The phylogenealogy of R-L21:

four and a half millennia of expansion and redistribution

Joe Flood*

* Dr Flood is a mathematician, economist and data analyst. He was a Principal Research Scientist at CSIRO and has been a Fellow at a number of universities including Macquarie University, University of Canberra, Flinders University, University of Glasgow, University of Uppsala and the Royal Melbourne Institute of Technology. He was a foundation Associate Director of the Australian Housing and Urban Research Institute. He has been administrator of the Cornwall Y-DNA Geographic Project and several surname projects at FTDNA since 2007. He would like to give credit to the many 'citizen scientists' who made this paper possible by constructing the detailed R1b haplotree over the past few years, especially Alex Williamson.

ABSTRACT: Phylogenealogy is the study of lines of descent of groups of men using the procedures of genetic genealogy, which include genetics, surname studies, history and social analysis.

This paper uses spatial and temporal variation in the subclade distribution of the dominant Irish/British haplogroup R1b-L21 to describe population changes in Britain and Ireland over a period of 4500 years from the early Bronze Age until the present. The main focus is on the initial spread of L21-bearing populations from south-west Britain as part of the Beaker Atlantic culture, and on a major redistribution of the haplogroup that took place in Ireland and Scotland from about 100 BC.

The distributional evidence for a British origin for L21 around 2500 BC is compelling. Most likely the mutation originated in the large Beaker colony in south-west Britain, where many old lineages still survive. From that spread point it was carried rapidly by sea into north-west France, Ireland, north-west Spain and the Middle Rhine, which today have a high incidence of L21, and into Northern England and Scotland. Of about 45 known early Bronze branches or subclades of L21, almost all are found in Britain or in the English-speaking Diaspora. We are able to identify most of the larger subclades of L21 as 'Atlantic'—spread throughout the Atlantic Beaker range with a distinct presence in Cornwall-Devon in the early Bronze. Continental R-L21 has origins in small random samples from the extensive English distribution. While many studies have tried to identify continental contributions to Isles populations, here we suggest that the reverse was much greater, at least in the early Bronze Age.

The global distribution of L21 subclades is almost exactly Pareto, showing an entirely random expansion from an initial point of time, however that point is much later than the early Bronze. Around 100 BC a second major R-L21 expansion from a severe bottleneck was initiated in Ireland and Scotland, when a dozen residual 'deep' sub-branches sprang to life and came to dominate L21. This is consistent with a collapse in the effective population of Ireland, followed by a rapid expansion. The limited evidence suggests that a severe weather event, famine and/or epidemic occurred around this time. The strongly patrilineal nature of insular Celtic society helped to keep male lines culturally intact, so that these emergent deep subclades can still be identified with Irish clans to some extent.

Around 90 per cent of R-L21 individuals in Scandinavia have paternal-line relatives in Ireland and Scotland dating to Viking times. The distribution is random and involves small numbers and distinct lines, suggesting that some of these were taken to Scandinavia as prisoners and slaves.

The Great Migration of millions of people from Ireland and Scotland to North America in relatively modern times was so substantial that no founder effects can be discerned and the New World has acted as a growth matrix extending and preserving the pre-existing R-L21 distribution.

This paper introduces several 'skyline' methods to trace the development over time of the subclade distribution of L21. These show that the distribution in England has not changed a great deal since the Bronze Age, in stark contrast to the situation in Ireland and Scotland. England and the Continent now make a much smaller contribution to R-L21 than in the past, probably stemming from Roman and Germanic expansion that pushed L21-bearing populations westward.

The phylogenealogy of R-L21: four and a half millennia of expansion and redistribution

1. The discovery of L21

The Y-chromosome is passed more or less unchanged from father to son except for a few small mutations that accrue from time to time due to copying error. These mutations, known as single nucleotide polymorphisms or SNPs, define a unique family tree of all men, an ordered tree of descent known as a haplotree. This tree is quite detailed; and like the rings of a tree, it contains information about the broad demographics of mankind, revealing a great deal about population settlement, expansion and development. This may be studied using the statistical methods of genetics, or by methods rather similar to genealogy, in which genetic, archaeological and historical evidence is assembled relating to the timing and place of each SNP mutation and the history and distribution of descendants carrying the SNP. We call this latter approach phylogenealogy.

Haplogroup R-L21, the group of all men that have the SNP mutation known as L21, is the most common patrilineage in the British Isles. It is a major branch of the general Y-haplogroup R1b that has dominated Western Europe since the early Bronze Age. Around 37 per cent of men in the British Isles as a whole are R-L21, and two-thirds of the Irish. The coastal Atlantic areas in France across from Britain and an area on the Middle Rhine also have significant incidences of L21, but otherwise the presence in continental Europe is low. Because the British, particularly the Irish, have been such major contributors to the populations of USA and Canada, R-L21 is also one of the commonest lineages in North America. [1] It has sometimes been identified as a carrier of Celtic culture because of its high frequency in areas that once spoke Celtic languages, and details of the lineage have been eagerly researched by those claiming Celtic heritage.

The L21 SNP marker was located on the Y-chromosome in 2005.[2] The major interest of genetic genealogy at the time was not so much in SNPs, as testing for these was then unavailable to the public, but in finding sets of rapidly mutating multivalued short tandem repeat (STR) marker values that might provide a DNA signature for the ancient families of Ireland, in the same way that many modern surnames have STR signatures or haplotypes. Various STR clusters were identified, mostly informally, and associated with particular clans or districts and with traditional clan leaders from the early Christian period (see Wright 2009, Wilson et al. 2001).

The O'Neills of Ireland were the best known and important family in Irish history, descended from a long dynastic line that for centuries were Kings of Ulster and High Kings of Ireland. In 2006 a group of researchers from Trinity College Dublin reported a signature haplotype which was allegedly associated with some descendant families of O'Neill, and which they claimed was,

> a biological record of past hegemony and supports the veracity of semimythological early genealogies. The fact that about one in five males sampled in northwestern Ireland is likely a patrilineal descendant of a single early mediaeval ancestor is a powerful illustration of the potential link between prolificacy and power. (Moore et al. 2006).

This research was subsequently used to promote and badge DNA testing: *Are you descended from the legendary kings of Ireland?*

---

[1] Over 63 million men are L21: 21.5 million in Europe and 42 million in the English-speaking Diaspora. (estimated from the Origins Database and Hammer et al. 2005).
[2] Along with several other key R1b-subclade markers in Figure 5 including U106 and U152.

These assertions turned out to be over-enthusiastic in a number of respects,[3] but the results did demonstrate that the descendants of a single man could eventually come to dominate a large population. Other examples from the distant past were soon found; however the growth of single lineages to such an extent has not happened in historical times, and the circumstances in which it might occur remain uncertain.

After 2008 when SNP testing became commercially available, it became apparent that the various Irish clusters were actually associated with SNPs that defined deep subclades of L21 on the phylogenetic Y-tree. The O'Neill 'Irish modal' or 'Irish Type 1' cluster in particular was ultimately defined by M222 (Kennedy 2014), one of the original SNP markers named by Underhill (2003).

In the last few years, the advent of NextGen DNA sequencing and amplification technology, in which large sections of the human genome can be sequenced fairly cheaply and reliably, has diverted the attention of geneticists away from the Y-chromosome to whole-genome comparisons of current populations and ancient DNA. However NextGen complete or partial sequencing of the Y-chromosome has also greatly extended the scope of Y-chromosome analysis. Very recently affordable panels of SNPs have become available that have permitted the very large 'untested rump' of R1b samples in the public domain to be re-tested for subclade membership – and they have revealed an extraordinarily complex hierarchical structure within R-L21.

This paper largely deals with the changing incidence of R-L21 and its subclades, in the context of the four expansive phases during which it came to occupy its current spread:

- the founding of L21 and its major branches and settlement areas at the beginning of the Bronze Age around 2500 BC, followed by a long 2500 year interregnum with relatively little activity, the settlement of Scotland some time later;
- a major redistribution and subsequent expansion of L21 in Ireland and Scotland at the dawn of the Common Era 100 BC–700 AD;
- limited translation of L21 to Scandinavia as the result of Viking incursions; and
- the Great Migration to the English-speaking New World where most R-L21 is to be found today.

We develop phylogenealogical methods to investigate lifetime changes in distribution of the R-L21 haplogroup and the possible causes of these changes. In doing so we reach several substantial new conclusions about the populating of the Isles.

The Appendices contain most of the technical material and tables. Appendix A discusses the data, and Appendix B examines the constraints of using traditional population genetics variance methods for L21, because of its pronounced internal dynamics. Spreadsheet C has the larger tables used for the study.

2. Incidence of L21 and subclades

The incidence by country of origin of L21 is shown in Figure 1 (see spreadsheet Table C1, which calculates the distribution of Y-haplogroups in most European countries).

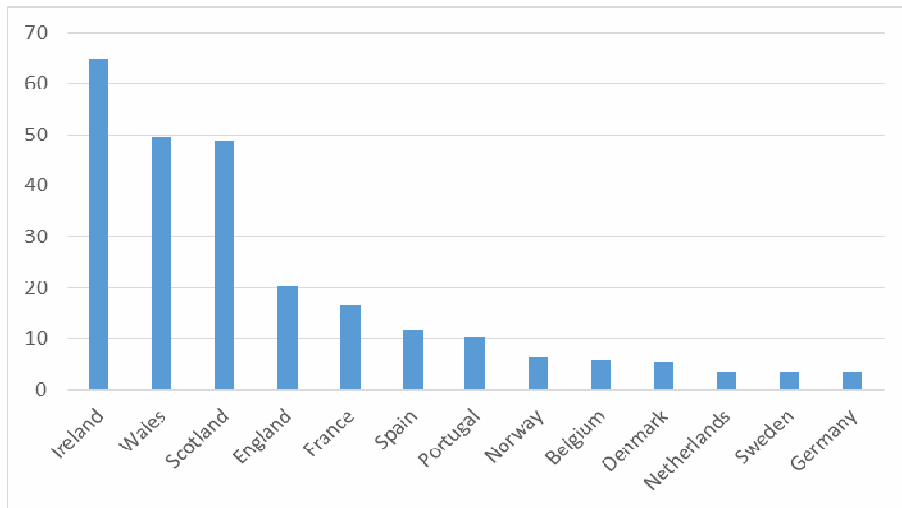[3] Critiqued in two chapters of Jaski (2013)

Figure 1. Incidence of R-L21 by origin
Source: European Origins database (Appendix 1, Table C1).[4]

A strong west-to-east decline is evident, and there is a heavy presence of 50 per cent or more in the traditional 'Celtic' locations, especially Ireland. France has about 16 per cent R-L21 while Iberia has 10 per cent. The incidence in the 'Germanic' countries is low, about 4 per cent. There is a residual presence throughout most of central and eastern Europe.

The distribution is heavily regionalised. Within England, as Figure 2 shows, London has the highest incidence of L21, presumably due to immigration from the 'Celtic' areas, followed by the North and Midlands, because of proximity to Scotland. The south coast has a relatively low incidence, with a pronounced west-to-east gradient, because of the mix of other subclades of R1b—especially U106 and DF27.
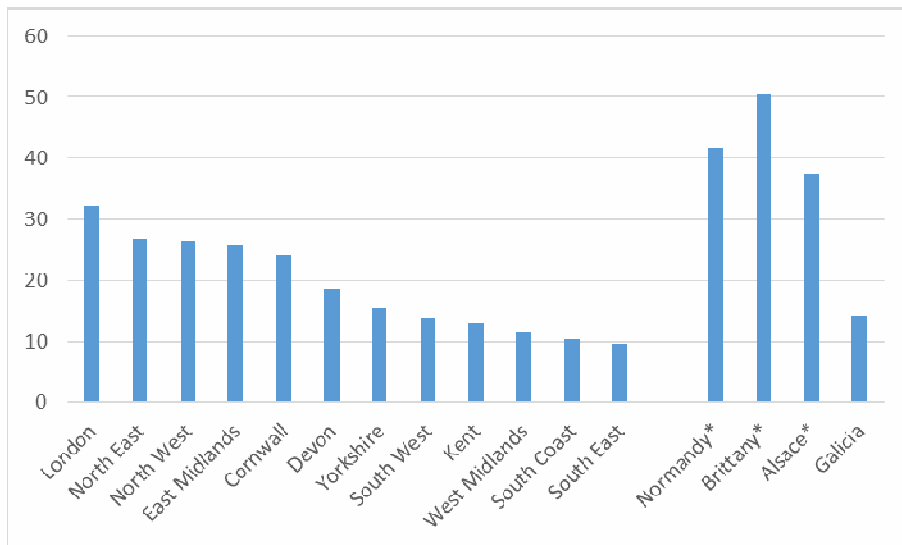

Figure 2. Incidence of L21, regions of England, France and Spain
Note *: Small samples.

---

[4] The database shows a lower incidence of R1b than some other sources – which translates to a lower incidence of L21. See Appendix A.

Galicia in the north-west of Spain has a similar incidence as the south-west of England. Although the samples are small, it appears that Normandy, Brittany and Alsace in France have a high incidence similar to the Celtic fringe of the Isles.[5] It is believed the Rhineland of Germany also has a high incidence similar to Alsace, though we have not been able to locate specific data for the area.[6]

L21 has about 45 subclades immediately below it or its major branch DF13.[7] Some of the subclades also immediately branch rapidly, so that L21 has 75 known branches that survive from 2200 BC or earlier.[8] So many of the early Bronze subclades of L21 have survived to the present that any sufficiently widespread data collection will contain many of these. For example, the 1000 Genomes Consortium (2010) collection from Cornwall and Kent contained 30 R-L21 men who fall into 15 different subclades including several rare lines.

The distribution of the sizes of the larger subclades of L21 is shown in Figure 3. It is extremely close to an idealised Pareto distribution (Fox and Lasker 1982) showing that what we see is a Yule process attributable to pure chance (Yule 1925, Rossi 2015) or genetic drift, which does not require any special assumptions about human behaviour. The distribution is self-similar as one continues downstream through the haplotree, since the number of descendants of any fairly large group of men, taken to enough generations, is always Pareto in the limit.
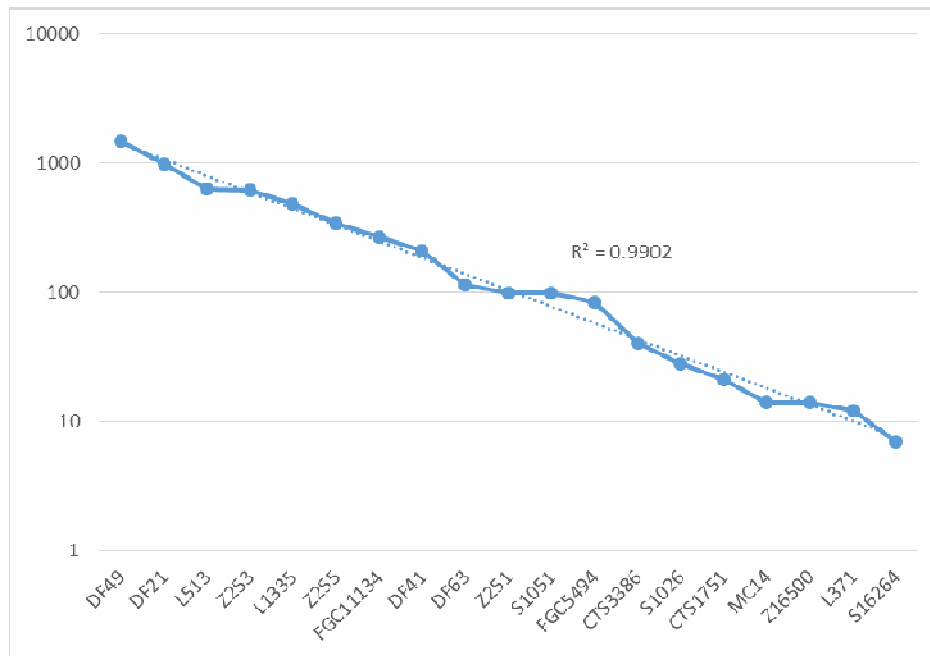

Figure 3. Log rank size of sizes of subclades of L21, with trend
Note: There are about 15 smaller known subclades.
Source: L21 full database, Appendix A, N=6619.

---

[5] Using the generic term 'the Isles' to refer to Britain, Ireland and associated islands.
[6] The FTDNA mapping facility shows considerably more L21 on the German side of the Rhine, along the stretch from Koblenz to Stuttgart.
[7] Also if below the early marker ZZ10_1. This includes 19 singletons with only one known representative.
[8] Branches are identified from their presence in the database, and confirmed in the websites www.ytree.net or www.yfull.com. Timing largely follows the methodology used in www.yfull.com, where each SNP mutation found in a Big Y test occurs on average every 140 years or 5 generations approximately. See Appendix A.

The tight conformity to the theoretical distribution confirms that our sample accurately represents the global distribution of L21 and is not subject to very recent bottlenecking or founder effects to any degree, at least as far as the sizes of subclades are concerned (see Appendix A).

The subclades of L21 have a very different incidence in different countries, which gives a strongly local character to L21, as Table C2 and Figure 4 show. The larger subclades are spread fairly evenly in England, suggesting it is a central place or point of distribution where subclades either developed first or subsequently mixed. In Ireland and Scotland, by comparison, the top five subclades account for about 80 per cent of the total, and the dominant subclades are different in each country. As we shall see, much of the difference in Ireland and Scotland is due to large founder effects in SNP lineages dated to the early Common Era, 2500 years after the formation of L21.
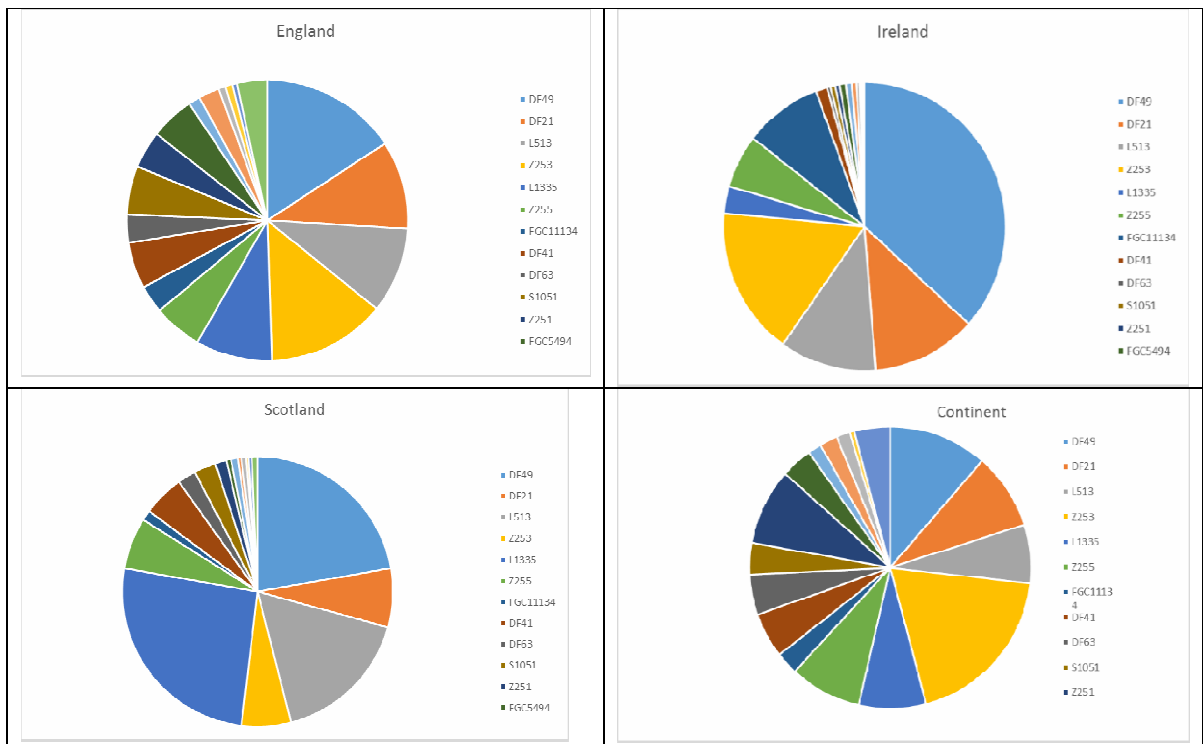

Figure 4. Distribution of largerL21 subclades: England, Ireland, Scotland and the Continent
Source: L21 database (Table C2a)

Although the distribution of subclades in individual European countries differs a good deal, the total distribution for mainland Europe looks rather like the English distribution – suggesting that European countries have been populated by small random samples from England. At least in the early years of L21, this appears to be true.

3. Early Bronze – the primary expansion of L21

3.1 The phases of expansion of L21.

There have been several periods when L21 expanded substantially, as shown by a very rapid increase in the number of sub-branches over a short period of time. In genetics, the *effective population* is estimated by the number of allele changes over a period of time, and rapid changes in the effective population are assumed to correspond to similar changes in the real population.

The genetic evidence for the expansion of L21 shows the following:

- In the first expansive period, from 2500 BC to about 2000 BC, L21 and its subclades were founded, split and expanded throughout the Atlantic Beaker range. Sometime later in the middle Bronze the expansion extended to Scotland.
- In the second period, the early Common Era from 100 BC to-600 AD, another large advance in Ireland and Scotland substantially reorganised the distribution of subclades. This was probably preceded by a very substantial fall in the population of Ireland, a near-extinction.
- In the Viking period, this new Irish-Scottish L21 was carried into Scandinavia and Northern Europe, probably by slaves taken in raids.
- Finally in the colonial period from 1600 to 1900, L21 moved freely throughout the English-speaking Diaspora, preserving the ancient distributions and preventing the extinction of some ancient lines.

The rest of this section examines the phylogenealogical evidence for these assertions, including evidence from the L21 haplotree and from archaeological and historical sources.

3.2 Beginnings - The Atlantic culture

The L21 mutation occurred during an extraordinarily rapid expansion of the effective population of the R1b haplogroup on the Atlantic seaboard. In only a few hundred years. 'Western R1b' formed over 300 Y-chromosome branches that survive to the present day and which define our current categories of Western R1b.[9] Batini et al. (2015, Figure 1) show that the branching of R1b at this time was spectacular, equal to that of all other European haplogroups taken together. No other effective male population expansion of this rate, magnitude and extent is known until the modern era.

The companion paper Flood (2016) proposes that the original Western R1b men were a closely related group of mariners and traders who came to the Atlantic seaboard from the Mediterranean around 2700 BC.[10] These invaders are often known as the 'Bell Beaker Folk' because of their distinctive drinking vessels. The Bell Beaker period marked a period of unprecedented cultural contact in Atlantic and Western Europe on a scale not seen previously nor seen again.

With boats as their major form of transport and trade as a major means of sustaining communities, the Bell Beakers established their initial colonies near to tradeable resources, on the coast and up major rivers. They appear to have leapfrogged to specific areas, probably to exploit valuable metals like gold, tin and copper- very much in the manner of their descendants in the New World four millennia later. The Beakers formed maritime colonies in quick succession in Iberia, southern England, Ireland, the Rhone Valley, Brittany and the Middle Rhine. These settlements grew together connected by the sea trade routes of the 'Atlantic culture' (Cunliffe 1994, 2001, 2010), with the south of England at the centre. As Bradley (2007: 26) puts it, 'The islands' distinct geography … allowed them to form links with regions of the European mainland that would not have been in regular contact with one another.'

---

[9] This compares with about 40 haplogroup lineages that survive in Europe from before the last Glacial Maximum, and an estimated 400 lines that developed across Eurasia from 25 000 BC to 6000 BC (Flood 2016), almost none of which are attributable to Western Europe.
[10] Flood (2016) critiques an alternative theory that R1b arrived at the Atlantic seaboard by land.
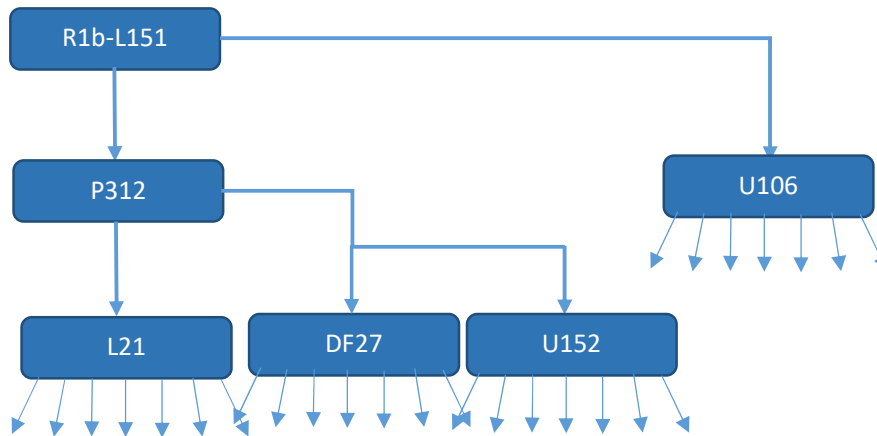
Figure 5. Atlantic R1b, major branches

In Figure 5 we see the different 'Western R1b' haplogroups R-DF27, R-L21, R-U106 and R-U152. Flood (2016) regards these as the genetic expressions of separate settlements - R-L21 in the south-west English mining and religious settlement, R-U106 (probably) in south-east England, R-U152 on the Rhone, in Lombardy and the Cisalpine area, while R-DF27 represents the original Iberian settlement.

The largest Beaker settlement, apart perhaps from the gold-tin Tagus valley settlement in Portugal, appears to have been in south-west Britain 'Wessex' in what looks very much like the world's first minerals rush, seeking the world's most valuable resources at the time, alluvial gold and tin. Standish et al. (2015) write, 'Southwest Britain would have been an extremely important region during the Bronze Age, as local populations would have had the ability to control the supply of two of the key materials in use at this time.'

NextGen sequencing of the rapidly branching R1b genome has allowed for reasonably accurate dating of L21 to about 2500 BC. This date has been supported by the presence of Bell Beaker sites all over Britain and Ireland dating from before 2400 BC. The Beaker constructions in Cornwall are the most extensive in Britain with an abundance of round barrows and cairns, henges, stone circles and stone cist graves.[11] The construction of Stonehenge II and III in Wiltshire, which required complex logistics and extensive manpower, was probably funded from the proceeds of the Cornwall-Devon mining bonanza. At Durrington Walls near Stonehenge, the largest village on the Atlantic seaboard was sited for a short time around 2500 BC, housing about 4000 people from all over Britain (Parker-Pearson et al. 2013).

Ireland is particular important for L21. A Bell Beaker arsenical bronze smelting industry at Ross Island in the south-west of Ireland dates to 2400 BC, when the local sulpharsenide ores were smelted to produce most of the arsenical bronze axes used in Britain. Traded artefacts from the site have been found in the south of Britain, while large numbers of artefacts using Cornish gold have been found in Ireland.[12] A long-suspected relationship between Bell Beaker peoples and R1b DNA has now been confirmed by the sequencing of the first ancient Bronze Age genome in the Isles (Cassidy et al. 2016). Remains at Rathlin Island off the north coast of Ireland have been dated to 2050 BC and are L21>DF21, the largest subclade of L21 prior to the Christian era. Rathlin was a production facility for

---

[11] Pevsner (1989: 27); http://www.historic-cornwall.org.uk/flyingpast/living.html Accessed April 2016.
[12] David Keys, *Cornwall was scene of prehistoric goldrush, says new research.* Daily Mail 5 June 2015; reporting on Standish et al. (2015).

porcellanite axe heads, a dense form of recrystallised basalt, and several Bronze Age gold artefacts have been found there (Jope et al. 1952).

After this early 'rush' of settlement Ireland seems to have been demically isolated from the main Bell Beaker culture. Once the metals rushes were over, the Irish Beaker period was 'characterized by the ancientness of Beaker intrusions, by isolation and by influences and surviving traditions of autochthons' (Osmon 2011).

Another very early L21 colony was founded on the Middle Rhine, and it is believed a high incidence of L21 occurs in the area even today.[13] The presence of unique early L21 subclade branches from the area suggests the settlement date is probably prior to 2300 BC. Cassidy et al. (2016) found a significant admixture closely related to Irish DNA in modern Germans (particularly visible as a Middle Rhine hotspot in their Figure 3). The presence of Bronze Age Wessex wheel-and-cross disks and Wessex-style pottery along the Middle Rhine,[14] coupled with the L21 genetic connection, make it likely that the colony was launched from south-west Britain.[15]

It appears that the Beaker expansion hit its carrying capacity quite quickly, because after about 2000 BC there are few new branches in the L21 haplotree until the Common Era. One exception is very significant branching in the L513/DF1 subclade about 6 SNPs or 800 years from the formation of L21. This might correspond to a mid-Bronze population expansion in Scotland; a late Scottish Bronze Age where arable land expanded at the expense of forests; perhaps because all suitable land had been cleared in England and by this time and settlers turned to more marginal land in Scotland. This may have occurred as late as the Bronze Age Climatic Optimum around 1600 BC, when climate change made settlement further north more practicable.

3.3 L21 Subclades in the early Bronze Age

Our principal technique for classifying the original subclades of L21 is to examine the number of early Bronze branches and their geographic extent. The subclades fall into three broad classes, which we call Atlantic, local or residual, depending on their size and spread.

Most of the major subclades are *Atlantic/Wessex.* These subclades branched widely and disseminated along the 'Atlantic culture' routes from south-western Britain to Iberia, Brittany, and the Rhine in the last half of the 3rd millennium BC. We define Atlantic subclades as satisfying three conditions:

- having many early branches, spread widely throughout Britain and Ireland; suggesting a spread during the initial colonisation by R1b. The number of early branches[16] typically determines the size of the subclade;

---

[13] The Beakers normally settled near tradeable resources and why they bypassed the lower Rhine is not clear. Given their interest in alcohol (Sherratt 1987) it is tempting to think they were interested in the wild grapevines that grew in the area.

[14] Flanagan (1998), Taylor (1980). Clarke (1970) attributes 'Wessex-style beakers' in south-west England to a 'rich and powerful group of settlers from the Middle Rhineland who came mainly from the area around Mainz and Koblenz'. The genetic evidence strongly suggests the reverse is the case.

[15] By an ironic quirk, some of the various Celtic and Germanic invaders of Britain over the millennia might be said to be 'returning home'.

[16] For the purpose of this exercise we count the number of 'early branches' that occur within 6 SNPs of the founder of the subclade, using www.ytree.net which has the most up-to-date, accurate and easily followed haplotree of L21. An 'early branch' should be standalone in origin, without other places obviously higher in the tree.

- two separate early Bronze lines (branching pre-2000 BC) in Cornwall or Devon, signifying the likelihood of a very early ancestor located there;
- early Bronze lines in two or more of France, Germany or Iberia. L21 spread to these places very early but never really 'took off' as far as we know; the continental lines of L21 we observe are very long and thin and their branches are early. However the relatively low numbers tested in mainland Europe hampers analysis.

Condition b) locates the subclade in south-western England. The presence of two widely separated[17] members of the same subclade in Cornwall or Devon provides a good chance their common ancestor lived in the area before 2000 BC. We know that the Cornwall/Devon area was a major dissemination point for R-L21, and likely had the first large settlements in Britain, Beaker or otherwise. That was where tin and gold were found in most abundance, and where the largest number of Beaker sites are found. The significant presence of about 15 per cent of the brother Iberian haplogroup R-DF27 in Cornwall, the only place in Northern Europe with a significant presence, confirms the scope of early population exchange with Spain. It is known that Cornwall and Devon are genetically differentiated from the rest of Britain (Leslie et al. 2015) and very early differences in genome have persisted there.

Of these Atlantic subclades:

- *DF21* was comfortably the largest subclade overall until the Common Era (see Table 1 and Section 5) and was particularly dominant in Ireland. Although it is regarded as 'original Irish' initially it was a classic Atlantic subclade, with 28 early branches including one each in Iberia (*Torres*), the Rhineland (*Fuston*) and Normandy (*Montgomery*), and it remains the largest subclade in England and second in Ireland (Table C2a). There are five members of DF21 in Cornwall and Devon, within two different early branches.
- *Z253* branched 27 times during the early Bronze era and is the most widespread subclade and the most common on the Continent and in England before the Common Era. It remains the second-largest subclade in England and third in Ireland (see Table C2a). Our sample has six widely separated men in Cornwall/Devon,[18] and early branches are found across Britain and around the Atlantic horizon in France, Germany and Spain. As well Z253 found its way to Ireland and Scotland very early and half a dozen branches are Irish (some of these might be continental incursions).
- *Z251* is also widespread with 15 early branches, mostly English but with representatives having ancestry in Scotland, Portugal, Germany and even Poland and Mexico. Like other English subclades it has fallen in importance. The branch S9294, dating to 2000-2300 BC, is probably originally Cornish as it has two widely separated representatives *Millett* and *Watty*.
- *DF41* apparently did not branch for 9 SNPs (the late Bronze), after which it formed 15 branches across the Atlantic horizon from Germany to Iberia. Two of these branches are found in Cornwall/Devon. A subclade A40 participated in the post-Roman growth spurt in Scotland, while the Royal Stewart line, which is from Brittany, expanded rapidly in mediaeval times.
- *DF49* began as a fairly sizable Atlantic subclade, branching 23 times in the early Bronze. It was widespread across the Isles with traces in France (Normandy and Poitou) and in Iberia. It is found in widely separated branches in Cornwall as *Coad-Coode* and *Biddick* and in Devon

---

[17] Defined as being in two separate Bronze branches, or having a genetic distance measured by 67 SNPs of 18 or more.

[18] Unfortunately these have not done sufficient testing to provide good evidence that Z253 was present in Bronze SW Britain. However the high STR variance of Z253 there is indicative.

as *Hicks* and *Woolcott*. Three thousand years later during the Irish Golden Age this subclade found real prominence as M222 and became the largest fraction of L21.

- *FGC11134*. This subclade had eight early branches, in France and Portugal as well as the Isles. Most of its development occurred much later in Ireland, probably in concert with DF49.
- *DF63* is regarded as the 'earliest subclade', because it branched six ways directly from L21 without the intervening DF13 mutation that is upstream of most other subclades. It has a fairly large early continental component making it truly Atlantic – Spain, Portugal, France, Netherlands and Germany, and is found in Cornwall, though not in particularly early branches (*Hicks* and *Trengove*). The 'Lennox Cluster' Z16506 participated in the post-Roman expansion in Scotland, probably as indigenous p-Celtic.
- *FGC5494.* This middle-sized subclade branched very early at least 14 ways and is found across Britain and Ireland, with small lines in Germany, France and Iberia. In the Isles it has continued down very long thin lines to a number of discrete surnames in the present: *Kenyon, Collings, McLaren, Maynard, Lunney, Maxwell, Phillips*. There are three early lines in Cornwall/Devon.
- *S1051* is a poorly studied middle-sized subclade. It has early branches in Iberia and Denmark, two widely spaced branches in Cornwall (*Priske, Medland*), and is spread across the Isles. It was once the fifth largest subclade but now is eleventh. It was subject to an early RecLOH[19] which gives it a fairly distinctive STR signature.

Table 1. Estimated pre- Roman incidence of major L21 subclades by Isles countries.

Subclades	England	Ireland	Scotland	N
DF21*	12.6%	27.5%	24.5%	147
Z253*	15.2%	16.1%	16.4%	104
L513*	7.9%	15.4%	11.3%	80
DF49*	12.0%	13.4%	8.2%	77
S1051	8.9%	2.6%	11.9%	44
Z251*	6.3%	3.9%	5.7%	33
FGC5494	8.9%	3.9%	1.9%	32
DF41	6.8%	3.0%	3.1%	27
Other	21.5%	14.1%	17.0%	111

Note: *) These subclades are enumerated net of the contribution of the deep subclades in Table 2, as a partial proxy for the early distribution. Alternative methods are explored in Section 5. Source: From Table C2c, using men in the L21 database who have 67 markers tested.

These nine Atlantic subclades formed part of the same general population, and they lived and migrated within the Atlantic geographical extent. Along with L513, they are estimated to comprise about 80 per cent of the pre-Roman R-L21 population.

A few small subclades are also probably Atlantic, such as S1026 and Z15600, found across the Isles and in France and Germany, CTS3386, which is mostly in Ireland but also Italy and Finland, and FGC35995/Y14240, which has only been found in France, Sweden and Mexico.

---

[19] A RecLOH occurs when one arm of a palindrome 'writes over' or replaces the other, making both values of a two-valued marker the same. This may be a big jump, and may affect more than one marker if they are on the same palindrome. A similar RecLOH on YCA (jump from 19-23 to 19-19) occurs in half a dozen other subclades of L21, but not so early.

The second class of subclades has been largely confined to one location and is not seen around the Atlantic horizon. It has only one significant member and several small members:

- *L513* is out of synchronisation with other subclades, expanding at rather different times and places. It seems that it became embedded in northern populations of the Isles and its expansion relates to periods of warmer weather. It first branched after eight equivalent mutations, somewhere about 1600 BC in the Bronze Climatic Optimum (Minoan Warm Period). It was then very vigorous, producing 27 early branches, including several found only in Scotland, giving the general impression of expansion into new territory. These early branches are equally spread between Scotland and Ireland, with some presence in England. Somewhere around 100 BC, in the Roman Warm Period, the branch L193 that is today the largest began to expand. It is tempting to regard L513 as the eponymous early Bronze Scottish subclade, just as DF 21 is Irish, however it may also have been strongly represented among the pre-Goidelic peoples in Northern Ireland.
- *MC14* is largely Scottish and branched at the same times as L513.
- *CTS1751* has an uncertain distribution, and several old lines in Devon suggest an origin there; but it is also found in Yorkshire and Lancashire. Its deep subclade BY595 is mostly Irish.

Table 2. Residual subclades of L21 and locations of origin

Subclade	Location of origin
15049032 A-G	England, Scotland
A5846	England, France, USA,
A7900	Wales, France, USA
BY2868/BY2899/A4556	England, Scotland, Wales, USA
BY4045	USA
BY575	England, Ireland, Wales, Finland
FGC13742	England, Wales, USA
FGC13780	Cornwall, USA
FGC21979/A9607	Ireland, Scotland, USA
FGC5924	Cornwall
L371	Scotland, Wales, Ireland
S16264	Devon, France, USA
Y14240	Wales, France, Sweden, Mexico
Z16500	England, Ireland, Scotland, France, Germany
Z17300	England, Scotland, Wales
Z39589*	Romania (originally German from the Rhine)
DF13*	England (4) Scotland (2) Ireland France Germany, Italy Norway USA (5)
ZZ10*	England Scotland Canada

Sources: L21 database, www.ytree.net

The third class consists of about 16 small 'residual' subclades, for which we do not yet have sufficient samples to deduce location or timing (see Table 3 for a listing). Some of these may be specific to certain locations, such as A7900 and L371 in Wales, while A5846 (*Blanchett* and *Berry*) might be French, FGC13780 is known only in Cornwall and the USA and BY4045 in the USA. Only one is not known in Britain.

There are also about 19 singletons currently designated DF13* or ZZ10*, with lines of descent within the Atlantic spread that have apparently survived only a few men from extinction for 4500 years.[20]

The distribution of L21 subclades across the British Isles for the first 2500 years seems to have been reasonably uniform compared with what came later, though Table 1 shows DF21 and L513 had a considerably higher incidence in Ireland and Scotland than in England, which had a much higher contribution from smaller subclades of L21 (see Section 5 for more detailed analysis).

The final group of subclades were late bloomers that came apparently 'from nowhere', expanding very rapidly in Ireland and Scotland from about 100 BC. These had existed as small background lines at the tail of the Pareto distribution for thousands of years, until a major redistribution of R-L21 gave them the opportunity to increase their numbers very rapidly.

4. Dark Age to Diaspora
We resume our historical account in this section, with a description of what happened to L21 in the period 100 BC to 1800AD, beginning with a very major redistribution.

4.1 The Dark Age - Golden Age redistribution and expansion in Ireland and Scotland
Something quite remarkable happened in Ireland and Scotland around the beginning of the Common Era. A very substantial makeover of the structure of R-L21 occurred, so large as to resemble a population recovery from some kind of disaster in which the Irish population was nearly wiped out. This was followed by very rapid growth. Two main phases appear to be involved in this expansion: one in the 'Dark Age' from 100 BC, then a consolidation and faster growth from 400 AD.

4.1.1 Dark Age collapse and emergence of deep subclades
First, somewhere around 100 BC, two residual subclades appeared from obscurity and began branching. They were L1335 'Scottish modal' and Z255 'Irish Sea', to use the STR cluster names under which they were first discovered. These were accompanied by nine equally obscure branches of major Atlantic subclades that had also been residual[21] since the early Bronze Age—DF49>M222 'Irish type I', DF21>Z3000 'Clan Colla', FGC11134>CTS4466 'Irish Type II', Z253>L225 'Irish Type III', Z253>CTS9251 'Irish Type IV' and L513>L193 'Little Scottish' (see Table 3a).[22]

Table 3a shows the principal 'deep subclades' involved in the Irish and Scottish Dark Age expansion, in descending order of variance. Excel Table C2d shows their detailed count by country. The subclades are all about 15 SNPs from the present on average, or 2100 years old, a period associated with the late Iron Age and Celtic culture.[23] The STR variances of the subclades shown in Table 3 are only about a half of those of the original Bronze Age subclades (see Table B1), which is to be expected as these deep subclades are less than half the age. However even though all these subclades are about the same age, the variances are different depending on the internal structure of the lineage (see Appendix B).

Table 3a. Deep subclades and clusters of the late Iron Age/early Common Era.

[20] As the sample expands it is expected that more will appear. In fact there are already a number of untested isolates within the database that are probably residual singletons.
[21] "Residual' here means having a long thin unbranching line with 15 or more equivalent SNPs (M222 has 45).
[22] The timing on these breakouts is important but not critical, Although yfull.com dates these subclades to 100-300 AD, we calculate using the same method on the larger sample in www.ytree.net that the coalescent for each of these branches is an average 15 SNPs or 2100 years
[23] The mean age of the subclades is calculated as in Appendix A. www.yfull.com dates them all to about 200-400 AD using the same method, but their L21 NextGen sample is quite small with relatively few branches compared with the.

Branch	L21 Subclade	Variance	N*	Cluster name
Irish				
CTS4466	FGC11134	0.159	186	Irish Type II
Z255	Z255	0.159	265	Irish Sea
M222	DF49	0.147	966	Irish Type I, Ui Neill
P314	DF21	0.140	53	P314 Project
L1336	DF21	0.135	42	Clare
CTS9881	Z253	0.135	35	Irish Type IV
Z3000	DF21	0.124	320	Clan Colla
Z16282	DF21	0.101	58	Carroll
Scottish				
L193/S176	L513	0.204	160	McLean- Little
A71/S190	DF21	0.192	109	Little Scottish
L1065	L1335	0.142	429	Scottish Modal

Note: *) Taken from Table C2d The lines/clusters selected are those of the right age greater than 0.5 per cent of the sample—all but two of which are very well known.

What makes these subclades 'deep' is that they straddle a long thin line of many equivalent SNPs, so that after the branch formation in the early Bronze there were no further known branches for typically 1500–2000 years, when they suddenly sprang to life. Because of this very long isolation during which the STRs mutated, the new founder of the deep branch might quite possibly have a very different haplotype than the founder of L21, and so therefore do all his descendants. This means that members of deep subclades can often be distinguished by STRs alone.
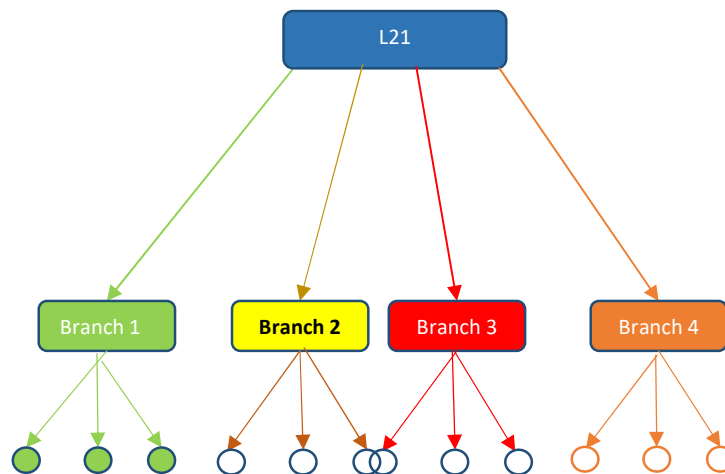

Figure 6. Deep branches, clusters and overlap

Figure 6 shows how these STR clusters form. The founders of deep branches 1 and 4 have haplotypes far from the modal, so their descendants do not have overlapping STRs. Deep Branches 2 and 3 however are near the modal and have some overlap in the STRs of their descendants, so they do not form a distinct cluster. Clearly, the branches have to be 'deep' (at least half way down the timescale) for this to work, otherwise their lines of descent will overlap.

Most of these Irish late–blooming subclades have been associated with particular clans and legendary leaders from the 'Dark Age' and in fact they do show some degree of correspondence with

particular surnames traditionally associated with the clans. However the correspondences are less than perfect. Some of the larger 'deep subclades' are:

- *M222* is the extreme case however considered. It has 45 equivalent SNPs in its lead-in and its haplotype is so distinctive it can be picked up with no error by only a few SNP markers (which is how it was discovered). It is common among surnames such as *Gallagher, Boyle, Doherty* and *O'Donnell* purportedly associated with the legendary Ui Neill, High Kings of Scotland. It is almost three times as large as the next deep subclade, and it is the flagship subclade of L21 (especially in Ulster, where Busby et al. (2012) and Myres et al. (2011) found local incidences of over 40 per cent). It also spread aggressively into Scotland, in moderate quantities to England, and in trace amounts to Northern Europe. As we shall see it is possible that it is foreign to Ireland.[24]

- *Z3000,* known as *Clan Colla,* is one of a number of deep lines of DF21, the largest subclade in Ireland before the Common Era expansion. The lineage is supposed to be descended from the Three Collas, warlike chieftains who conquered Ulster in the early part of the 4th century, one of whom became the first King of Airgialla in southern Ulster. The *Maguires, MacMahons* and other surnames are supposed to be descended from Airgialla, though this is only party demonstrated by DNA (O'Hart 1892, Biggins 2016.). This is the only late subclade to be found in Wales in significant numbers; perhaps associated with the Kingdom of Dyfed and the colony of the Deisi in Pembroke, or with the Irish colony on the Llyn Peninsula in the north-west.

- *Z255* 'Irish Sea' is one of the residual subclades from the tail of the original L21 distribution that suddenly sprang to life in the Christian era. Surnames such as *Byrne, Gleeson, Fitzpatrick* and *Beatty* are involved. It has a good representation in Scotland and England, and also in Scandinavia.

- *CTS4466* 'Irish Type II' is numerous in southern Ireland but rarely occurs elsewhere. It includes surnames such as *Collins, Donohue* and *Sullivan*. Like M222, it sprang from a fairly small Atlantic subclade, in this case FGC11134, down a long line of equivalent SNPs.

- *CTS9881* 'Irish Continental' shows the inaccuracies that may arise when only a small number of STRs are used to try to identify clusters. It was initially thought[25] that the subclade included a number of Norman-English surnames and that it must have come from the Continent through the Pale. Most of these matches disappeared once SNPs were tested. There is one small English branch BY412 but it is recent, probably from Irish migration.

A similar but less extensive expansion began around the same time in Scotland, presumably under the influence of some of the same drivers.

- *L193,* a fairly deep branch of L513, expanded rapidly in Scotland in the Celtic period, along with the 'Little Scottish' branch A71 of DF21. It seems that these formed and expanded in the original p-Celtic Brythonic-speaking population of Scotland.

- *L1335* Almost as spectacular as the M222 tsunami was the sudden rise of a new branch of the background subclade L1335, which had only branched once since its inception. In the span of six SNPs, from about 100 BC to 700 AD it branched an extraordinary 38 times, the record for an L21 subclade. This founder extravaganza brought L1335>L1065 from a single man to the largest L21 subclade in Scotland, so that it is now known as 'Scots modal'. There is no clue to its origin except for a small branch from about 1800 BC living on the remote Llyn

---

[24] http://www.ancestraljourneys.org/irishsurnames.shtml
[25] https://sites.google.com/site/irishtype4/irish-type-4-sub-clade

Peninsula in north-west Wales. This may be misleading as the Llyn was the site of an Irish colony in the early Christian period.

4.1.2 'Golden Age' expansion and overflow

About 500 years later, after the Romans departed from Britain, this expansion continued very vigorously in the early Christian period in Ireland, flowing through to Scotland and later to England and Wales. The expansion was still centred on the deep subclades but was more broadly based. This surge from about 300–700 AD laid down the subclade distribution we see today. About 80 per cent of present-day L21 from Ireland and Scotland derives from deep branches at this time, and 40 per cent of L21 elsewhere – England, Wales and the Continent (see Table C2c and C2d). Overall, about 70 per cent of L21 is from the period, so that L21 is much more a Common Era phenomenon than a Bronze Age one.

Following the departure of the Romans from Britain, a further expansive surge appears to have occurred in what became the early Christian 'Golden Age' of Ireland.[26] This later expansion appears to be a largely random outcome of rapid growth within the redistributed Irish population. Some of the larger new 'deep' branches involved are shown in Table 3b. As well, M222>DF104 was formed, which contains the bulk of the M222 expansion. These subclades all have about the same STR variance as the Dark Age subclades, but their average SNP count is several less. They all occur in the big subclades DF21, Z253 or L513.

Table 3b. Subclades of the Irish Golden Age

Branch	L21 Subclade	Variance	N*	Mean SNP count[a]	Cluster name
CTS3087	L513	0.156	26	12	
S7898	Z253	0.150	22	12	Corofin
Z16372	L513	0.146	29	13	Shaw Nicholson
Z23532	L513	0.141	23	11	
L226	Z253	0.132	187	13	Irish Type III, Boru
L1402/A385	DF21	0.131	34	13	Seven septs of Laois
L1336	DF21	0.101	42	13	

Note a) Big Y counts are 140 years per SNP, so 13 would have TMRCA about 180AD.

Some of the better known of the 'Golden era' subclades in Table 3b are:

- *L226* 'Irish Type III' is the largest branch of the eclectic subclade Z253. It begins with a 25-SNP lead-in which probably is native to Ireland but might be foreign. It is found in Munster, especially in Tipperary, Clare and Limerick. (Wright 2009). It is associated with the Dal gCais or Dalcassians, a tribe who designated themselves as descendants of a semi-legendary Dark Age king of Munster, Cormac Cas. The most famous member of this tribe was King Brian Boru who ended the line of Ui Neill High Kings in the early 11th century. Common surnames in the clan include *O'Brien, MacNamara, Kennedy, Grady*, *McMahon Hogan and McGrath* (O'Hart 1892). The Kingdom of Dyfed in Pembroke is supposed to be Dalcassian but there is no sign of this lineage in Wales.
- *'Cruithen' L513.* Three separate lineages of L513, which we normally associate with early Scottish DNA, are found in Northern Ireland dating from the Golden Age (Table 3b). Along

---

[26] Again, www.yfull.com shows the TMRCA for some of these subclades as 650-800 AD, but from the SNP count, 200-450 AD is more likely.

with DF21, they are probably part of the original p-Celtic population in Ireland and Scotland. Although these subclades are predominantly Irish (see Table C2d) they have some Scottish presence, suggesting a continuous link. A people known as the Cruithen lived in Ulster at the dawn of recorded history, and the Irish used this same term for the Picts of Scotland. The Cruithen were ultimately overcome by the Northern Ui Neill in the 7th Century and many were driven to Scotland. Although some scholars have rejected the tie,[27] the presence of expanding Golden Age L513 lineages in Ireland lends some support to the Pictish connection.

In Scotland the main event of the period was the overflow of Irish excess population into the Western part of Scotland as the Goidelic-speaking 'Scots' who eventually overcame the confederation of original Brythonic-speaking tribes known as Celts (Broun 1999). Most of this 'invasion' is easily visible as M222. Given its peculiar origins and extraordinarily rapid growth from a single man, it is possible that L1335 is also Dark Age Irish and accompanied the Scots.

4.1.3 The stripping of the tail
Further evidence for a near-extinction event in Ireland is the stripping away of the tail of the Pareto subclade distribution there. In a population bottleneck, a small sample of men survive including only a few subclades from the long tail. These have no competition so they may rapidly become major subclades, further stripping away the tail. This is what appears to have happened in Ireland.

As we saw in Figure 4, middle-sized subclades are poorly represented in Ireland. This continues into the tail of residual subclades: Table 2 shows 15 of the residual subclades and singletons in England but only four in Ireland; Wales and even Cornwall-Devon with much smaller samples have more small subclades. A standard property of the Pareto distribution is that doubling the sample size will increase the number of subclades present by a fixed amount, The Irish sample is more than four times as large as the English but the number of residual subclades is much less, suggesting the tree has been severely pruned.

A possible alternative is that Ireland was originally settled by only a small number of L21 men so that the bottleneck occurred at the beginning. However the Continent and Scotland have a reasonable share of small subclades showing from the Atlantic expansion and so should Ireland; as well it should have developed its own mini-branches since we know it was settled very early in the Beaker expansion. The anomaly is a prime sign of a bottleneck event.

4.2 Collapse and recovery?
The post-Roman expansion within Ireland was preceded by a substantial collapse, recognised as a period of fortified warfare from 100 BC to 300 AD, which Charles-Edwards (2000) has called the 'Irish Dark Age'. The genetic record suggests that this was more prolonged and far darker than anyone has previously considered. The substantial reorganisation of male DNA lines in Ireland over a few centuries, coupled with the pruning of the residual subclades, is consistent with a catastrophic decline in the effective male breeding population of Ireland, perhaps to only a few hundred breeding men—followed by a very rapid recovery and a growing population that increased well beyond the original level. These bottlenecks appear to be quite common in the ancient genome and are responsible for intermittent severe pruning of the phylogenetic tree – explaining why so very few lineages have come down to us from ancient populations that numbered in the millions. This is the first bottleneck event to be so clearly identified in time and place.

That there should have been population growth in Ireland and Scotland at this time is not unexpected. From 250 BC to 400 AD the climate in Northern Europe entered what has been called

---

[27] Ó Cróinín (1995), Jackson (1956).

the Roman Warm Period (Bianchi and McCave 1999). For ancient peoples, population carrying capacity is largely dependent on food availability, and the warmer climate was beneficial for crop production in colder areas such as Scotland and Northern Ireland. It is likely that the warmer climate also led continental Celts further north, and eventually the Romans.

We know that Ireland on the edge of the known world was very backward throughout the Bronze Age, and when Iron Age technology finally arrived, this must have provided a considerable boost to food production. Ireland sat outside of the Roman world, but the slow diffusion of Roman advances in crops and farming, probably through the agency of Christianity, must have kept the momentum up for a considerable time.

It is not the fact of expansion, but the extraordinarily sudden appearance of 11 new rapidly expanding subclades that lead us to suspect a catastrophic event. Y-lineages normally maintain their relative proportions once they are of sufficient size (and in England, the relativities were largely maintained within R-L21, as we show in Section 5). In a large established population, any new haplogroup will not randomly reach a sufficient size in competition with other lineages to make much impact. The only ways for a lineage to break out from obscurity are: in the aftermath of a severe bottlenecking event, through founder settling of new territory, through differential population growth rates, or if the expansion is not random. These are the possibilities we now consider.

It should be stressed that as far as we know we are not talking about an external invasion. The L21 subclades that appeared in the Dark Age had ancestor lines in the Isles for thousands of years but in tiny quantities. The simultaneous expansion of a small number of randomly chosen subclades in Ireland (and to a lesser extent in Scotland), apparently without interference from existing large lines, appears very similar to what occurs when a small population settles new territory. The relative absence of tiny subclades from the Pareto tail in Ireland also suggests a major dieback. The only real way for sudden marked founder effects to take place in an established population is for genetic diversity to be very substantially thinned before a major population expansion; then any man who by chance has many sons may make a large impact in the genetic pool as there is little competition.

4.2.1 Plague and famine
The DNA redistribution suggests a catastrophe, probably a famine and/or plague that more than decimated the Irish population. Exactly what plague or other disaster might have done the thinning is unknown. Among calamities, plagues generally take the highest toll. We speculate that as Ireland opened up to the outside world, diseases were introduced that had been present in Britain or the Continent for some time. The locals had no resistance - which incidentally would give resistant newcomers a considerable advantage in breeding for a few generations.

If Ireland had been isolated for millennia there were plenty of diseases to which the Irish would have not acquired immunity, such as those that later devastated the New World—smallpox, influenza, typhoid, yellow fever and pertussis—which taken together have typically wiped out 95 per cent of newly exposed populations (Diamond 2005). The arrival of Iron Age people, Roman traders or slaves taken in raids would have been sufficient to break the isolation and spread contagion.

Events of this kind were not unfamiliar to the ancients. The Irish pseudohistory *Lebor Gabala* tells of an early post-Deluge settlement in Ireland, about 9 000 people led by one Partholon, who all died of plague in a single week. A similar fate befell the people of Nemed some decades later, with all but 30 subsequently being killed in a battle with the Fomorians, aboriginal inhabitants of Ireland. While

these tales are not to be regarded as historical events, the possibility of plague devastating the whole island community was clearly familiar to the 11th-century authors of *Lebor*.

Famine, plague and war go hand in hand. Ireland has suffered regular severe famines, as has Northern Europe more generally. Myllyntaus (2009: 80) estimates that crop failures in Northern Europe occurred with much greater frequency than usual in certain periods. For example, torrential floods and rains caused a famine from 1315–22 in the Isles and Northern Europe that was the greatest on record, causing widespread starvation, violent social conflicts, ruthless crimes, epidemic diseases and very high mortality (Jordan 1997). Whatever happened in Ireland during the Dark Age must have been considerably worse than this, a 2500 year event probably reducing the population by more than two orders of magnitude.

Ireland is exposed to weather events and other extreme conditions because of its proximity to the Atlantic. The warmer climate between 300 BC and about 100 AD produced frequent extreme weather events on the Atlantic seaboard. Strabo wrote that around the years 120 to 114 BC (exactly in our genetic timeframe) storm surge from the North Sea covered large areas along the coasts of Denmark and northern Germany with water, permanently altering the coastline and forcing the Cimbrians, Teutones and Ambrones into the lands of the Romans. Similar events probably happened in Ireland during the same period, reducing the population, which could not so easily migrate.

Severe bottlenecking events associated with climate change or epidemic appear to have taken place with some regularity in prehistory, and periodic cullings such as this one are probably responsible for the very small number of ancient lines that have come to down to us on the Y-haplotree. The sweeping clean of populations by plague or disaster in ancient times was not however entirely without benefit. Subsequently, the land could be resettled more efficiently and new technology introduced, eventually producing a larger, healthier and more prosperous population.[28] Ireland had been particularly backward as a remote island at the edge of the known world, and this resettlement appears to have been very advantageous, both culturally and economically.

It is not necessarily a fatal objection that this proposed catastrophe is not recorded by either Irish or Roman sources, except in cryptic terms. Irish history in the first millennium is at best unreliable, a 'mixture of truth, lies, myth and legend'. The Irish were fond of compiling long genealogies that claimed descent for most of the common surnames from legendary heroes and kings. While the Irish genealogies do extend back into the Dark Age, they do not go as far back as 100 BC, apart from the tantalising clues of the *Lebor.* The Roman chronicles too were silent, as Ireland and Britain were outside of their sphere of influence at this time.

4.2.2 Invasion and war?

One thing we do have both archaeological and semi-historical evidence for is war. War eradicates the male population much more thoroughly than the female, Warfare then as now was bloody and brutal, but in these times it involved slaughter and selling off the survivors and their families.

Some of the newly expanding lines may have been foreign. O'Rahilly (1946) believed that the first Goidelic speakers in Ireland arrived from Aquitaine in south-western France about 100 AD in several groups including the Connachta under Tuathal Techtmar, who carved out a territory in Meath, the kingdom of the southern Ui Neill, fighting 'a hundred battles'. The Irish Chronicles describe Tuathal as a legendary figure descended from a long Irish lineage. Aquitaine is an unlikely site for L21>DF49;

---

[28]. Bell et al.(2007) suggest that while the Black Death of the 14th century decimated the English population, it cleared the land for the wool industry which later supported the Industrial Revolution and eventually England's colonial empire.

an alternative O'Neill lineage is R-DF27>>Z37492, much more likely to have come from southern France. The 17th century historian Geoffrey Keating compiled a tale that Tuathal was a leader of exiles in Scotland assisted to return by the Roman governor Agricola in the hope that raiding would stop under his rule. Part of this legend, again in the *Lebor*, refers to a severe famine that struck Ireland around 56 AD in punishment for the unseating of Tuathal's father, the rightful High King. However, critics have discounted this legend as propaganda designed to legitimise a foreign invading force. The advance of the Ui Neill seems to have eliminated much of the indigenous opposition or driven them off to Scotland. Tuathal's 'hundred battles' might have been a metaphor for Dark Age brutality and the elimination by war of large groups of people during a Goidelic-speaking Iron Age invasion.

A people called the *Fir Domnann* were said to have landed in Ireland and settled in several locations, and it has been proposed that they were Dumnonians from Devon/Cornwall. Branches of DF49 and divergent early branches of M222 are present today in Cornwall and Devon. Yorke (1995: 18–19) has speculated that emigration from Dumnonia to Armorica in the 5th and 6th centuries that led to a Breton-speaking sister kingdom in Brittany was an opportunistic expansion rather than a response to Saxon harassment, so an earlier intrusion bringing the M222 forerunner to Ireland is quite possible. M222 settlers in Scotland were prominent in the *Damnonii* tribe of Argyll, and the naming might once again be more than coincidence, reflecting an actual tribal source.

Earlier social historians were aware that during the Dark Age the population of Ireland had been severely reduced. Charles-Edwards suggested that the sale of prisoners to Roman Britain by local chiefs reduced the population. However this is disputed and it seems more likely that extensive Irish slaving raids to Britain actually augmented the population as the *pax Romana* declined (St Patrick was taken by slavers from England to Ireland as a young boy).

The advent of Christianity in the early 5th Century accompanied a flowering of local culture, and at this time the growth in effective population accelerated. Religion pacified the warring people and gave hope and focus after centuries of fear and withdrawal.[29] The monasteries were large organised farms, able to combine small holdings of land and engage in intensive farming and the cultivation of grains. Due to their regular interchange with Rome the monks may have brought newer crops and farming methods, and they were probably more prepared and had more time to experiment. Food production must have risen, permitting a rapid population advance above pre-crisis levels. It is certainly known that most of the forests of Ireland were cleared during the post-Roman period, a definite indication of a population surge. In this period, vows of celibacy would also have removed men and women from the breeding population, maintaining some continuing pressure on genetic diversity, though not sufficient to produce the wholesale distributional changes of the Dark Age.

Whatever the circumstances, and whether or not foreign expeditionary forces took advantage of a population collapse, the Irish expansion was both visible and forceful, and soon spilled over into neighbouring countries, most of whom were in disarray after the collapse of the Roman Empire. To a fair extent, L21 as we know it is not an early Bronze but a Common Era phenomenon, since that is when the distribution took its current form.

4.2.3 The influence of the Romans and Irish-Scottish interchange
The picture in Scotland is not as clear as in Ireland, although some of the circumstances are similar. Certainly the presence of factors encouraging rapid population growth were the same – the Roman

---

[29] It is one of the basic principles of demographics, almost a biological imperative, that when people feel safe and secure they will have more chilren.

Warming and the late arrival of Iron Age technology. However it is harder to explain why the growth was so uneven and initially restricted to the deep subclades L193, A71 and especially the upstart subclade L1335 (see Tables 3 and C3d). The same disaster that almost wiped out the Irish population might have also affected the West and the Lowlands of Scotland. However we think this is unlikely as the subclade tail has not been thoroughly stripped in Scotland—and it seems that a founder effect involving the settling of vacant lands is more likely.

The evidence for invasion and warfare is more substantial in Scotland than in Ireland. Iron Age invasions from the Continent were more likely to have taken place following the wholesale displacement of populations by extreme weather in Northern Europe. The wide distribution of circular broch towers in Northern Scotland and around the Forth, built from around 100 BC as apparent defensive structures, points to social disturbance or warfare.

The continuing harassment of Scotland by the Romans after 71 AD may have been sufficient to cause a founder effect in the Scottish Lowlands once they departed. Although they held Caledonia for only about 40 years, their repeated invasions cleared the buffer zone between Hadrian's Wall and the Antonine wall (see Figure 7). Tribes that opposed the Romans could be almost eliminated: Caesar claimed his conquest of Gaul killed a million people, mostly civilians. Severus invaded Scotland in 209 AD with 40 000 men, claiming to have committed genocidal depredations on the natives. Whether this is accurate, the Roman military presence had been preventing large areas of Scotland's fertile land between the Walls from being developed, acting as a population constraint.

After the Romans left in AD 419, the vacuum between the Walls was filled by incoming populations from all directions. The situation in Scotland is complicated by what has commonly been regarded as an invasion of the west coast of Scotland by Irish Goidelic speakers, who ultimately wrested power from the Pictish autochthonous majority and gave Scotland its rulers, its Gaelic language and its name. The new arrivals were known as *Scots* (a name previously applied by the Romans to Irish raiders).

If one regards the carrying capacity of a region to be largely determined by food production, the population always tends to overshoot following the introduction of new crops and technology, to the point where it can only be sustained in the best seasons, or beyond. Then the excess population can only be relieved by famine and/or by exodus.[30] As the Irish resettlement and cultural renaissance progressed, the excess population spilled over into Scotland and the Western Isles.

The indigenous inhabitants had fiercely resisted Roman expansion for centuries and prevented their settlement of Romanised populations between the Walls, so a settlement in Scotland by less formidable Gaelic-speaking 'Scots' must have been partly consensual. This was facilitated by the adoption of Christianity influenced by Iona and Ireland, into the 'Brythonic enclosure' of Strathclyde in the 6th century AD.

---

[30] Exactly as occurred in Ireland in 1845-9 when the population of Ireland was 30 per cent higher than it is today.

Figure 7. Hadrian's and Antonine Walls, Scotland and Northern England
Source: created by Norman Einstein 2003, Creative Commons.

The expansion of M222 in Scotland is the only clear example we have of what might be a fairly concerted move by a single lineage, a deliberate move by the M222 Ui Neill group into Scotland, establishing the Dalriada overkingdom of Argyll and Antrim and over time gradually eliminating the Pictish A71/L511 p-Celtic confederation from positions of influence. If the move had been random with regard to lineage, we would expect to see a more balanced move of Irish DNA into Scotland.

The alternative to invasion is that M222 is autochthonous, not from England, and grew from its beginnings astride both Ireland and Scotland in the Dalriada area across areas vacated by the putative disaster of 100 BC. There it was later co-opted into Christian and Goidelic culture just as the Picts were. This is less satisfactory however as it does not explain the clear distinction between the two peoples or the apparent advance of M222 northwards and eastwards.

The sudden rise of the mysterious subclade L1335 from residual status to a point where it exceeded native DF21 and L513 and even M222 in Scotland defies explanation. A further complication is the extensive presence of the other major R1b branch U106 on the east coast of Scotland, which was probably associated with intrusions from the Continent, also suffering from post-Roman unrest.

A further complication is the arrival of Saxons from the Continent, who according to Gildas (Frazer 2009:43) were brought in to stop the ferocious Pictish and Scottish invaders boiling from the north and west, but who ultimately revolted and formed their own settlements. There is ample evidence of population exchange between lowland Scotland and the Continent over an extended period, visible within L21 and other haplotypes, but these matters are beyond the scope of this study.

4.2.4 The elite hypothesis
Since genetic genealogists became aware of the presence of the extensive M222 lineage in Northern Ireland, a 'hypothesis of elites' has been advanced as responsible for its prevalence. This is crudely expressed as 'chieftains with many wives and brave warriors with many sons', or more precisely on the ability of elites to afford polygamy and to exclude other men from breeding, in line with an ancient prejudice that virility and manly virtue expresses itself both in battle and male-line procreation. A similar idea features in many myths and pseudohistories within patriarchal societies.

Specific lineages may certainly have a continuing impact within patriarchal cultures with endogamy, where specific bloodlines may be roughly preserved as cohesive entities and may be able to sustain

an elite caste or status. There has probably been no society that meets these requirements more thoroughly than the Celtic Irish.

The Irish derbfine system was one of the most agnatic on record, designed so that land and power would not pass out of the hands of a single male lineage. The derbfine was the set of patrilineal descendants of a common great grandfather. When one of the members died, property was passed and often divided among the rest. Contracts could only be undertaken with the consent of the whole collective, and new chiefs or kings were always elected from within the derbfine of the last leader. This tended to keep the clans or septs geographically segmented. It also kept power fragmented—Ireland had about 150 petty 'kings'.

It cannot be denied that particular families, clans or races have been able to gain and hold power and dictate the directions of a society. However, this does not automatically convert to a 'selective breeding advantage' in the way that Moore et al. (2006) proposed for Ireland or Thomas et al. (2006) for Anglo-Saxon England. If it did, alleged 'inferior races' would have been outbred the world over, but they never have been.[31]

In fact the 'elite hypothesis' has little or no empirical support. Unlike herd animals, humans are poorly equipped for selective breeding. It is not an easy matter to expand a human male-line deliberately or even to sustain one: the history of Eurasian elites is littered with failed dynasties that could not produce heirs. There is nothing to suggest from genealogical studies that the number of male-line descendants of a single man correlates with socioeconomic status or any other variable—except perhaps the negative impact of having a dangerous and life-shortening occupation, such as being a miner, a seafarer or a 'brave warrior'.

The preferred alternative to the 'elite hypothesis' is random expansion from a bottleneck. The Yule process in the limit creates the Pareto outcome of Figure 3, where a few men randomly have large numbers of descendants and many men have few descendants. As the number of descendants is random, and elites are by definition small, it is much more likely that large lineages will descend from poor men than from kings – and this is what the recent genealogical record shows with few exceptions.

The status theory does have a bearing on the structure of growth out of a bottleneck, in that any man who randomly has many sons during the chaos of an unregulated society may be able to secure resources and gain a temporary advantage for his family. Woolf (2007: 21) states, 'In a world in which masculine physical strength counted for much in both labour and coercion, a band of brothers may have brooked little resistance'.[32] He cites, 'It is preferable that a man's lord should be his kinsman', quoting MacFirbhisigh's Law: 'It is customary for great lords that when their families and kindreds multiply, their clients and followers are oppressed, injured and wasted.'

Social constructs like property ownership, religion and dominant language can be changed quickly by incoming elites. These may confer some temporary advantage in breeding for their followers, if for example women only marry men who speak the dominant language or follow the dominant religion. M222 in Scotland was Christian and Goidelic-speaking, which probably enhanced their status.

---

[31] This depends to some extent on whether children of mixed marriages are absorbed into the subject or the master group. However, as an example aboriginal peoples have been able to retain a fair degree of autochthonous Y-DNA (Hammer et al.

[32] There are modern counterparts: in Albania in the lawless period after the collapse of Communism, families moved *en masse* onto public land and began to build houses. The author was told that the size of the land holding depended on the number of sons to defend it.

However any hereditary elite will have many followers with different lineages, and they all will join and benefit from any 'selective breeding advantage', so in the end the random will triumph. It is possible the M222 founders had some sort of assistance from a status advantage in the early years; but more likely the penetration of M222 and the other deep subclades is entirely a random phenomenon, which is strongly supported by the natural Pareto distribution shown in Figure 3.

In short, the trick to a 'founder effect' is to have many sons and grandsons randomly at a time when the effective male population is very low, just preceding a major population advance. This is what the M222 founder must have done – and also his ancestor the L21 founder, thousands of years earlier. They may have been someone of significance, but in all likelihood they were not.

So we contend that status does not create large lineages but the reverse. Relatively large lineages in periods of crisis form a core or seed for clans that are able to claim power in certain types of patriarchal societies. Following disasters that suddenly reduce the breeding population and eliminate social controls, families that by chance have large numbers of sons are able to take and hold empty territory and form local elites. They then place their own junior relatives on adjacent lands and build up a clan holding. Eventually they may employ new languages and religions to support their social status and keep the momentum going. This is what we see following the Irish Dark Age, when new clans suddenly sprang up all over Ireland. A dozen different lineages were affected covering most of the country. The size of the lineages is a chance occurrence, but the fact that many are associated with particular hereditary clans at one time holding regional hegemony is not.

4.3 The slavery syphon

The opposite breeding strategy to being of high status is to be a slave. For the selfish gene that cares nothing for freedom and little for quality of life, slavery can be an effective strategy for settling new territory – just as captive domestic species have been have flourished in locations far from their origins. Slaves are protected by the captors who assist them to survive in a foreign land. They are encouraged to breed to increase their numbers for work and sale, and they till the land so they are usually well-fed. The men may have last choice of the women, but they do not have to engage in punitive wars and dynastic struggles that reduce their numbers, or go on long expeditions that keep them from their wives. Consequently, the genes of slaves may survive and spread well in new lands.

From about 793 AD Viking raiders from Scandinavia[33] began to assault the coastline of the Isles; perhaps the long Christian peace where disputes could be resolved through church and court rather than by brute force had reduced the ability of the Britons to defend themselves. The Vikings occupied most of the Scottish Isles and the Isle of Man initially as 'pirate retreats' and they established large port settlements at York, Dublin and along the south and east coast of Ireland. Much of the Hebridean archipelago became Norse-speaking.

Viking society was a slave society, 'thralls' worked the land allowing the freemen to sail in search of plunder. One of the main reasons for the raiding, according to Woolf (2007) was the partible system of land inheritance, where all brothers inherited the land, so that in a growing population plots soon became so small they were unviable. Partible inheritance and slavery have often gone hand in hand, because slaves work the land but cannot inherit.

The Vikings took vast numbers of slaves to run their agricultural holdings, mostly from now overpopulated Ireland and Scotland. In a single day it is reported they took 1000 slaves from Dublin. These slaves were brought to the Viking homelands, and their genetic inheritance is visible.

---

[33] Originally from Norway, but by 850 AD from Denmark.

Observing the Norse genetic contribution to the larger British population has proved elusive,[34] but the reverse impact of slaves from Britain on Scandinavia's population is easier to see. The L21 incidence in Scandinavia is only about 4 per cent, but still visible and significant. Of 48 samples of L21 from Scandinavia, none are unequivocally Atlantic - in place since the early Bronze. About half belong to post-Roman growth branches like M222, CTS4466 or L1335. All but five show a fairly close relationship (57/67) with Irish or Scottish men. Table 4 gives the breakdown of counts of Nordic L21 in the L21 database by origin and destination.

Table 4. Early and late L21 arrivals in Scandinavia

Destination	*Early*	*Late*		*Total*
	Atlantic	*Irish*	*Scottish*	
Denmark	2	2	1	5
Norway		21	6	27
Sweden	3	10	3	16
Total	**5**	**33**	**10**	**48**

A reasonable conclusion is that about 90 per cent of Nordic L21 men are descended from slaves taken in raids. The samples however are small and this requires further examination.

Germanic societies on the Continent were organised on a similar basis to the Scandinavian. There is some similarity between the L21 distributions in Scotland, the Low Countries and Germany (see Table C3), which may be due to documented exchanges in the early Middle Ages or to earlier slaving raids from the Continent.

A proper investigation of the complex history of the continental Insular Atlantic L21 enclaves would require better data and deeper investigation. It is from the Continent that the Saxon and Norman invasions of Britain were launched, by people who were in part descendants of the Atlantic civilisation, and it is not surprising it has been difficult to isolate their genetic heritage in Britain.

4.4 The Diaspora

A thousand years after the post-Roman expansion, the British Empire began in 1607 with the permanent settlement of the colony of Virginia. This was soon followed by about 20 more colonies on the eastern seaboard of North America. By this time, Britain's forests and other natural resources of the time were essentially exhausted and investment companies formed with the intention of profiting from the virgin land in the New World. By 1670, about 120 000 British were in the New World, and by 1770 2.1 million.[35] A considerable number of men in our L21 database are descended from settlers in 17th century Virginia and Carolina.

From the 1840s, much of the population of Ireland, Scotland and Cornwall proceeded abroad as economic refugees. About 10 million Irish have emigrated, considerably more than the current population of Ireland, and today over 40 million North Americans claim Irish heritage. Similarly, following the Highland Clearances and the dissolution of the Clans around 1750, the Scots began to emigrate (see Beaty 2009 for a good account). About 50-million people identify as being of Scots or Scots-Irish heritage, even though the population of Scotland is only 5.3 million. One could say that if slavery is good for spreading genes, eviction and persecution may be better.

---

[34] Except in the case of the Shetland Islands and Orkney (Wilson et al. 2001, McEvoy et al. 2006). The problem lies in finding lines that are demonstrably Norse.
[35] https://web.viu.ca/davies/h320/population.colonies.htm

This large influx is an excellent sampler of Britain's population from the 1600s to 1800s, and it is not surprising that the distribution of L21 subclades in North America is very much like that of the British Isles as a whole (see Appendix A). While founder effects in North America are not apparent at the broad level, they are readily apparent at the surname level (see Flood 2013 for a good example). As always the statistical rule of colonisation is: few men, founder effects and redistribution; many men, expansion of the original distribution.

5. The changing distribution of L21 – Skyline methods

In the last two sections we made an attempt to establish the pre-Roman distribution of L21 subclades simply by removing the major expansive subclades from the post-Roman era. This is a fairly limited methodology as quite a number of other branches expanded late in a small way, and truncating only the large branches will underestimate their contribution. We have good reason to believe for example, that England had a considerably larger proportion of L21 in early times.

Another way to analyse the data is through branch analysis methods that fall under the general heading of skyline analysis (Drummond et al 2005). This can be done either with SNPs or STRs.

5.1 Skyline and SNPs

With SNPs the method is straightforward in theory. We presume the population of each subclade at any time is proportional to the number of branches in existence at that time, as a skyline plot does (see Batini et al. 2015).

Table 5. Distribution of L21 subclades now and around 50 AD using SNP and STR skylines

	Incidence % of R-L21 (SNP)[a]		Incidence % of R-L21 (STR)[b]	
Subclade	Present	50AD	3	25
DF49	15.0	9.8	22.7	11.4
DF21	15.9	14.4	19.7	15.1
Z253	11.9	13.4	11.4	14.1
L513	9.1	10.4	10.6	8.2
L1335	7.4	0.8	9.6	0.9
Z255	4.7	1.8	5.4	1.4
FGC11134	9.9	7.3	4.9	3.5
DF41	4.6	5.5	3.3	6.0
S1051	1.1	2.4	2.2	4.7
Z251	5.5	8.7	2.1	7.5
FGC5494	2.9	6.7	2.1	6.6
DF63	2.9	3.5	1.7	5.2
Other	8.8	16.1	4.2	15.3
N	1291	492	2383	348

Source: a) http://www.ytree.net/; b) L21 database, 111 marker.

The main points of interest in the first two numeric columns of Table 5 are the advance of DF49 (M222) and the two formerly residual subclades L1335 and Z255, using share lost to the English subclades and the remaining residual subclades.[36]

5.2 Skeleton skylines and STRs

A similar skyline process may be followed with STRs by creating a 'skeleton' which is an extension of the process we go through to weed out close relatives. For a test depth k, within each subclade we include only one of any pair with genetic block distance (GD) < k, so that all genetic distances in the skeleton are k or greater. As k increases we go progressively back in time to see only the branches that existed at that time. This allows us to use the full L21 dataset with country of origin signifiers.[37]

The simplest skeleton algorithm has been used, in which all records are removed with GD closer than k to the first item in the list, proceeding iteratively through the remaining records in the list. This process gives non-unique skeleton solutions, depending on the order of the list and favouring the first items; with larger k the results can be fairly different. The process can be bootstrapped by using different orderings of the records and taking the average.[38] The resulting skyline distributions approximate a time progression that is not expected to be accurate but will illuminate the basic trends.
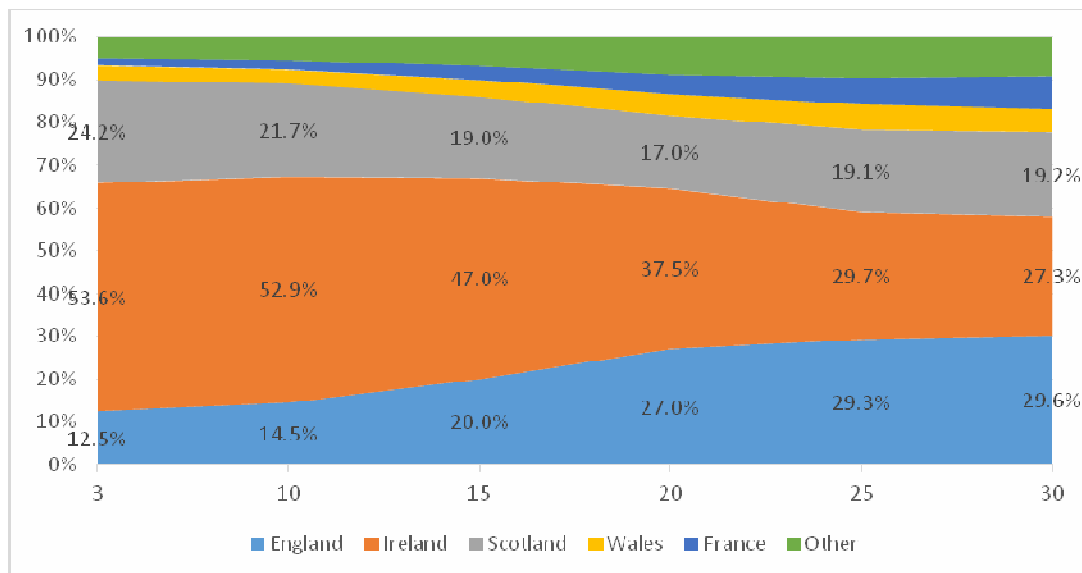

Figure 8. Skeleton skyline distributions, 111 markers, country of origin
Note: Rough time equivalents (k = 3, 10, 15, 30, 25, 30) (AD 1750, 1250, 900, 550, 200, BC 150).

Figure 8 shows the fraction of European L21 in different countries as the skyline progresses backward. At k=25, which approximates the pre-Dark Age distribution of L21, England and Ireland each have about 30 per cent of L21 and Ireland slightly less, while Scotland has 19 per cent and the

---

[36] By comparison with our other sources we see that the two largest deep subclades L1335 and DF49>M222 are undertested in NextGen, while the smaller subclades are relatively overtested, probably because the customers have been notified they have something unusual that needs testing.

[37] In fact only the 2137 records with 111 STR markers tested are used, because the finer resolution of the larger set of markers reduces the time spread around k,

[38] Two extreme solutions are given by putting the most isolated records at the top or at the bottom while progressively stepping outwards. The first solution marks out the extreme boundaries of each subclade, while the second marks central points within the structure. The results here are the average of these two, plus three random distributions.

Continent about 16 per cent (see Table C4 for details). The continental contribution falls steadily from *k*=25 onwards. From *k*=20 or post-Roman times, the English share rapidly slides down to the present-day level of 12.5 per cent and the Continent to residual levels,[39] while the Irish contribution doubles and the Scottish share increases.

The total population probably grew at a fairly constant rate across the geographic range, so we are not witnessing differential rates of population growth; but most probably a pushing back westwards of high-L21 Insular Atlantic populations. In line with the historical record, the Continental decline might be attributed to movements of tribes in Gaul displaced by the Romans, and then followed by the expansion of Germanic peoples to the Atlantic and into eastern England – carrying high proportions of other branches of R1b and other Y-haplogroups. However while we indicate the possibility of further investigation of continental movements, our data are not adequate for this purpose.
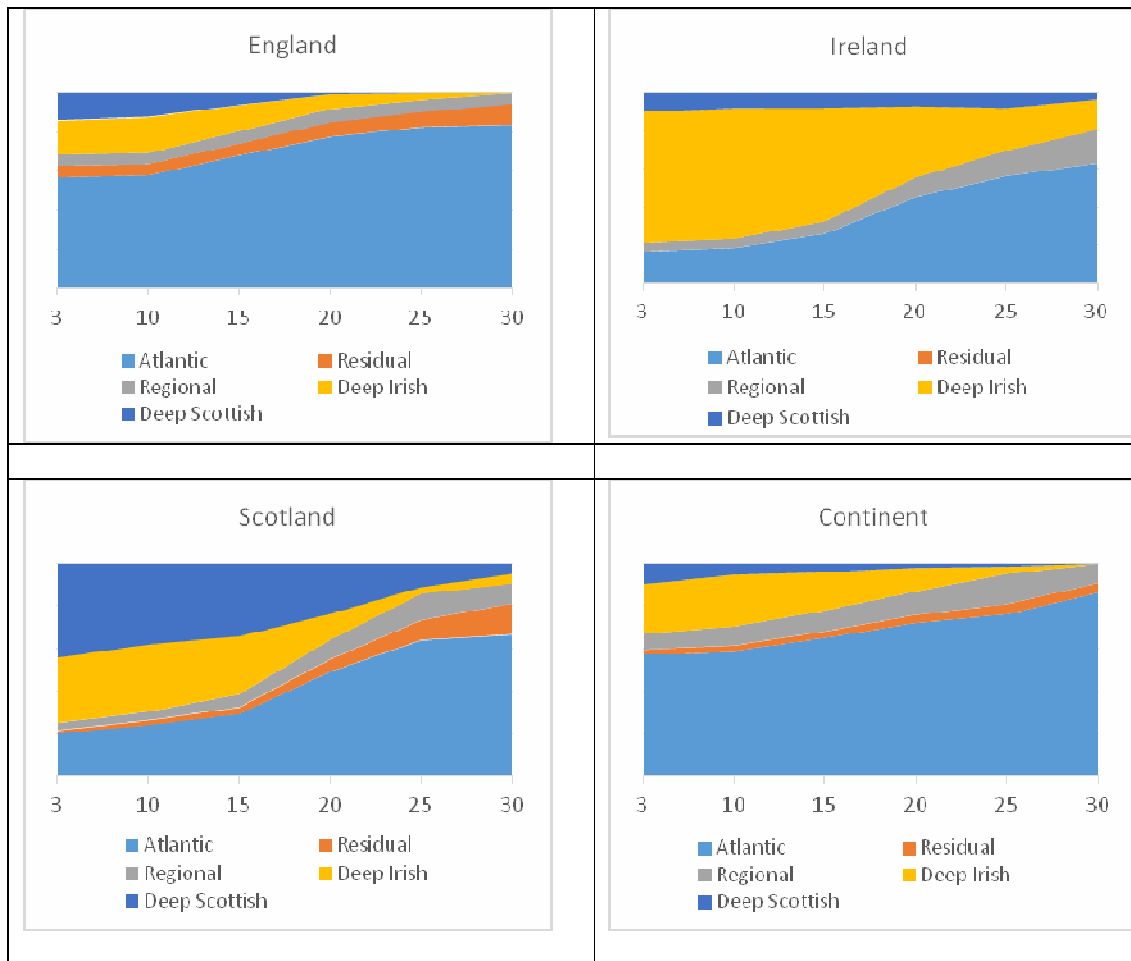

Figure 9. Skeleton skyline, types of L21 subclade

The skeleton skylines for England, Ireland and Scotland are shown in the Excel Table C4, and a summary for the different classes of subclade is given in Figure 9. Some observations are:

---

[39] Note the underrepresentation of the Continent; in Table A3 we recommend it should be multiplied by a factor of 5 in the present. However here we are not so interested in the absolute level but in the fact that it has fallen.

The skeleton skyline also shows the change of the distribution of the various subclades of L21 over time. The last two columns of Table 5 show the equivalents for STRs as the SNP NextGen distributions, and it is reassuring to see they are very close for the two methods (correlations 0.89 and 0.94, for the distributions at present and 50 AD)—an extremely good fit for the earlier distributions, given that they have been obtained by different methods using different data. The skeleton estimates are probably more accurate, because of NextGen testing bias (see Appendix 1).

- The overall distribution at k=10, the beginning of the Middle Ages, is similar to the present day;
- England and the Continent taken as a whole have a similar subclade distribution, which stays fairly constant over the full skyline, except for recent creep of the deep subclades at the expense of smaller Atlantic subclades;
- Initially, the distributions in England, Ireland and Scotland are not dissimilar except that Ireland has considerably more DF21 and (apparently) very few residual subclades, while England has little L513 and twice as much of the mid-sized Atlantic subclades Z251 and FGC5494;
- In Ireland and Scotland the rise of the deep subclades at the expense of the Atlantic subclades is shown very clearly. Most of the change takes place between r=25 and r=15 (approximately 200-900 AD). The Irish deep subclades have a strong presence in Scotland, presumably due to overflow from Ireland. Again, the results are consistent with severe shocks and foundation events in Ireland and Scotland about 2000 years ago, from which the L21 distribution is still recovering.

It is tempting to use the skeleton skyline to try to attribute an origin to various branches, as the process eventually throws up a single representative or several widely separated representatives of each branch; however the method tends to deliver the most common geographical presence in each branch, which may not be the place of origin.

Conclusions

In the words of Wang et al (2014), 'the Y-chromosome is a superb tool for inferring human evolution and recent demographic history from a paternal perspective.' Our picture of the past prior to written history is incomplete and based on very limited evidence and many preconceptions. However just as genetic genealogy can assist in unravelling the recent past when records fail, phylogenealogy and the human Y-haplogenetic tree can reveal details of unsuspected population events that may have affected the later history of nations. Using the Y-haplotree, we have been able to reveal substantial changes and trends that have not been obvious using other methods of analysis.

We presume that the major haplogroup known as Western R1b was spread throughout Atlantic Europe by the people known as the Beaker Folk, who were seafarers seeking tradeable resources. The expansion of the Beaker people over a narrow period of a few hundred years, establishing widely separated colonies in key locations by boat, has meant that different subclades of Western R1b have become associated with particular settlements, and we associate L21 with south-west Britain. Within the L21 lineage we can see evidence of an extremely rapid expansion of descendants of a single man, who peopled Ireland and a good part of Britain, northern France and the Middle Rhine within the space of a few hundred years, apparently meeting little opposition from the existing inhabitants. This occurred within the context of a wide ranging water-based network of trade and culture.

The evidence for a dispersal from Britain includes:

- the larger variance of L21 in England than anywhere else, and the presence of about 30 distinct subclade branches;
- south-west England as the major focus of Beaker activity in Northern Europe, at the centre of the L21 range, and having large easily accessible alluvial deposits of key metals;
- L21 on the Continent taking the form of small random samples from the English distribution;
- all but one of the major L21 subclades being 'Atlantic'—spread throughout the range of the Atlantic culture, with an early Bronze presence in south-west England defined by two or more widely separated examples;
- no support for any other origin for L21 or its subclades.

During the period of the Atlantic culture, L21 was a primary signifier of an Insular Atlantic people in the Isles and beyond. Beaker settlements on the adjacent Continent seem to have been small client arrangements, though the presence of L21 today in the key areas shows their genetic influence persisted. Subsequent invaders of Britain such as the Belgae, Saxons and Normans had a British admixture from the early Bronze Age, making their DNA rather difficult to distinguish from the English population.

Several major subclades DF21 and L513 found their way into northern and Irish populations in greater proportions. There is some indication from the Y-haplotree that Scotland underwent a population expansion during the Bronze Age Climate Optimum from about 1600 BC, so differential rates of population growth involving founder effects may be responsible for this limited early differentiation of subclades.

It has been suspected for a long time because of the very few male lineages hat have come down to us that the human genome has been subject to intermittent pruning involving substantial decreases in genetic diversity probably resulting from natural disasters, epidemics or extremes of warfare. For the first time one of these has been pinpointed, in Ireland around 100 BC, also affecting Scotland but not England to any degree.

At this time, more than half way through its history, a major reorganisation of L21 took place in Ireland and parts of Scotland. The Irish male effective population fell to very low levels, so significant as to be equivalent to a near-extinction and resettlement by survivors. This hitherto undocumented population collapse is probably due to an extreme (2500 year) weather event accompanied by famine, epidemic and opportunistic invasion and warfare. In support of this:

- a dozen L21 'deep subclades' that had been residual for millennia suddenly appeared at the same time from nowhere and grew rapidly. One of these (M222) grew extremely rapidly over a few centuries to become the largest subclade of L21;
- the long tail of the Pareto distribution of L21 subclades was almost extinguished in Ireland;
- the isolation of Ireland and its exposure to the Atlantic has made it vulnerable to weather, famine and epidemics, as strongly hinted in the Irish pseudohistories;
- severe storm surge and famine were recorded by the Romans in Northern Europe from 120-114 BC, sufficient to displace a number of tribes and change the coastline;
- historians have made reference to a 'Dark Age' of fortified warfare in Ireland and a population decline prior to 400 AD.

This decline was followed by a period of rapid population growth and re-peopling which took the population above pre-disaster levels, during which Ireland underwent a major cultural renaissance and the current subclade structure of L21 was laid down. About 70 per cent of L21 belongs to 'deep subclades' from this period. The size of particular subclades has been established randomly, however

the patrilineal and fragmented nature of Irish society caused these new deep subclades to be confined to particular areas and clans to a fair extent, a spatial legacy that is still visible today. The very 'long thin' lead-ins to the new subclades meant that their founders had drifted far from the Atlantic Modal in many cases, so that the 'clusters' could be easily identified by STR signatures.

In Scotland much the same happened with the emergence of several new large 'deep subclades' especially L1335. However here the situation was complicated by prolonged harassment by the Romans that kept the population 'between the Walls' at minimal levels. With the departure of the Romans, the gap was filled by Irish populations that had overextended their local carrying capacity, and invasion from the Continent and England. During the same period Insular Atlantic DNA seems to have been pushed back on the Continent and in England. This substantially lowed the contributions of England and the Continent to the total L21 population.

Around 90 per cent of L21 in Scandinavia dates from the Viking period and is probably attributable to prisoners and slaves brought back to Scandinavia.

We have uncovered no evidence of any significant back-migration of L21 from mainland Europe to the Isles,[40] though this must have occurred in small quantities.

Phylogenetic methods such as variance, PCI and admixture analysis are found to be unsuitable in their usual form for analysis of L21 because of the multi-staged dynamic subclade development and the *in situ* alterations the haplogroup has undergone.

The final legacy of L21 is that it was carried to the English-speaking Diaspora in such great numbers that no bottlenecking is evident; in fact the structure of L21 seems to have been better preserved abroad than in the Isles These foreign descendants of an ancient Bronze Age lineage have come together collectively to test their own DNA and construct the detailed Y-haplotree that has made this analysis possible. Embedded in that haplotree is a rich imprint of events that happened before literacy or recorded history in the Isles, which this paper has begun to explore.

---

[40] With the exception of the Royal Stewart line DF41>L726, who are known to have come from Brittany in the Middle Ages. The most likely back-migrants would be Z253 or Z251, which have a wide continental presence.

7. References

Batini, C, Hallast, P, Zadik, D, et al. (2015). Large-scale recent expansion of European patrilineages shown by population resequencing. *Nature Communications* 6, Article 7152.

Beaty, K G (2009). Finding the 'Scot' in the Scottish-American: an investigation of Scottish identity through mitochondrial DNA and Y-chromosome markers. MA Thesis, University of Kansas. kuscholarworks.ku.edu/bitstream/handle/1808/5976/Beaty_ku_0099M_10671_DATA_1.pdf, Accessed April 2016.

Bell, A R, Brooks, C and Dryburgh, P R (2007). *The English Wool Market, c.1230–1327* Cambridge: Cambridge University Press,

Bianchi, G G and McCave, I N (1999). Holocene periodicity in North Atlantic climate and deep-ocean flow south of Iceland. *Nature* 397 (6719): 515–7.

Biggins, P (2016). DNA of the Three Collas. www.peterspioneers.com/colla.htm#multiplesepts.

Boattini, A, Lisa, A, Fiorani, O, et al. (2012). General method to unravel ancient population structures through surnames, final validation on Italian data. *Hum. Bio.* 2012, 84, 235–270.

Bradley R (2007). *The prehistory of Britain and Ireland.* Cambridge University Press.

Broun, D (1999). *The Irish identity of the Kingdom of the Scots in the twelfth and thirteenth centuries.* Boydell, Woodbridge.

Busby, G B, Brisighelli, F, Sanchez-D P, et al. (2012) The peopling of Europe and the cautionary tale of Y-chromosome lineage R-M269. *Proceedings Biological Sciences/The Royal Society* 279: 884–892.

Campbell, K D (2007). Geographic patterns of haplogroup R1b in the British Isles. *J Genetic Genealogy* 3 (1), 3.

Cassidy, L M, Martiniano, R, Murphy E M, et al. (2016). Neolithic and Bronze Age migration to Ireland and establishment of the insular Atlantic genome *PNAS* 113(2) 368–373.

Charles-Edwards, T M (2000). *Early Christian Ireland.* Cambridge University Press.

Clarke, D L. (1970). *Beaker pottery of Great Britain and Ireland*. Cambridge University Press.

Cunliffe B (1994). *The Oxford illustrated history of prehistoric Europe*. Oxford University Press.

Cunliffe, B (2010). *Celtic from the west. Chapter 1: celticization from the west: the contribution of archaeology*. Oxbow Books.

Cunliffe. B (2001). *Facing the ocean: the Atlantic and Its peoples, 8000 BC to AD 1500.* Oxford University Press.

Diamond, J (2005). *Guns, germs, and steel: the fates of human societies.* W W Norton & Co.

Drummond A J, Rambaut, A, Shapiro, B and Pybus, O C (2005). Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol Biol Evol* 22(5):1185–1192.

Flanagan, L (1998). *Ancient Ireland: life before the Celts*. Dublin: Gill & MacMillan.

Flood, J (2013). *Unravelling the code: the Coads and Coodes of Cornwall and Devon*. Deluge Publishing.

Flood, J (2016). The conquest of the Atlantic seaboard: Beaker Folk and Western R1b. In preparation.

Fox GW and Lasker GW (1982). The distribution of surname frequencies. *International Statistical Review* 51: 81-87.

Frazer, J E (2009). *From Caledonia to Pictland: Scotland to 795*. Edinburgh University Press.

Hammer, M F, Chamberlain, V F, Kearney, V F, et al. (2005). Population structure of Y-chromosome SNP haplogroups in the United States and forensic implications for constructing Y-chromosome STR databases *Forensic Science International* 164: 45–55.

Jackson, K H (1953). *Language and history of early Britain*. Edinburgh University Publications.

Jaski, B (2013). *Early Irish kingship and succession*. Four Courts Press.

Jope, E M, Morey, J E and Sabine, P A (1952). Porcellanite Axes from factories in north-east Ireland: Tievebulliagh and Rathlin. *Ulster Journal of Archaeology* 15: 31–60.

Jordan, W C (1997). *The Great Famine: Northern Europe in the Early Fourteenth Century.* Princeton: Princeton University Press.

Kennedy, I (2014). The history of M222: a story in six parts. http://www.kennedydna.com/HistoryOfM222.pdf, accessed April 2016.

Leslie, S, Winney, B, Hellental, G, et al. (2015). The fine scale genetic structure of the British population. *Nature* 519: 309–314.

McEvoy, B, Brady, C, Moore, K L T and Bradley, D G (2006). The scale and nature of Viking settlement in Ireland from Y-chromosome admixture analysis. *European Journal of Human Genetics* 14, 1288–1294.

Moore L T, McEvoy B, Cape E, et al. (2006). A Y-Chromosome signature of hegemony in Gaelic Ireland. *Am J Hum Genet* 78: 334–8.

Myres, N M, Rootsi, S, Lin, A A, et al. (2011). A major Y-chromosome haplogroup R1b Holocene era founder effect in Central and Western Europe. *European Journal of Human Genetics* 19**:** 95–101.

Myllyntaus, T (2009). Summer frost: A natural hazard with fatal consequences in pre-industrial Finland. Chapter 3 in Mauch, C and Pfister (eds). *Natural disasters, cultural responses: case studies toward a global environmental history*. Lexington Books.

Ó Cróinín, D (1995). *Early medieval Ireland 400–1200*. Longman.

O'Hart, J (1892). *Irish pedigrees or the origin and stem of the Irish nation*. James Duffy and Co.

O'Rahilly, T F (1946). E*arly Irish history and mythology.* Dublin Institute for Advanced Studies.

Osmon, R (2011). *The graves of the Golden Bear: ancient fortresses and monuments of the Ohio Valley.* Grave Distractions Publications.

Parker-Pearson, M G, Pollard, J, Richards, C., et al. (2013). 'Stonehenge', pp 159–78 in Harding, A and Fokkens, H (eds).*The Oxford handbook of the European Bronze Age*. Oxford University Press.

Pevsner, N (1989). *Cornwall*. Yale University Press.

Pfister, U and Fertig, G (2010). The population history of Germany: research strategy and preliminary results, Max Planck Institite for Demographic research. Working Paper WP 2010-035

Roewer, L, Croucher, P J, Willuweit, S., Lu, T T, Kayser, M, Lessig, R, et al. (2005). Signature of recent historical events in the European Y-chromosomal STR haplotype distribution. *Human Genetics* 116: 279–291.

Rossi P (2015). Self-similarity in population dynamics: surname distributions and genealogical trees. *Entropy* 17: 425–37.

Sherratt, A. G. (1987). Cups that cheered: the introduction of alcohol to prehistoric Europe. In Waldren, W H and Kennard, R C, *Bell Beakers of the Western Mediterranean: definition, interpretation, theory and new site data. The Oxford International Conference 1986*. Oxford: British Archaeology Reports: 81–114.

Standish, C D, Dhuime, B, Hawkesworth, C J and Pike, A W G (2015). A non-local source of Irish chalcolithic and early Bronze Age gold. *Proceedings of the Prehistoric Society* 81: 149–177.

Taylor, J J (1980). *Bronze Age goldwork of the British Isles.* Cambridge University Press.

The 1000 Genomes Project Consortium. (2010). A map of human genome variation from population-scale sequencing. *Nature* 467: 1061–1073.

Thomas, M G, Stumpf, M P H and Härke, H (2006). Evidence for an Apartheid-like social structure in early Anglo-Saxon England. *Proc. Biol. Sci.* 273: 2651–7.

Underhill, P A (2003). Inferring human history: clues from Y-chromosome haplotypes. Cold Spring Harbor Symposia on Quantitative Biology, Cold Spring Harbor Laboratory Press: LXVIII, 487–493.

Wang, C-C, Thomas, M, Gilbert, P, et al. (2014) Evaluating the Y-chromosomal timescale in human demographic and lineage dating. *Investigative Genetics* Dec 2014: 5–12.

Wilson J F, Weiss, D A, Richards M, et al. (2001). Genetic evidence for different male and female roles during cultural transitions in the British Isles. *Proc. Natl Acad. Sci. USA* **98**, 5078–83.

Winney, B, Boumertit, A, Day, T, et al. (2012). People of the British Isles: preliminary analysis of genotypes and surnames in a UK-control population. *Eur J Hum Genet* 20(2): 203–210.

Woolf, A (2007). *From Pictland to Alba, 789–1070.* Edinburgh University Press.

Wright, D M (2009). A set of distinctive marker values defines a Y-STR signature for Gaelic Dalcassian families. *Journal of Genetic Genealogy* 5: 1–7.

Yorke, B (1995).*Wessex in the early Middle Ages*. Continuum International Publishing.

Yule, G U (1925). A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, F.R.S.. *Philosophical Transactions of the Royal Society B* 213: 21–87

Zhivotovsky, L A, Underwood, P A, Cinnoglu, C, et al. (2004). The effective mutation rate at Y-chromosome short tandem repeats, with application to human population-divergence time. *Am. J. Hum. Genet.* 74: 50–61.

.

APPENDIX A. Data

This project has assembled as many L21 and other European Y-haplotype data records as possible. To obtain the fine resolution we need for examining the very detailed and 'bushy' branching of R-L21 in its early years and in Ireland and Scotland nearly 3000 years later, we need to know the proportions of the early Bronze L21 subclades deriving from each country, as well as the 'deep subclades' that expanded in the early Christian era. There are no publicly available academic studies that have produced sufficiently detailed data or that have tested British data sufficiently well.[41]

The FTDNA Y-project commercial database is by far the largest and most comprehensively tested for Y-chromosome STRs and SNPs. It can only be publicly accessed in a partial way, by taking records from various public projects and combining them.

Two datasets from FTDNA were compiled for the project in this way. A general European 'origins' dataset to calculate incidence of L21 and other haplogroups was obtained by combining all the European geographic projects with all the haplotype projects for Europe and eliminating duplicates. To the commercial core was added the databases of a few research studies where these contained L21 typing—which has helped to improve the measures of incidence in Spain and Italy. A total of 27264 records are in this European Origins database. The estimated incidences of all different Y-haplogroups by country are shown in spreadsheet Table C1, along with a list of sources. We only use the L21 results in the paper, but as all haplogroups and major subgroups of R1b and P312 had to be enumerated in doing so, we present the full results.

The R1b incidences shown are somewhat lower than other estimates, for example Campbell (2007) in the British Isles. This appears to be due to more men joining the smaller haplogroup projects than the R1b projects, because R1b is thought to be not 'interesting'. The geographical projects are more balanced, but they are limited in size and usually not as well tested and assembled.

The second dataset is the L21 database, which is entirely from FTDNA projects with the addition of the 1000 Genomes set (as this has been analysed for subclades of L21). It includes STR markers, and is much more thoroughly cleaned than the European Origins dataset.

First, all descendants of the same man have been removed unless they are fairly distant (same ancestor, genetic distance 65/67 or closer are removed), keeping only the record with the most testing. This prevents the database being crowded out by near-relatives. After this procedure, 6276 L21 records remained (Table C2a), including 5002 with 67 STR markers or more (Table C2b). To give an idea of the scale of this collection, Busby et al. (2012) claimed to have 'the largest collection of R-M269 yet assembled' with 2000 records and 10 STRs. Yet although the L21 sample here is nearly eight times as large and much more extensively tested, for some purposes (particularly on the Continent) our sample is still not large enough.

Second, the country of origin has been corrected where possible. Participants in FTDNA testing are asked to provide their earliest confirmed paternal-line ancestor and their 'country of origin'. However only a few geographical projects, CORNWALL and DEVON, formally vet these 'distant ancestor' assignments by checking the existence of the ancestor and the line of paternal descent. The main problem is the descendants of early settlers in North America, particularly from the Virginia

---

[41] The People of the British Isles collection from the Wellcome Trust (Winney et al. 2012) is believed to have good information and was the source of the 1000 Genomes data in Cornwall and Kent, but the database has not been made available for research.

and Carolina colonies, who do not know their European 'roots' and often guess at their origin (Campbell 2007). They may either give England, Ireland, Scotland depending on their surname or some family tradition, or else state 'United Kingdom' or 'Unknown'. These all have to be corrected to 'United States' or 'Unknown', which is a fairly laborious task.

If an ancestor with a date and place of birth is given, the country of origin is set to that country in preference to what is stated (it quite frequently differs, such as ancestors from Northern Ireland who are stated to be from 'Scotland'). If a name and date but no place is given, the place of origin of many of these ancestors may be found in online databases.[42] After this correction about 60 per cent of the records provide a European ancestor/place of origin (see Table A1). However this does not turn out to be as serious an impediment to analysis as one might expect, since the missing values are random and unsystematic. As Table A1 shows, the subclade distribution for both groups with no known European ancestor correlates strongly with the total, showing no bias. The 'Isles Diaspora' group, which consists largely of descendants of 17th century settlers, correlates well with England and Ireland, whereas the 'Unknown Ancestor' group correlates much better with Ireland and Scotland. This is reassuring.

Table A1. Pearson correlations of subclade incidence vectors for the British Isles and missing ancestor categories

	England	*Ireland*	*Scotland*	*English-speaking Diaspora*	*Unknown ancestor*	*Total*
England	1					
Ireland	0.494	1				
Scotland	0.473	0.680	1			
Isles Diaspora	0.809	0.871	0.717	1		
Unknown ancestor	0.454	0.872	0.921	0.799	1	
Total	0.701	0.930	0.840	0.962	0.928	1

The absence of geocoding of ancestry is not the only problem; 670 records have no L21 subclade recorded (see spreadsheet Table C2a).Unfortunately the absence of subclade is systematic because men of different origins have engaged in different amounts of testing, and also because many of the large 'deep subclades' of Ireland and Scotland can be identified easily from STRs alone, without SNP testing. Only 6 per cent of Irish records and 9 per cent of Scottish records do not show a subclade, whereas about a quarter of records from England and Wales have no subclade and 30 per cent of those from the Continent. Thus the places that already have a paucity of data and which are important to the Atlantic hypothesis have the deficiency aggravated by minimal testing.

With 67 markers we are able to increase the size of subclades by searching for clusters (taking as a clustering threshold a genetic distance of 8 or less on 67 markers). This only works on deep subclades with a distinctive founder haplotype signature. This approach is even more accurate with 111 markers, as employed in Section 5.

Timing the Y-haplotree
Vitally important for this paper has been the accurate Y-haplotree for L21 and the timing of key SNPs—most notably L21 itself and its Dark Age subclades. Our original awareness of the implications

---

[42] Such as Geni, Ancestry, Worldconnect.

of haplotree analysis came from scrutinising coalescence times of various SNPs on the www.yfull.com site, which happened to coincide with known archaeological and historical events.

The methodology used in yfull is an operationalised version of the standard method of SNP dating, which depends on counting the average number of downstream mutations within a specific part of the Y-chromosome and then applying standard mutation rates (see Batini et al. 2015, for example). The average number of SNP mutations occurring on the 8-10 million base pairs in a typical Big Y test is one per 140 years or five generations.[43] This method is inexact for calculating the TMRCA for two men (who can each have very different numbers of mutations occurring since their common ancestor). It becomes much more exact as the number of men with a particular mutation increases, because the variance of the mean changes inversely with the number of samples, becoming exact for large haplogroups with many men.

The problem is the pruning of the tree, with most side branches disappearing, leading to long stretches with no branch having large numbers of 'equivalent mutations'. Along such a stretch there might as well be only one descendant. Therefore accurate results may only be obtained for 'bushy' SNPs with many branches. Fortunately L21 and its Dark Age subclades are of this kind. The Rathlin sequencing of ancient R-DF21 genomes matched the prior estimates almost exactly, which is most reassuring.

The yfull L21 sample is quite limited, therefore we found it necessary to turn to the much larger BigTree sample[44] to estimate the age of the Dark Age subclades in Table 3 with accuracy (on yfull these subclades all fell into the post-Roman Golden Age, which made little sense from a historical perspective). A recount was made of all the deep subclades branch by branch, using the SNP information provided in BigTree, and the coalescents were adjusted upwards accordingly.

Another problem is testing, if one wishes to use Big Y SNPs for skyline analysis as in Table 5. Big Y may be taken as a separate test by FTDNA customers; it is expensive so is only undertaken by those with great interest. Some subclades have accordingly undertaken NextGen testing more frequently than others – probably due to advocacy by project administrators or to solve a specific problem. Also, people having 'rare branches' are encouraged to take Big Y, which may cause overrepresentation of the small subclades. Accordingly, the distribution of subclades of L21 from Big Y in Table 5 is significantly different from that shown in Table C2a, although the fit is much better in the past. However these distributional issues are not of concern to the present project.

Weighting correction for spatial bias

It has not been found necessary to correct for spatial bias during the course of this paper, because the randomness of the distribution of subclades as shown in Figure 3 and Table A1 is sufficiently reassuring, at least within the Isles. However there may be circumstances where this is necessary and the appropriate country weightings for L21 are shown in Table A3, for the benefit of other researchers.

The Table says that roughly speaking Irish descendants need to have a weighting factor of 0.6 applied, Scotland 0.25, France 4 and Germany 8, when examining the significance of subclades etc. However this does not take into account the extended and repeated emigration from Ireland and Scotland, whereby more than the current population of these countries actually emigrated over

---

[43] https://www.yfull.com/faq/what-yfulls-age-estimation-methodology/ outlines the method and its sources.
[44] The outputs of Big Y are largely unintelligible in isolation, even to well-informed laymen, so most either upload their results to yahoo chat groups, where they are analysed and incorporated in BigTree, or else they pay for analysis by yfull, which is better on other haplogroups. Both sites make their trees publicly available.

time, while Figure 3 inclines us to the conclusion that the sample is actually a fair representation of global L21, at least as far as the British Isles and English-speaking Diaspora are concerned.

Table A2. Population weightings for L21 by country

	Population 1841 million	L21	Population L21 million	Sample size	Weight
England	13.7	0.202	2.85	398	1.00
Scotland	2.6	0.488	1.27	701	0.26
Wales	1.8	0.495	0.89	81	1.58
Ireland	8.1	0.65	5.27	1327	0.57
France*	4.6	0.41	1.89	67	4
Germany*	7.0	0.35	2.45	44	8

Note: *France here consists only of Normandy, Brittany, Picardy and Alsace-Lorraine. Germany contains the Rhineland; its L21 fraction is taken to be the same as Alsace. These regional populations are taken pro-rata from national 1840 figures, using the current regional population distribution. Source: Online Historical Population Reports www.histpop.org, Pfister and Fertig (2010).

Disciplinary bias – strengths and weaknesses of the dataset

Population genetics is a very different discipline from genealogy. The first is concerned with alleles and their transmission and development, mostly using statistical programs to compare DNA within and between populations. Genealogy is based on historicity and the construction of the family tree using records and making educated assumptions using family tradition and social trends.

Different disciplines have surprisingly different norms and paradigms relating to data. The natural sciences such as genetics prefer primary data, although this is changing as large collections of genomes are established. Social sciences such as economics and geography largely use secondary published data, while genealogy uses administrative data.

From an operational point of view, commercial DNA databases are similar to other 'found datasets' such as administrative data, which generally have to be cleaned and converted to purpose. The FTDNA dataset is a particularly interesting example of an evolving self-contributed online database of a kind that is becoming more and more common, and it is a worthy subject of research investigation in its own right. Unlike special-purpose datasets which are mostly single-use and often discarded, the evolving dataset may be applied to all kinds of problems, whenever the number of records and the ever-increasing level of testing are sufficient for the purpose. Like most evolving administrative or 'found' datasets, extra care and different kinds of data cleaning are required than with primary special-purpose samples. If necessary weightings may be applied to limit bias.

Despite the disciplinary preference for self-collection, many formal published phylogenetic Y-chromosome studies have used the FTDNA database or its more easily accessible Ysearch offshoot to fill out their data. What is often not appreciated or is glossed over is that many research collections do not actually use a formal population sampling strategy.[45] They designate a few geographical areas or population groups and obtain a number of volunteers in each group, usually applying minor screens to improve randomness and local focus. Surprisingly, the manner in which these local samples are collected is rarely stated. The lack of attention to this point is startling, but it is presumed some form of informal submissions model is used, either advertising for participants, or initiating contact until a sufficient number of participants is reached. This is very similar to the way

---

[45] Those that do employ a selection strategy generally have much better data sets – a fine example being Boattini et al. (2012) in Italy who performed a prior surname analysis and selected their sample size on this basis.

FTDNA projects obtain their members – either they are pro-actively approached or they respond to various forms of publicity.

Because of the very limited level of testing in most formally published research databases, the only ones we have been able to find to supplement our databases are the 1000 Genomes Consortium (2010) and Boattini et al. (2012).

The strength of the FTDNA 'crowd-funded' database is the very large number of records, the inclusion of the English-speaking Diaspora, its 'living' evolving nature and the breadth and depth of DNA testing, well beyond what most research studies have been able to afford. The weaknesses are the sporadic geotyping in most FTDNA projects (notable exceptions being CORNWALL and DEVON where applicants are formally vetted) and the strong spatial bias and moderate genotype bias with regard to numbers of participants and levels of testing, which leads to some concerns about representativeness.

We know from Table A2 that the FTDNA database is heavily loaded in favour of English-speaking countries, particularly the USA, and also (apparently) heavily in favour of men of Irish and Scottish descent. This can in theory be corrected by weighting; however it turns out to not be a great concern for R-L21, the pre-Diaspora structure of which appears to be representative and actually appears to have been better preserved in the Diaspora than in Europe, if the number of residual subclades found there is any guide.

APPENDIX B. When STR variance fails

Before the widespread availability of whole-genome testing, one of the principal techniques of phylogenetics made use of the spatial variance of a small number of SNPs as a proxy for the age of the coalescent (TMRCA) of a group of men. This is because of a standard theorem in genetics that the variance of an allele is an unbiased estimator of the number of generations since the common ancestor (Zhivotovsky et al. 2004). Inflated claims were made for the technique: Sun et al. (2009) stated 'microsatellites are accurate molecular clocks for coalescent times of at least 2 million years'.

Phlyogeographers attempted to prove a cline in STR variance across Europe from East to West existed, as proof that agriculturalists settled from the Middle East in this way (Roewer et al. 2005, Rosser et al. 2000). However Busby et al. (2012) refuted a paper by Balaresque et al. (2010) that attempted to use R1b to confirm this cline, and in the process showed that the results depended on the properties of the STR markers used, concluding 'existing data and tools are insufficient to make credible estimates for the age of this haplogroup'.

Table B1. The major L21 subclades up to 100 BC, with variances

Subclade	67 marker variance	111 marker variance	N
DF49*	0.342	0.326	157
L513/DF1*	0.340	0.330	129
FGC11134*	0.332	0.282	20
DF41*	0.329	0.305	58
Z253*	0.326	0.293	215
FGC5494	0.319	0.305	69
S1051	0.290	0.318	74
DF21*	0.271	0.287	333
Z251	0.269	0.292	79
DF63	0.263	0.269	37
S1026	0.252	0.258	26

Note *: As in Table 1, deep subclades have been removed.

The variances of major subclades of early L21 are shown in Table B1. All these subclades are almost the same age as measured by SNPs, so the variances should be the same; however even with the large deep subclades removed there is still a very substantial difference in variance between subclades (which is much worse if the deep subclades are included).

Despite Busby's reservations, STR variance does broadly correlate with average time to coalescence; however it is not a robust or particularly accurate measure. In practice there are three classes of errors that may arise when using variance of STRs for dating coalescence times or the length of time a particular haplogroup has been present in a locality:

- variance depends much more on the mix of subclades than on geographic factors;
- the presence of large recent lineages within a particular subclade will void normality and give false answers;
- particular STRs can show sudden leaps, and data errors or outliers can substantially modify variances.

The main reason that STR variance is a poor predictor of regional difference or of origin is shown in Table B2.

Table B2. L21 variance explained* by region and subclade

Source of variance	Contribution	% of variance explained
Region	0.002957	2
Subclade	0.114356	76
Region*Subclade	0.033216	22
Var(Error)	0.289278	

Note: *Minimum Norm Quadratic Unbiased Estimation

Of the non-random variance in STR markers, 76 per cent is explained by the mix of subclades, 22 per cent by the interaction between subclade and region (for instance, the presence of 'short fat' sub-branches in some regions), and only 2 per cent by country differences (for example, the length of time R-L21 as a whole has been in the country). So three-quarters of the explained variance is due to the presence of different subclades (particularly the deep subclades that have their own STR signature) and the remainder is due to combined effects (the different properties of subclades by region). Only a negligible proportion of variance is due to the region alone.

Table B3 shows the variance of L21 in various countries or regions for different numbers of markers (the results are fairly sensitive to what markers are used), in descending order for 67 markers. Since each country contains more than one subclade, the coalescent (TMRCA) is actually the same; at about 4500 years. The variance gives some idea of the diversity of L21 in each place (by coincidence, it also describes what we suspect was the order of settlement in Europe). The different variances do not imply different lengths of time, they represent the relative diversity of the L21 distribution—'long and thin' or 'short and fat'.

Table B3. STR variances of L21, various locations and regions, 67 and 111 markers

Country/region	67 markers	111 markers
England	0.3439	0.3374
France	0.3373	0.3167
Ireland	0.3366	0.3202
Unknown	0.3300	0.3206
English-speaking Diaspora	0.3295	0.3227
Germany	0.3183	0.3397
Wales	0.3183	0.3397
Scotland	0.3177	0.3242
Hispanic	0.3108	0.2953
Mediterranean	0.2958	0.3002
Scandinavia + Low countries	0.2923	0.3237
All	**0.3352**	**0.3271**

It is reassuring that the variance in the two categories of missing origin (*Unknown* and *English-speaking Diaspora*) is very close to that of the whole sample, showing first that the sample is representative and second that the Great Migration had no obvious founder effects. It is noteworthy

that L21 has only been in North America for 400 years yet its variance is greater there than in Scotland, where it has been present for 4500 years.

Finally, variance is unfortunately a non-robust measure. It is calculated as squares of differences and bad data or outliers can have a substantial effect in a small sample. Measurement error, data transcription errors, or sudden leaps in the values of particular markers such as RecLOHs have to be monitored carefully. Changing the markers used can also produce different results, as Tables B1 and B3 show.

So—while STR variance tells us something useful, it is not the amount of time L21 has been present in a location. It gives a better indication of the age of a particular SNP, but even there it is modified by the internal structure and the proportion of recent expansion in the subclade.

The problematic nature of STR variance as a tool implies that many of the standard methods of phylogeography also do not work on closely related populations and are at best only mildly indicative of any relationship. In particular admixture analysis using Y-STRs, as was commonly employed throughout 1995-2010 in academic papers (e.g. McEvoy et al. 2006), does not work on L21.

For instance, if one was to try to infer present-day proportions of Irish, Scottish or English populations on the Continent using the pre-Roman distributions in Table 5, the results would have no meaning because the populations in Ireland and Scotland had not yet differentiated to any significant degree. In the very long term there are likely to be several widely separated periods of interaction between neighbouring countries, and if the distributions have changed internally in the meantime, the overlays will be hard to interpret by a single measure.

The various correlations between modern distributions are shown in Excel Table C3, which indicates that most places on the Atlantic culture spread are more closely related to England as the original point of distribution than to each other, However there are some correlations between Germany, the Low Countries and Scotland because of late interchange between those places, which deserves investigation using other haplogroups as well as L21 and better continental data.
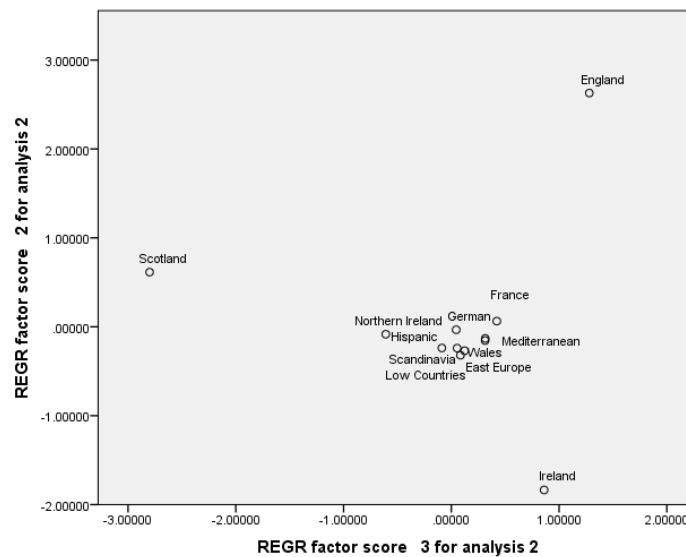
Figure B1. Country scores on Principal Components 2 and 3[46]

A similar concern applies to Principal Components Analysis (PCA) on L21. Figure B1 graphs the country placements on the principal components of variance. The results are similar whether we use STRs or subclade vectors (the proportions in each subclade), but the latter give much better fits, as one would expect from Table B2.

The first three components separate out England, Ireland and Scotland as expressing much of the variance in L21. All the other countries, even Wales, remain tightly clustered in the middle. This is because their subclade distributions were mostly established through the Atlantic culture long before the Isles differentiated.

The same thing will happen using SNPs, so that one must be careful in interpreting any form of variance-based procedure for geographical differentiation of closely related populations undergoing significant internal change over time, including admixture analysis and PCA.

---

[46] Component 1 separates England from Ireland and Scotland and is otherwise uninformative.