tree. This base haplotype

13 25 15 11 11 14 12 12 10 14 11 31 -- 15 8 10 11 11 23 14 20 31 12 15 15 16

is related to the whole tree. Therefore, it does not mean much in this case, because the tree has a number of "the most recent common ancestors" as it was shown in this study. The shown base haplotype is just the prevalent one, but still is a part of the mix. By applying the logarithmic method we obtain [ln(253/22)]/0.046 = 53 generations not corrected for back mutations, or 56 generations corrected, that is 1400 years to a phantom "common ancestor". A phantom – because we know that the tree as a whole is much "older", as some haplotypes in it are derived from common ancestors who lived as long ago as 4000 years before present and earlier, and the tree does not have just one common ancestor. It is, indeed a phantom object.

However, unlike in the case of the whole tree, we do know that the Donald Clan branch has one common ancestor, allegedly John Lord of the Isles. We know that the most recent common ancestor of that branch lived between 25 and 29 generations before present (625 to 725 years ago, plus-minus the margin of error). This branch on the right-hand side in the 25-marker tree contains 112 haplotypes, and 34 of them are base, identical haplotypes

13 25 15 11 11 14 12 12 10 14 11 31 -- 16 8 10 11 11 23 14 20 31 12 15 15 16

Indeed, they are identical with the first 25 markers in the 37 and 67 marker Donald Clan base haplotypes, obtained from the 67 and 37 marker trees. This by itself is a rather trivial observation, and it would be surprising not to observe it. What is not trivial, and is in line with the procedure explained in detail in this study, is that by employing the logarithmic method we obtain [ln(112/34)]/0.046 = 26 generations to the Clan Donald common ancestor, which is 650 years before present. This is a complete fit with the data obtained previously after tedious calculations of mutations in the branch. Here we needed just to calculate a number of identical haplotypes in the branch, and we achieved the same result.

DISCUSSION

There are at least two important subjects to discuss in this study. One is an employment of DNA for genealogical studies, and another is the results of the study itself with respect to Somerled and his alleged contribution to Highland Clansmen Y-chromosomal haplotypes.

1. DNA Genealogy.

This is apparently the first study in the new area of DNA genealogy, in which detailed analysis of extended haplotypes was performed using as extended as a 67 marker haplotype tree for separating branches (lineages) of a mixed population, and employing the mutation rate constants for calculations of timespans to the most recent common ancestors of the identified lineages – both by counting mutations and without it, from a number of the base haplotypes only. The study showed that 67 marker haplotypes can be successfully used for such a study, while 37 marker ones do not possess (or do not always possess) a good

resolving capability to provide results as reproducible as those obtained with more extended haplotypes.

It was also shown that 25, 37 and 67 marker haplotypes of the resolved lineages commonly provide quite reproducible and mutually compatible data. This is despite the fact that 25 and 37 marker haplotypes do not contain as many as 42 and 30 additional markers compared to 67 marker haplotypes. In other words, once lineages are resolved, calculations can be conducted almost equally well with haplotypes of different length. Essentially, 67 marker haplotypes are needed not that much for the calculations, but for initially resolving branches in a haplotype tree. However, a higher number of alleles available for analysis, provided by more extended haplotypes, noticeably reduces the margins of error of an obtained timespan to a common ancestor of the population (branch, lineage).

A reasonably good fit between time spans obtained from 25, 37 and 67 marker haplotypes illustrates that the average mutation rate constants employed in this study (and obtained earlier using calibration of historical data and extended genealogies, see Klyosov 2009a,b) can be successfully used for DNA-genealogical studies.

The fact that the time span to a common ancestor for the Donald Clan members consistently points to between 625 and 725 years before present (plus-minus a margin of error, such as 625±90 and 725±105 years with 95% confidence), while the records show that John Lord of the Isled died 624 years before present (d. 1386), serves as a calibration itself.

The fact that the base haplotypes obtained from resolved branches (lineages) in 67 and in sometimes 37 marker haplotype trees were reproducible not only within and across this study, but also compared favourably with published data, illustrates that the method is robust and has a reliable quantitative ground.

As a brief summary, the method of haplotypes analysis involved the following principal steps:

(a) Composition of a haplotype tree based on a haplotype dataset,

(b) Dissection of the obtained haplotype tree onto branches (lineages),

(c) Identification of a base haplotype for each branch as a maximally equidistant one from haplotypes in the branch. In practice a base haplotype (or a base allele) has a minimal number of mutations from all other alleles in the same marker,

(d) Counting the total number of mutations in a branch,

(e) Calculating an average number of mutations per haplotype (or per marker in the haplotype),

(f) Dividing the average number of mutations per haplotype (or per marker in the haplotype) by the mutation rate constant for the haplotype (or for the marker),

(g) The mutation rate constants for 25, 37 and 67 marker haplotypes are equal to 0.022, 0.046 and 0.145 mutation per haplotype per generation of 25 years,

(h) The mutation rate constants for 25, 37 and 67 marker haplotypes are equal to 0.00183, 0.00243 and 0.00216 mutation per marker per generation of 25 years,

(i) If a timespan to a common ancestor exceeds 25 generations (625 years), a correction for reverse (back) mutations should be introduced; the respective table is given in[14],

(j) A margin of error is recommended to be introduced into the final timespan to a common ancestor; a method for the calculation of margins of error is given in[14]; the method is based on a total number of mutations in the branch in a haplotype tree, or in

a dataset, provided that all haplotypes in the dataset are technically descended from the same most recent common ancestor,

(k) There are several criteria in order to determine whether all haplotypes in the dataset descended from the same common ancestor; one criterion is that the haplotype tree does not consist of two or several distinct branches. However the most direct and the most reliable criteria is that a timespan to a common ancestor, calculated as described above (the "linear" method, based on an average number of mutations per haplotype or per marker, that is the method based on mutation counting) should be equal (well within the margin of error) to a timespan to the common ancestor determined by the logarithmic method,

(l) The logarithmic method does not operate with mutation counting, it counts a number of base haplotypes in the dataset (see Fig. 2 in this study and explanations in the text and in the next item)

(m) The linear (mutation-counting) method employs the formula $M/N/k = t$, where N is the total mutation count in a haplotype dataset descended from the same most recent common ancestor, N is the number of haplotypes in the dataset, k is the mutation rate constant, and t is the timespan to the common ancestor (in generations), not corrected for back mutations (if more than 25 generations)

(n) The logarithmic (base haplotypes counting) method employs the formula $[\ln(N/n)]/k = t$, where n is a number of base haplotypes in the dataset (or in the branch of a haplotype tree),

(o) t's (a number of generations) should be approximately the same, when calculated using the linear (item m above) and logarithmic (item n above) methods.

The items listed above essentially describe a basis of DNA genealogy, which is a blend of the DNA sequencing (identification of [preferably extended] haplotypes and haplogroups) and chemical kinetics.

2. An alleged Somerled genealogy

To sum up the results of this study, a common ancestor of the most of European (including West, Central and Eastern Europe) R1a1 haplotypes lived around 4800 years ago. He had the following haplotype[15]:

13 25 16 11 11 14 12 12 10 13 11 30 -- 15 9 10 11 11 24 14 20 32 12 15 15 16 – 11 11 19 23 16 16 18 19 34 39 13 11 – 11 8 17 17 8 12 10 8 11 10 12 22 22 15 10 12 12 13 8 14 23 21 12 12 11 13 11 11 12 13

His descendant some 700 years later was a carrier of the Old Scandinavian haplotype:

13 25 **15** 11 11 14 12 12 10 13 11 30 -- 15 9 10 11 11 24 14 20 32 12 15 15 16 – 11 12 19 23 16 16 18 19 **35 38** 13 11 – 11 8 17 17 8 12 10 8 11 10 12 22 22 15 10 12 12 13 8 **13** 23 21 12 12 11 13 11 11 12 13

During those 700 years his haplotype had acquired four mutations (shown in bold), one of them, DYS19 = 16→15, is a very characteristic one for contemporary Scandinavian haplotypes. However, YCA II were still equal to 19, 23, more typical to East European and Central European haplotypes, unlike 19, 21 in many Scandinavian haplotypes. Theoretically, four mutations randomly accumulated in the 67 markers correspond to a 725 year time span. That is what we see in reality. It is not clear yet where a carrier of the Old Scandinavian haplotype lived, either still on the Russian Plain, or he had already moved to the Northern Europe in the course of those 700 years.

Another descendant of the Russian Plain haplotype, a thousand years later (**3870±460** ybp) after him had the following haplotype (mutations are marked), in which another set of mutations randomly happened:

13 25 16 11 11 14 12 12 10 13 11 30 -- 15 9 10 11 11 24 14 20 32 12 15 15 16 – 11 12 19 23 16 16 18 **18** 34 **38 12** 11 – 11 8 17 17 8 12 10 8 11 **11** 12 22 22 15 **11** 12 12 13 8 14 23 21 12 12 11 13 11 11 12 **12**

During those thousand years the haplotype had accumulated six mutations in 67 markers, which indeed corresponds to about a thousand year period (6/0.145 = 41 generations without a correction for back mutations, and 43 generations with a correction, that is 1075 years). Among descendants of that ancestor were the MacLeods, the Cowans, the Frasers, the Campbells, most of the Douglas family, many Perkins, most of the Ulster Heritage Ireland group, some MacNeills, Lamonts, McCains, etc. This haplotype still had a characteristic East European pair 19, 23 at YCAII, DYS458=15 and DYS19=16. Hence, it can hardly be called a typical "Scandinavian" haplotype. Data shows that 3870 ±460 years before present carriers of those haplotypes and their nearest mutations were still on the Russian Plain, between the Carpathian Mountains, Southern Black Sea steppes, Central Asia further on the Eastern South, and the Ural mountains on the East.

A thousand more years later, at **2550±330** years before present, the "Scandinavian" base haplotype appeared in Europe:

13 25 **15** 11 11 14 12 12 10 13 11 30 -- 15 9 10 11 11 **23** 14 20 32 12 15 15 16 --11 12 19 **21** 16 16 **17 17** 34 38 12 11 – 11 8 17 17 8 12 10 8 11 **10** 12 22 22 15 11 12 12 13 8 14 23 21 12 12 11 13 11 11 12 12

It was closer to the above haplotype (6 and 7 mutations in 37 and 67 markers, in bold), compared to the Russian Plain base haplotype (9 and 11 mutations, respectively). Furthermore, it had acquired YCAII = 19, 21, typical for many recent Scandinavian haplotypes. These were the times of re populating Europe by R1a1 tribes, in the middle of the 1st millennium BC. Those tribes came from the Russian Plain, from Central Asia, from the Baltic region, from the North to the Black Sea; among them were the Scythians, Ants, Venets, Sklavens, and other Slavic tribes, including, possibly the Scotichi. Alternatively, a carrier of the above Scandinavian base haplotype was not among them, but descended from the North European Old Scandinavian haplotype carrier, who lived there since about 4100 years before present, that is some 1600 years earlier plus-minus a margin of error of some 300-400 years. There are close to nine mutations between their base haplotypes (if to calculate without rounding them up), which corresponds to 1650 years between them. Hence, the "Scandinavian" haplotype might well be a direct descendant from the Old Scandinavian haplotype, either on the Russian Plain, or in the North Europe, or on it its way to Europe.

Around 1900±400 years before present the "Young Scandinavian" base haplotype has appeared[15]:

13 25 15 11 11 14 12 12 10 **14** 11 31 -- 15 9 10 11 11 23 14 20 32 12 15 15 16 --11 12 19 21 16 16 17 17 34 38 12 11 – 11 8 17 17 8 12 10 8 11 10 12 22 22 15 11 12 12 13 8 14 23 21 12 12 11 13 11 11 12 12

It had only one mutation, in DYS389-1 = 14 (the allele in DYS389-2 = 17 in the both base haplotypes), compared to the Scandinavian base haplotype. It is essentially the same base haplotype, and their common ancestors present the same person (indeed, "ages" of these two haplotypes are within the margin of error), or the descendant had – by statistical chance – the same haplotype. A similar base haplotype is observed in the branch that includes haplotypes

of Campbell, MacEacharn, and a descendant of Sir David Home (b. 1382). Since their branch contains only five haplotypes, a base haplotype is a little distorted, but preserves the main features of the "Young Scandinavian" base haplotype (see above), among them DYS19 = 15, DYS389 = 14,31, YCAII = 19,21.

This haplotype (**1850±380** and **1525±280** years to the common ancestor, see above) is ancestral to the Donald Clan haplotypes (**625±90** to **725±105** years to a common ancestor)

13 25 15 11 11 14 12 12 10 14 11 31 -- **16 8** 10 11 11 23 14 20 **31** 12 15 15 16 – 11 12 19 21 **17** 16 17 **18** 34 **39** 12 11 – 11 8 17 17 8 12 10 8 11 10 12 22 22 15 11 12 12 13 8 14 23 21 12 12 11 13 11 11 12 12

They are separated in time by approximately a thousand years, and have six mutations between them. As it was shown above, six mutations per the 67 markers correspond to about 1075 years between common ancestors of the populations.

As a result of this study, we traced a DNA-genealogical sequence between the ancestral haplotype of the Russian Plain some 4800 years before present, from which most of the Highland Clansmen families descended, then the Young Scandinavian ancestral haplotype in turn descended around 2300-1900 years before present. In the first half of the 1st millennium AD another branch has descended from the latter (MacEacharns and some Campbells belong to that branch), from which John Lord of the Isles had apparently descended in the 14th century, and haplotypes of his apparent descendants point to the time period of his life between 625±90 and 725±105 years ago. He and/or his closest relatives were ancestor(s) of many Highland Clansmen, among them the Clan Donald multiple members, some Perkins, some MacCains, McDaniels, McAllisters, McInnes, McNeills, Douglas, Mathesons, and some other families. All (or part) of them descended from the same common ancestor, whose haplotype was as follows:

13 25 15 11 11 14 12 12 10 14 11 31 -- 16 8 10 11 11 23 14 20 31 12 15 15 16 – 11 12 19 21 17 16 17 18 34 39 12 11 – 11 8 17 17 8 12 10 8 11 10 12 22 22 15 11 12 12 13 8 14 23 21 12 12 11 13 11 11 12 12

and he lived in the same time period, around **625±90** years before present.

As one can see, there is no unambiguous room for Somerled in the above considerations and conclusions. The most recent common ancestors, identified in this study with Highland Clansmen, form a sequence starting with the most recent lineages of some 300 years before present (they represent offshoots from older lineages), some of them, including the Donald Clan, lead down to **625±90** to **725±105** ybp, allegedly John Lord of the Isles, then to the Young Scandinavian ancestral haplotype of **1900±400** ybp, then to Scandinavian ancestral haplotype of **2550±330** ybp (the two last ones can be the same thing), then to the Old Scandinavian ancestral haplotype of **4100±700** ybp, and finally to the ancestral R1a1 haplotype of the Russian Plain of **4750±500** ybp (Klyosov 2009b). The last two are also overlapping within the margin of error, and a bearer of the Old Scandinavian haplotype could have resided on the Russian Plain.

It is not the most ancient R1a1 base haplotype, however. It is merely (but importantly) the most ancient haplotype for the majority of the European R1a1 haplotypes. Some R1a1 haplotypes of a common ancestor of 10,000 to 12,000 ybp were identified in Europe[8,15], and even more ancient, to 12,000 - 21,000 ybp in Asia[8]. However, their presence in Europe is minimal, no more than one per cent in the population.

Web Resources

http://www.electricscotland.com/history/articles/norse.htm

http://dna-project.clan-donald-usa.org/

http://www.clandonald-heritage.com/

http://www.clanmacalistersociety.org/historians-corner/article1.htm

http://www.macdougall.org/heritage.html

http://macinnes.org/info.php

http://en.wikipedia.org/wiki/MacInnes

http://en.wikipedia.org/wiki/MacDonald_of_Ardnamurchan

http://en.wikipedia.org/wiki/Scoti

http://www.royal.gov.uk/HistoryoftheMonarchy/Scottish%20Monarchs(400ad-1603)/TheearlyScottishMonarchs/TheearlyScottishmonarchs.aspx

http://www.archaeology.org/0107/abstracts/scotland.html

http://macinnes.org/info.php

http://en.wikipedia.org/wiki/Clan_MacAlister

http://en.wikipedia.org/wiki/John_of_Islay,_Lord_of_the_Isles

http://en.wikipedia.org/wiki/Somerled

http://thepeerage.com/p2457.htm

References

1 Skene WF: Celtic Scotland A History of Ancient Alban, 1890; 33.

2 Donald MacDonald: Clan Donald. Pelican Publishing Company Inc., Louisiana, 2008, p.24

3 MacKenzie A: History of the MacDonalds and Lords of the Isles with Genealogies of the Principal Families of the Name. Kessinger Publishing, LLC, Whitefish, Montana, 2007, p. 2

4 The Book of Veles. Translation from the proto-Slavic language by Georgy Z. Maximenko. ISBN 978-5-91047-001-3, Moscow, 2008, 370 pp (in Russian)

5 Clackson J: Indo-European Linguistics. An introduction (Cambridge University Press, New York); 2007: pp. 198, 207.

6 Neil J, Ritchie G, Ritchie A: Scotland Archaeology and Early History. Thames and Hudson, London; 1991, p. 154

7 Donald MacDonald: Clan Donald. Pelican Publishing Company Inc. Louisiana; 2008, p. 8

8 Klyosov AA: DNA Genealogy, mutation rates, and some historical evidences written in Y-chromosome. II. Walking the map. J Genet Geneal 2009; **5**: 217-256.

9 Bittles AH, Black ML, Wang W: Physical anthropology and ethnicity in Asia: the transition from anthropology to genome-based studies. J Phys Anthropol 2007; **26**:77-82.

10 Underhill PA, Myres NM, Rootsi S *et al*: Separating the post-Glacial coancestry of European and Asian Y chromosomes within haplogroup R1a. Europ J Human Genet 2010; **18**: 479-484.

11 Keyser-Tracqui C, Crubezy E, Ludes B: Nuclear and mitochondrial DNA analysis of a 2,000-year-old necropolis in the Egyin Gol Valley of Mongolia. Am J Hum Genet 2003; **73**:247-260.

12 Kim K, Brenner CH, Mair VH *et al*: A western Eurasian male is found in 2000-year old elite Xiongnu cemetery in Northeast Mongolia. Amer J Phys Anthropol 2010, on-line publication, DOI 10.1002/ajpa.21242

13 Li C, Li H, Cui Y *et al*: Evidence that a West-East admixed population lived in the Tarim Basin as early as the early Bronze Age. BMC Biology 2010; on-line publication, **8**:15, doi:10.1186/1741-7007-8-15

14 Klyosov AA: DNA Genealogy, mutation rates, and some historical evidences written in Y-chromosome. I. Basic principles and the method. J Genet Geneal 2009a; **5**:186-216.

15 Rozhanskii I, Klyosov AA: Haplogroup R1a1: its haplotypes, lineages, history, and geography. Proc Russian Academy of DNA Genealogy (ISSN 1942-7484) 2009; **2**: 974-1099.

16 Klyosov AA: A comment on the paper: Extended Y chromosome haplotypes resolve multiple and unique lineages of the Jewish Priesthood. Human Genetics 2009c; **126**:719-724.

17 Felsenstein J: Phylip, the Phylogeny Inference Package. PHYLIP, version 3.6. Department of Genome Sciences, University of Washington, Seattle 2005.

Titles and legends to figures

Figure 1. 67 marker haplotype tree for the "Highland Clansmen", haplogroup R1a1, and their allegedly related haplotypes. This 116 haplotype tree was composed of haplotypes obtained from public projects, as described in the text. The tight, compact, hence "young" branch on the right encompasses most of the Donald clan haplotypes (30 of them are on this branch, out of a total 49 haplotypes (61%) of the Highland Clansmen; the rest belongs to the Perkins (two haplotypes out of six in their family), MacCain (seven haplotypes out of eight in their family), three McDaniels, two McAllisters (out of four in their family), McInnes, McNeil, Douglas, Matheson. This "young", mainly Donald Clan branch descends from a common ancestor who lived 625±90 years before present. The "fluffy" and much more extended, therefore "old" branch on the left contains 43 haplotypes, descended from a common ancestor who lived 3870±460 years before present. A small separate branch of 8 haplotypes (at 5 o'clock) contains mainly haplotypes of Mathesons of Lochalsh (Scotland) and Mathesons from elsewhere, descended from a common ancestor who lived 900±200 and 825±160 ybp (calculated with 37 and 67 markers of the same haplotypes, respectively), however, has a quite a different origin than the rest of the R1a1 Highland Clansmen.

Figure 2. 25 marker haplotype tree for the right-hand side branch of 49 haplotypes in Fig. 1. The base haplotypes, that is apparent ancestral haplotypes, appear as a "comb" in the upper part of the tree. There are 20 of them in this particular case.

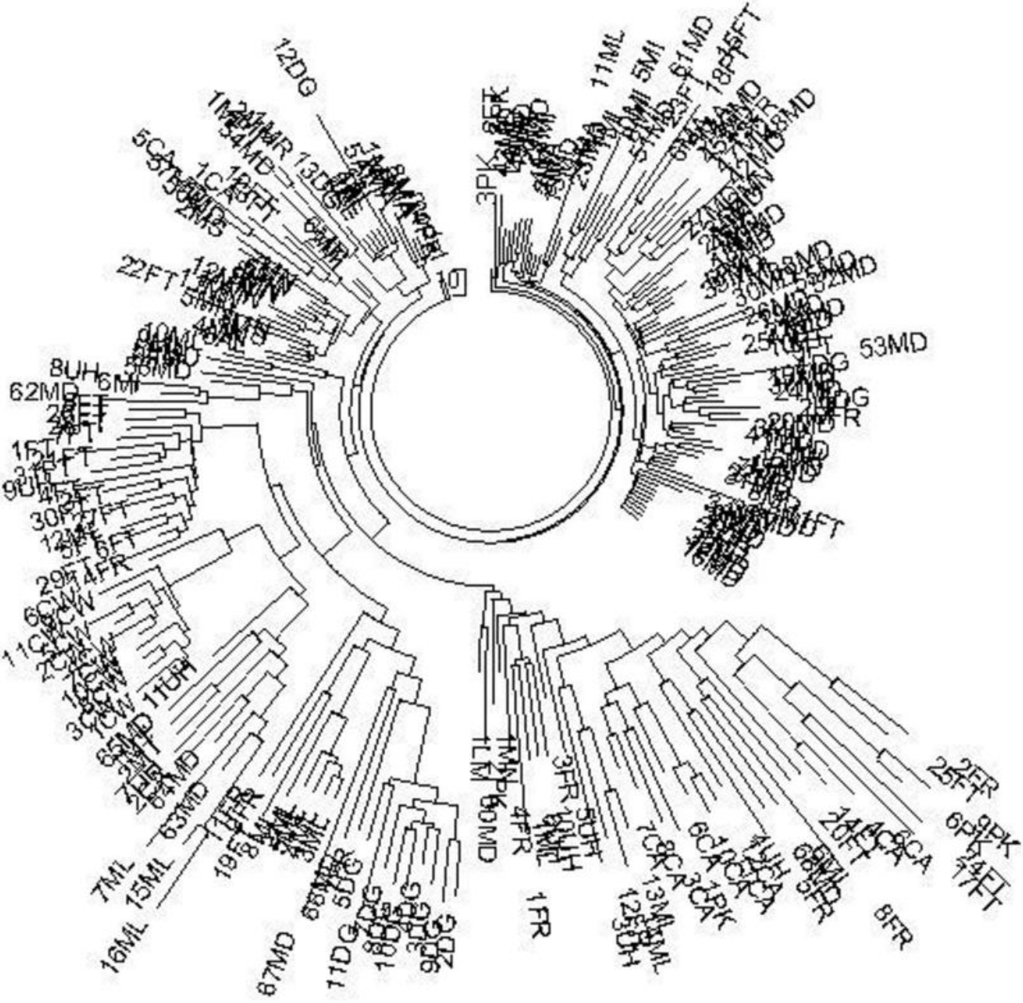
Figure 3. 67 marker haplotype tree for 65 Donald Clan haplotypes of haplogroup R1a1, combined with 40 of 67 marker R1a1 Scandinavian haplotypes having YCAII = 19,21. The

Scandinavian haplotypes were taken from YSearch public database, the Donald Clan haplotypes (all marked mc) were taken from the public site (Web Resources).

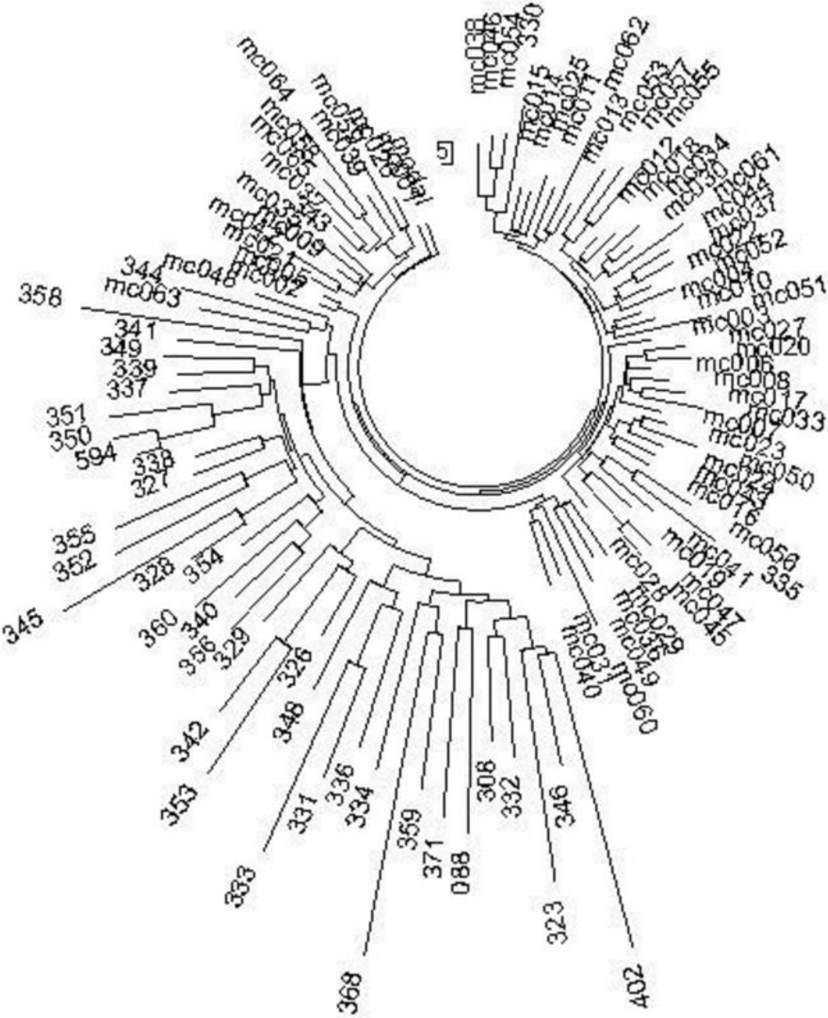
Figure 4. 37 marker haplotype tree for the "Highland Clansmen", haplogroup R1a1, and their allegedly related haplotypes. This 229 haplotype tree was composed of haplotypes obtained from public projects, as described in the text. The tight Donald Clan branch is in on the right-hand side. The older, fluffy branches on the left are "Old Scandinavian" and East European branches.

Figure 5. 25 marker haplotype tree for the "Highland Clansmen", haplogroup R1a1, and their allegedly related haplotypes. This 253 haplotype tree was composed of haplotypes obtained from public projects, as described in the text. The tight Donald Clan branch in on the right-hand side. The older, fluffy branches on the left are "Old Scandinavian" and East European branches.

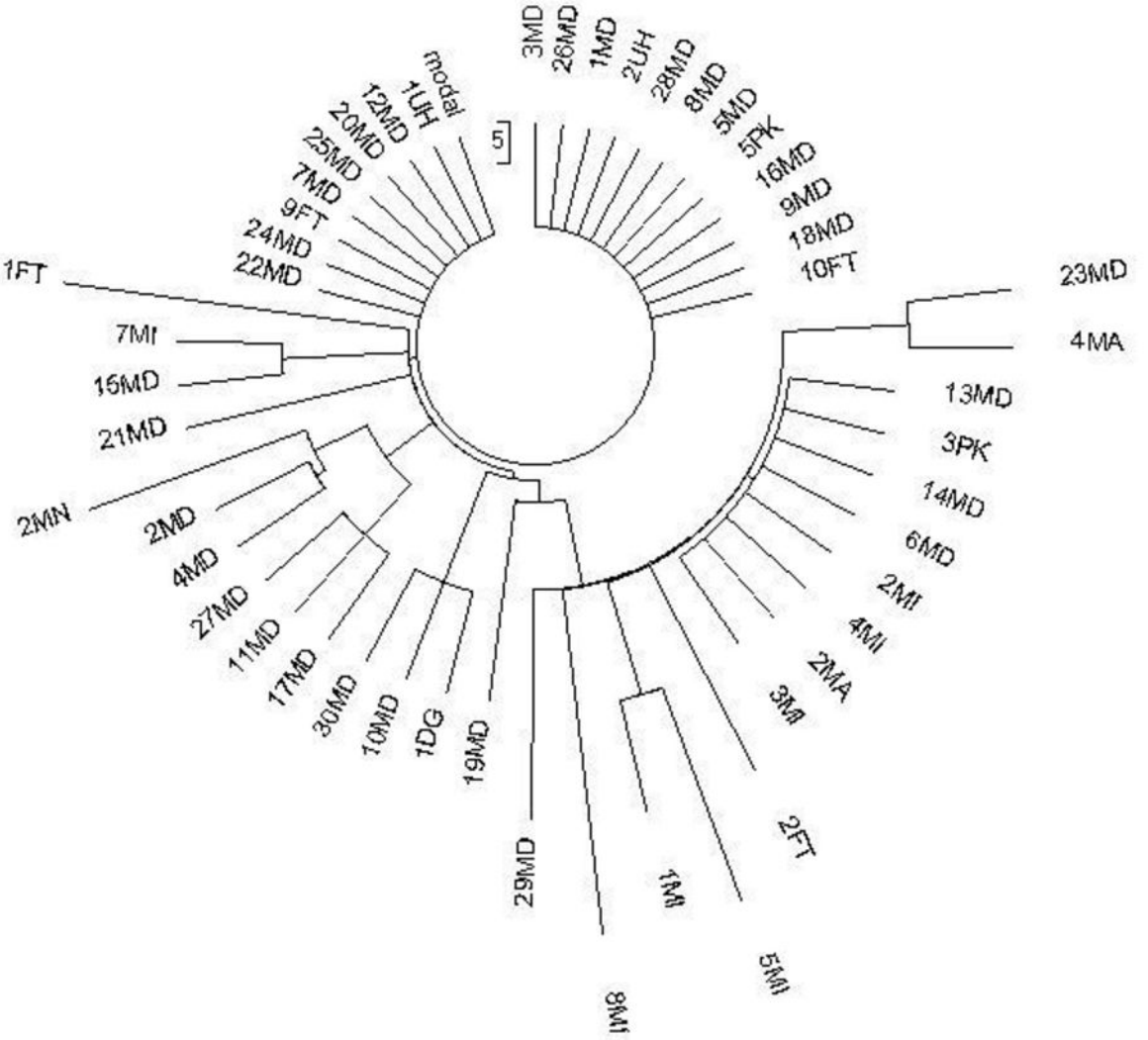
61MD
7ML
16ML
8FR
20MD

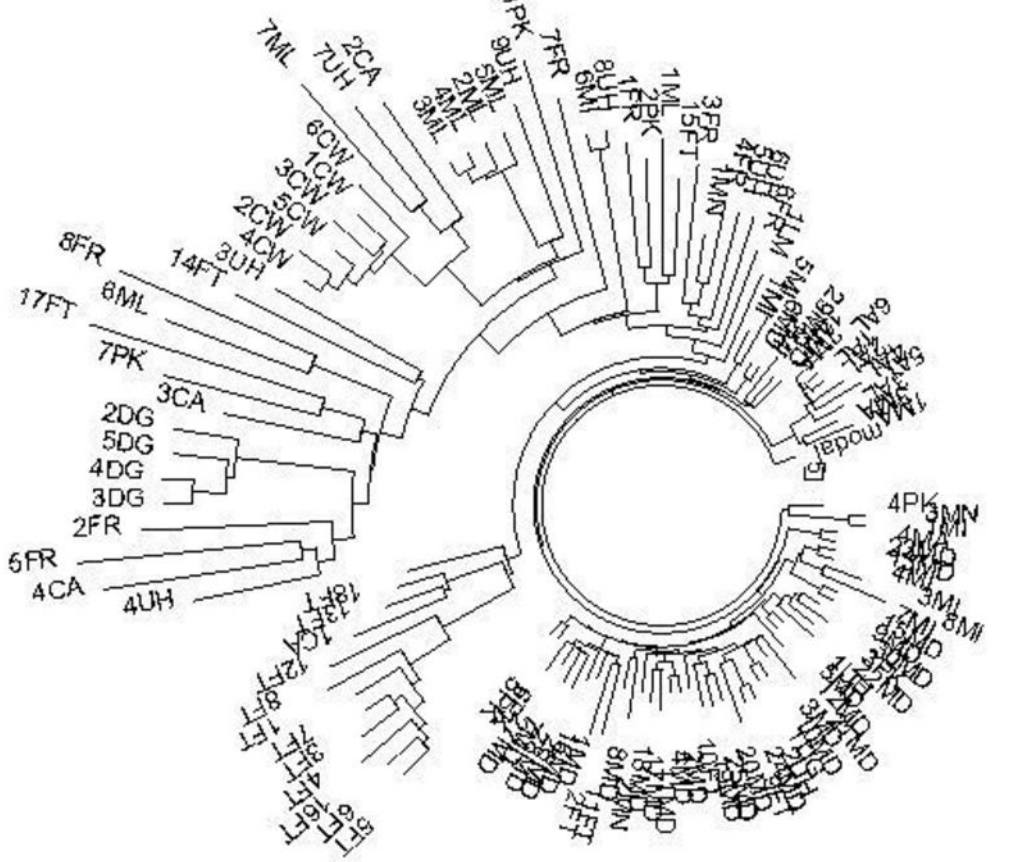


12DG
11ML
5MI
61MD
53MD
62MD
8UH
22FT
11UH
63MD
7ML
15ML
16ML
67MD
11DG
1FR
8FR
2FR
25FT
6PK
9PK
17FT



358
341
349
339
337
351
350
594
338
327
355
352
328
354
345
360
340
356
329
326
342
353
348
336
334
331
333
359
371
088
308
332
346
323
402
368
mc063
mc064
mc062
mc015
mc055
mc051
mc027
mc020
mc006
mc008
mc017
mc033
mc023
mc050
mc056
mc016
335
mc045
mc047
mc060
mc040
mc003


3MD
26MD
1MD
2UH
28MD
8MD
5MD
5PK
16MD
9MD
18MD
10FT
23MD
4MA
13MD
3PK
14MD
6MD
2MI
4MI
2MA
3MI
2FT
5MI
1MI
8MI
29MD
19MD
1DG
10MD
30MD
17MD
11MD
27MD
4MD
2MD
2MN
21MD
15MD
7MI
11FT
22MD
24MD
9FT
7MD
25MD
20MD
12MD
1UH
modal
5


7ML
7UH
2CA
1PK
9UH
5ML
2ML
4ML
3ML
7FR
8UH
6ML
1ML
2PK
1FR
3FR
15FT
6CW
1CW
3CW
5CW
2CW
4CW
3UH
14FT
8FR
17FT
6ML
7PK
3CA
2DG
5DG
4DG
3DG
2FR
5FR
4CA
4UH
4PK
3ML
8ML