

U106 explored: its relationships, geography and history

Report to the U106 group

April 2017 edition

Principal investigator: Iain McDonald

Contents

Part 2: Ancient DNA and migrations around the U106 origin	2
Introduction to ancient DNA testing	2
Ancient DNA results	3
<i>6500-4000 BC</i>	3
<i>4000-3000 BC</i>	4
<i>3000-2500 BC</i>	5
<i>2500-2000 BC</i>	6
Analysis	7

Ancient DNA

DNA extracted from archaeological remains are an invaluable source of the distribution of haplogroups in ancient and prehistoric times, and in the clades that were extant at those times. They provide a method to validate the ages of nodes in the haplotree, and they provide information we can use to deduce the migrations of people in these ancient times.

This document focusses on ancient DNA results in Europe between 6500 BC and 2000 BC. These cover the events both leading up to and shortly after the formation of U106, which most modern dating methods put at around 3000 BC. The maps on the following pages show these ancient DNA results, broken up by period, to highlight the differences between them.

The ancient DNA records used to provide the following maps come from Jean Manco's database:

<http://www.ancestraljourneys.org/ancientdna.shtml>

PROBLEMS WITH ANCIENT DNA

There are many difficulties in testing ancient DNA related to both the preservation of the sample and possible contamination from testers. At the time of writing (March 2017), extracting and sequencing DNA from 45,000-year-old remains is possible, and from 5,000-year-old remains is becoming routine. However, the quality of the results is still much lower than commercial tests. Typically, a modern sequencing test might reach a few dozen good-quality reads on the same base pair. If you find the same allele (A,C,G,T) at that base pair (nearly) every time, it gives you confidence that you have the correct one. With ancient DNA, you are lucky to get one or two reads, though calls are usually only made when a higher number of good-quality reads is reached (say 5x).

This means that the effective coverage of ancient DNA is very variable, and doesn't cover every SNP that a standard sequencing test would. Only in a small proportion of samples is the DNA of sufficiently good quality to call novel variants from the test, and often only broad haplogroups (e.g. "R1b") are published. Several dedicated individuals process the publicly-available raw results from these tests to extract further known SNPs. This may still only identify a test as, e.g. R1b-M269, rather than some more recent variant, but we have to do the best with what we have.

The problems associated with ancient DNA are summarised in the review article by Kivisild (2017):

<https://link.springer.com/article/10.1007/s00439-017-1773-z>

ASSOCIATING ANCIENT DNA WITH ARCHEAOLOGICAL CULTURES

What we get from ancient DNA is a sequence of called variants present in the genes of a single man. Normally, there is enough information to tie that man to a particular culture that lived at that time. Today, an entire culture cannot be adequately represented by a single person, or a single family. There is no particular reason to think that it was any different in pre-historic times. Any ancient culture is therefore going to be a mixture of haplogroups. However, each will have a different admixture of haplogroups, so tracking the ancient migrations of our ancestors requires tracking how these haplogroup admixtures change over geographical space and over time. Over ancient timescales, this is a case of looking at the ratio of haplogroup I to G, I to R, G to R, etc. In more modern times, we need to look at the ratio of more contemporary haplogroups, like R-U106 to R-P312, etc.

The sparse sampling of ancient DNA means that we do not get a complete picture of a haplogroup. For example, if a culture is sampled by a single haplogroup I individual, that doesn't mean the rest of the culture was haplogroup I. Similarly, if there are three or four R-P312 results from a culture, but no R-U106 results, that doesn't mean that R-u106 is not there, merely that it is likely to be present at a much lower percentage.

As with all datasets, our knowledge improves as more ancient DNA is sequenced. True sequencing of ancient DNA has only really been possible in the last handful of years, so we are only starting to get a real picture of how different haplogroups populated Europe. This is a fast-evolving field, with potentially game-changing discoveries being made every few months..

ANCIENT U106 RESULTS

A full analysis of the ancient results pertaining to U106 in particular is given at the end of this document. For now, it is worth forwarning the reader that there has only been one confirmed U106+ burial discovered that can be dated within the 6500 - 2000 BC time period we consider here. That burial is RISE98, who was found buried in the southernmost tip of Sweden, and appears to have lived at some point during the latter centuries of the third millenium BC.

After 2000 BC, the only currently (March 2017) confirmed U106+ burials are from a Roman-era cemetery in York, one of which is R-DF98 and one of which is R-DF96. It is likely that these burials represent gladiators, either from within Roman Britain or captured from nearby invading populations. The isotopic analysis of the R-DF98 burial suggests that he came from a more mountainous region. The paper's authors posit Wales, but other areas of the British Isles (e.g. southern Scotland) remain possible. Only one of the burials mentioned in that paper shows isotopic and autosomal DNA evidence from coming from outside the British Isles.

DESCRIPTION

This page details the archaeological DNA obtained from burials before 4000 BC, which show the context of Europe into which R-M269 arrived. Below are the symbols used in this page:

- Haplogroup I: ● ● ● ● ●
- Haplogroup C: ● ● ● ● ●
- Haplogroup E: ● ● ● ● ●
- Haplogroup G: ● ● ● ● ●
- Haplogroup R: ● ● ● ● ●
- Other haplogroup: ● ● ● ● ●

Opacity scales with age, such that the above represent (from left to right) ages of <6000 BC, 6000-5500 BC, 5500-5000 BC, 5000-4500 BC and 4500-4000 BC.

ANALYSIS

Early in the period where ancient DNA is well sampled, a very strong division is clear between the European part of the former USSR, where haplogroup R dominates, and the rest of Europe, where haplogroup G and I are dominant. There is not enough data to draw a clear line of differentiation between the two populations: likely it was fairly diffuse and dynamic anyway. However, by 4000 BC, it appears that haplogroup R was present on the shores of the Baltic Sea in the north, and the Dnieper-Don valleys in the south.

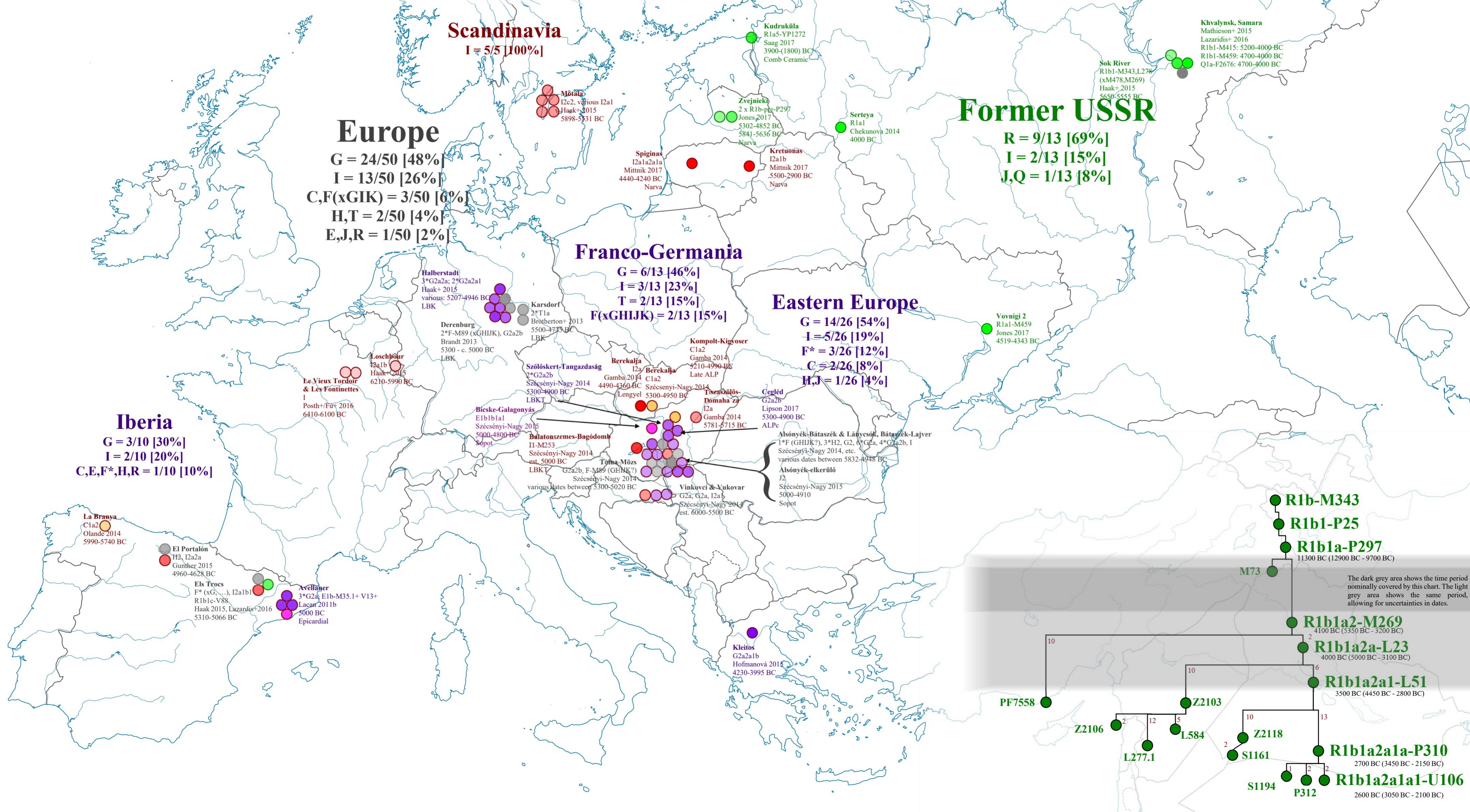
Sporadic outliers of haplogroup R exist in Spain (Els Trocs, c. 5200 BC) and Italy (Villabruna, c. 14,000 BC). The low frequency of these haplogroup R men, compared to their majority in Europe, suggest that these were not the people from whom modern R-M269 populations descend (indeed, the Els Trocs burial is R-V88, which is most commonly found in central and southern Asia these days).

Within the rest of Europe, Scandinavia appears largely haplogroup I, while continental Europe appears mostly haplogroup G, with some haplogroup I and minor haplogroups mixed in. The heterogeneity of the Hungarian results suggests that this mixing was fairly effective by about 5000 BC.

Within the former USSR, there is a fairly good mixture of haplogroups R1a and R1b. There is relatively little data, although there does appear some regional variation: R1b is found with R1a, R1a with R1a. This may be a true geographical difference, or a cultural bias based on individual, powerful families.

6500-4000 BC: the situation in Europe

Updated: 14 July 2016
Dr. Iain McDonald
on behalf of the U106/S21 group



Europe

- G = 24/50 [48%]
- I = 13/50 [26%]
- C,F(xGIK) = 3/50 [6%]
- H,T = 2/50 [4%]
- E,J,R = 1/50 [2%]

Scandinavia

I = 5/5 [100%]

Franco-Germania

- G = 6/13 [46%]
- I = 3/13 [23%]
- T = 2/13 [15%]
- F(xGHIJK) = 2/13 [15%]

Eastern Europe

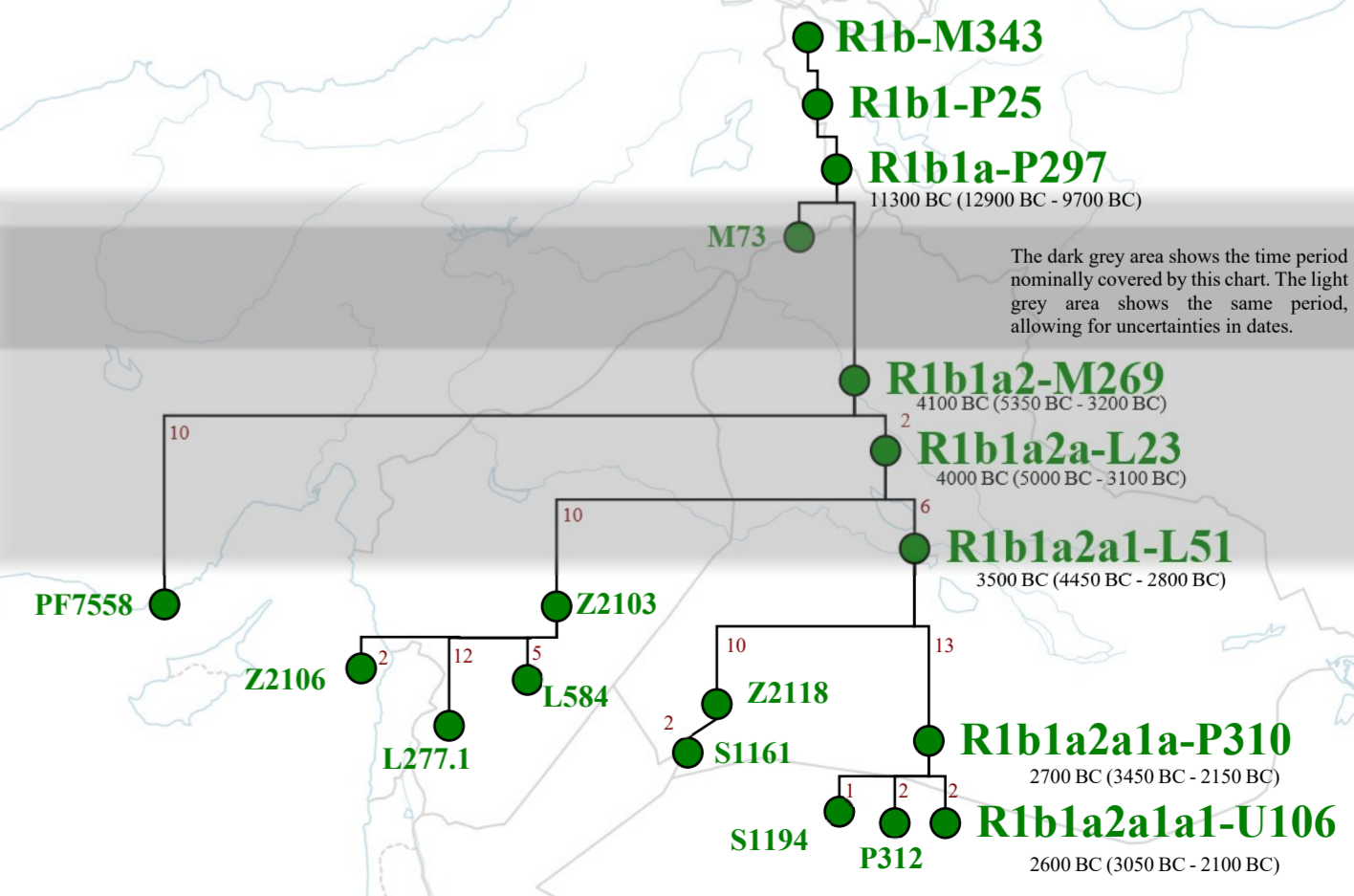
- G = 14/26 [54%]
- I = 5/26 [19%]
- F* = 3/26 [12%]
- C = 2/26 [8%]
- H,J = 1/26 [4%]

Former USSR

- R = 9/13 [69%]
- I = 2/13 [15%]
- J,Q = 1/13 [8%]

Iberia

- G = 3/10 [30%]
- I = 2/10 [20%]
- C,E,F*,H,R = 1/10 [10%]



The dark grey area shows the time period nominally covered by this chart. The light grey area shows the same period, allowing for uncertainties in dates.

DESCRIPTION

This page details the archaeological DNA obtained from burials between 4000 and 3000 BC, which show Europe during the initial R-M269 expansion, before U106 formed. Below are the symbols used in this page:

Haplogroup I: [Orange circle]

Haplogroup C: [Light orange circle]

Haplogroup E: [Pink circle]

Haplogroup G: [Purple circle]

Haplogroup R: [Green circle]

Other haplogroup: [Grey circle]

Opacity scales with age, such that the above represent (from left to right) ages of 4000-3800 BC, 3800-3600 BC, 3600-3400 BC, 3400-3200 BC and 3200-3000 BC. Many dates are uncertain, but the central (usually most likely) estimate is used to select the symbol colour. The Serteya burial, dating to 4000 BC, is carried over from the previous page. Burials dating to 3000 BC are carried over onto the next page.

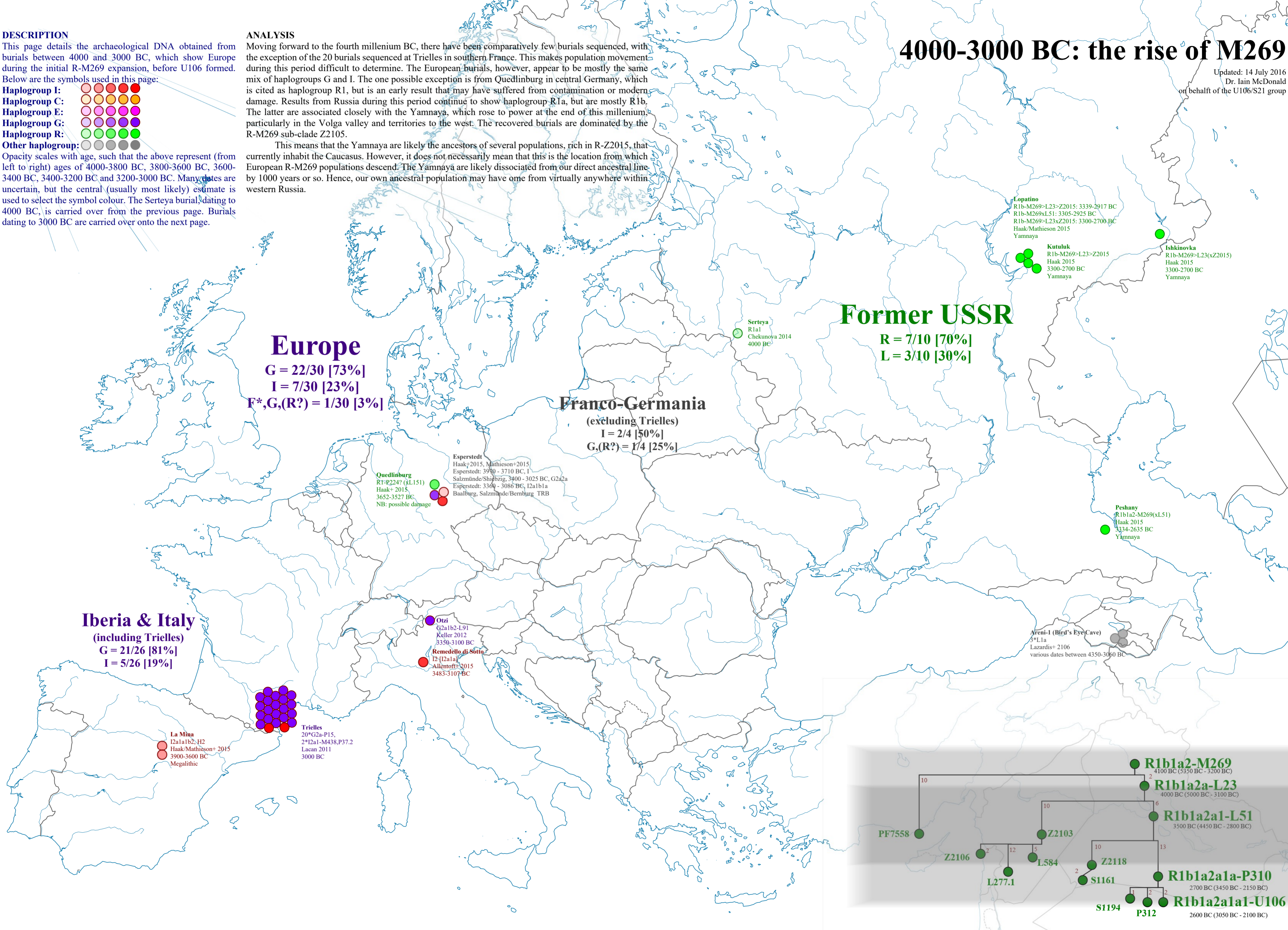
ANALYSIS

Moving forward to the fourth millenium BC, there have been comparatively few burials sequenced, with the exception of the 20 burials sequenced at Trielles in southern France. This makes population movement during this period difficult to determine. The European burials, however, appear to be mostly the same mix of haplogroups G and I. The one possible exception is from Quedlinburg in central Germany, which is cited as haplogroup R1, but is an early result that may have suffered from contamination or modern damage. Results from Russia during this period continue to show haplogroup R1a, but are mostly R1b. The latter are associated closely with the Yamnaya, which rose to power at the end of this millenium, particularly in the Volga valley and territories to the west. The recovered burials are dominated by the R-M269 sub-clade Z2105.

This means that the Yamnaya are likely the ancestors of several populations, rich in R-Z2015, that currently inhabit the Caucasus. However, it does not necessarily mean that this is the location from which European R-M269 populations descend. The Yamnaya are likely dissociated from our direct ancestral line by 1000 years or so. Hence, our own ancestral population may have ome from virtually anywhere within western Russia.

4000-3000 BC: the rise of M269

Updated: 14 July 2016
Dr. Iain McDonald
on behalf of the U106/S21 group

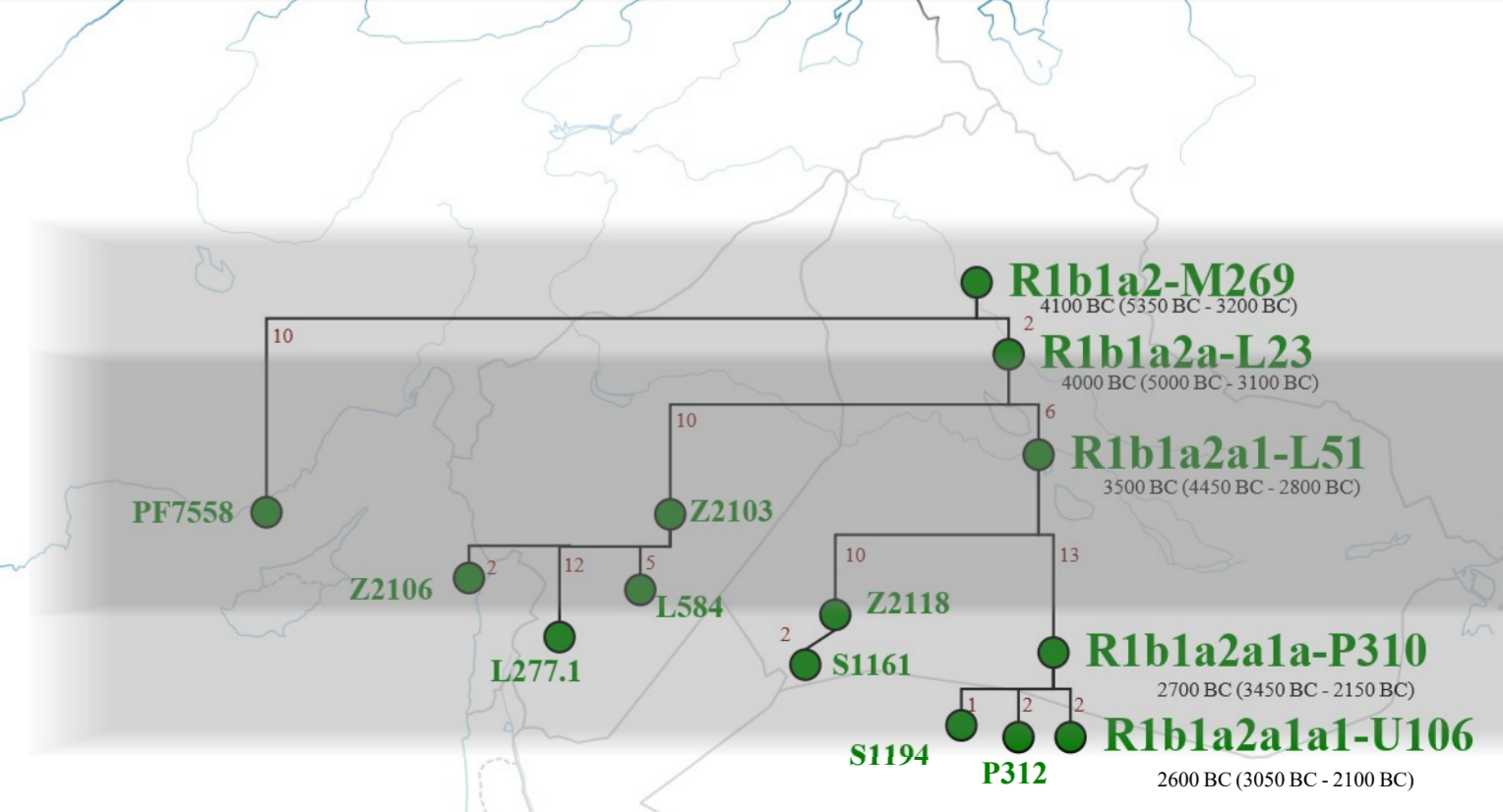


Europe
G = 22/30 [73%]
I = 7/30 [23%]
F*,G,(R?) = 1/30 [3%]

Franco-Germania
(excluding Trielles)
I = 2/4 [50%]
G,(R?) = 1/4 [25%]

Iberia & Italy
(including Trielles)
G = 21/26 [81%]
I = 5/26 [19%]

Former USSR
R = 7/10 [70%]
L = 3/10 [30%]



La Mina
I2a1a1b2, H2
Haak/Mathieson+ 2015
3900-3600 BC
Megalithic

Trielles
20*G2a-P15,
2*I2a1-M438,P37.2
Lacan 2011
3000 BC

Otzi
G2a1b2-L91
Keller 2012
3350-3100 BC

Remedello di Sotto
I2-[I2a1a]
Allentoft+ 2015
3483-3107 BC

Quedlinburg
R1-P224? (xL151)
Haak+ 2015
3652-3527 BC
NB: possible damage

Esperstedt
Haak+ 2015, Mathieson+ 2015
Esperstedt: 3990 - 3710 BC, I
Salzmünde/Shiebig, 3400 - 3025 BC, G2a2a
Esperstedt: 3369 - 3086 BC, I2a1b1a
Baalburg, Salzmünde/Bernburg TRB

Serteya
R1a1
Chekunova 2014
4000 BC

Lopatino
R1b-M269>L23>Z2015: 3339-2917 BC
R1b-M269>L51: 3305-2925 BC
R1b-M269>L23>Z2015: 3300-2700 BC
Haak/Mathieson 2015
Yamnaya

Kutuluk
R1b-M269>L23>Z2015
Haak 2015
3300-2700 BC
Yamnaya

Ishkinovka
R1b-M269>L23(xZ2015)
Haak 2015
3300-2700 BC
Yamnaya

Peshany
R1b1a2-M269(xL51)
Haak 2015
3334-2635 BC
Yamnaya

Areni-1 (Bird's Eye Cave)
3*L1a
Lazardis+ 2106
various dates between 4350-3060 BC

DESCRIPTION

This page details the archaeological DNA obtained from burials between 3000 and 2500 BC, which show Europe during the initial R-M269 expansion, before U106 formed. Below are the symbols used in this page:

- Haplogroup I:**
- Haplogroup C:**
- Haplogroup E:**
- Haplogroup G:**
- Haplogroup R:**
- Other haplogroup:**

Opacity scales with age, such that the above represent (from left to right) ages of 3000-2900 BC, 2900-2800 BC, 2800-2700 BC, 2700-2600 BC and 2600-2500 BC. Many dates are uncertain, but the central (usually most likely) estimate is used to select the symbol colour. Burials dating to 3000 BC are carried over from the previous page. The Ajvide burial on Gotland is included due to its large uncertainty, but good representation of the expected population (as observed in previous and later epochs). Tildes (~) prefix better-known SNPs which are phylogenically equivalent in tested modern populations.

ANALYSIS

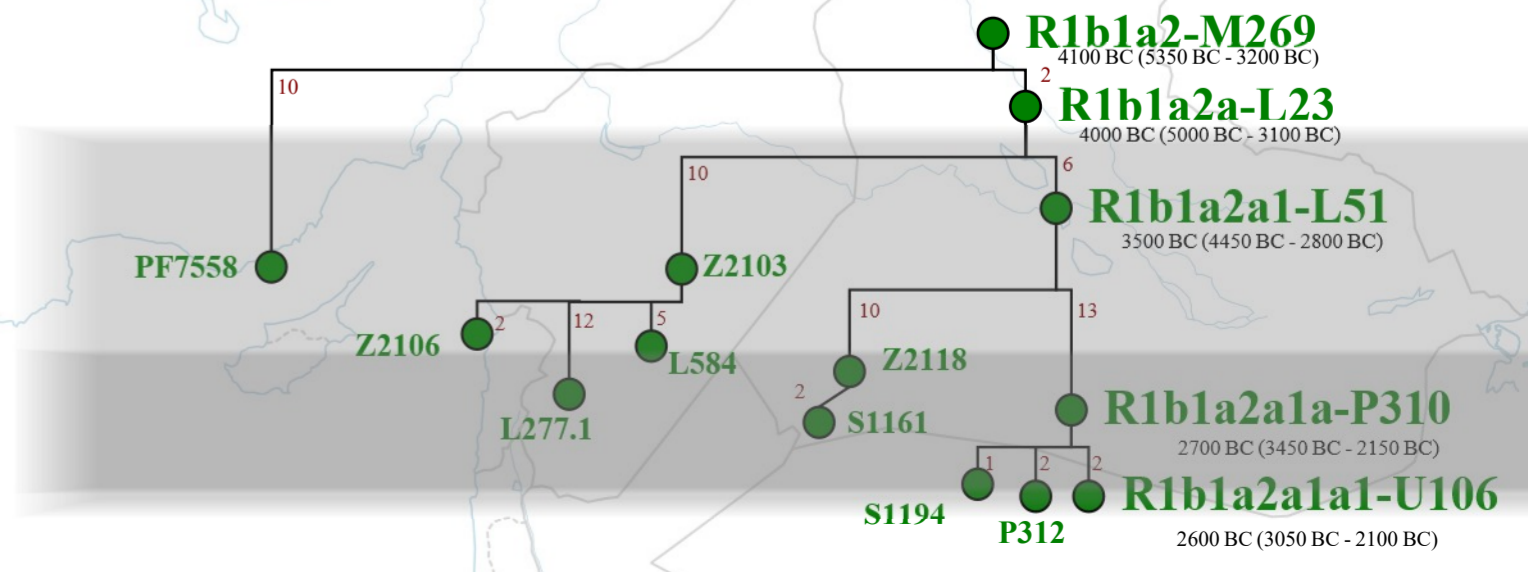
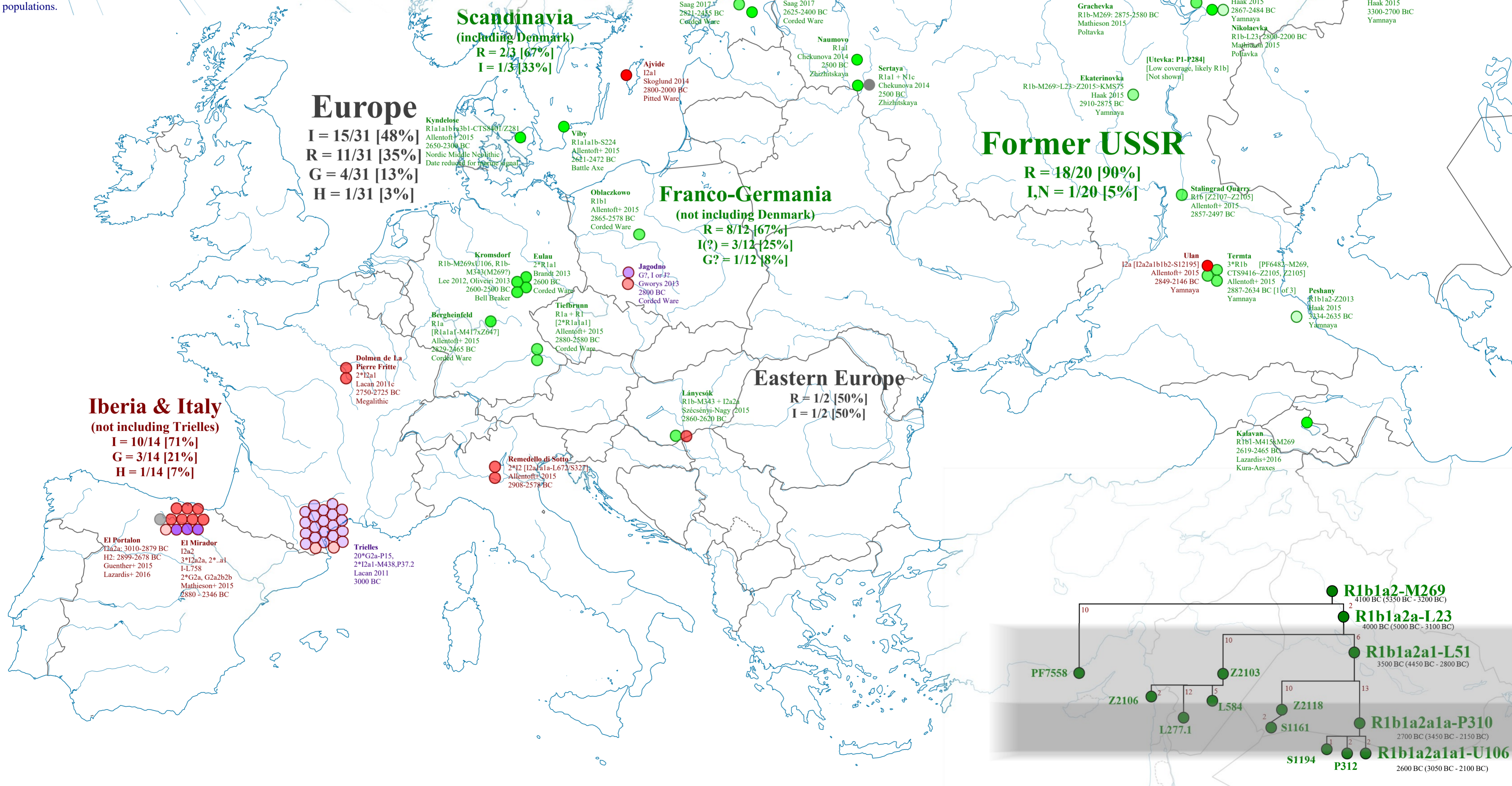
The period between 3000 and 2500 BC marks a seismic change for some of Europe's populations, and a continuation for others. The population of the former USSR remains largely unchanged. A population of R1a occupies western Russia and Estonia, the Yamnaya R1b-Z2015 burials keep going and spread down into the Caucasus. The population of southern Europe shows a strong concentration of haplogroups I and G.

By contrast, northern and eastern Europe experience a complete population replacement. It is unclear whether this replacement in the ancient DNA truly reflects a systematic removal (slaughter?) of the central European population, or simply a change in who warranted a sufficiently lavish burial. However, a haplogroup R population of a few percent suddenly becomes around two thirds of burials. These early burials are mostly R1a, belonging to the Corded Ware culture and associated Nordic Battle Axe and other late Neolithic cultures. The umbrella group of Corded Ware cultures crossed the northern regions of continental Europe around 2800 BC, occupying from Poland to the North Sea. They are a candidate group responsible for bringing the proto-Indo-European language and culture to Europe, which would be consistent with the power (and perhaps population) replacement they wrought. Unfortunately, no ancient DNA from other contemporary cultures (e.g. Baden, Globular Amphorae) exists for comparison.

It is not clear whether the R1b groups followed the R1a men as part of the Corded Ware culture, or whether they came in a separate migration. The R1b1 burial at Oblaczkowo, in central Poland, suggests they came with the Corded Ware migration. However, this is only one man, and his DNA was tested at sufficiently low resolution that no further haplogroup could be assigned. Alternatively, the R1b-M343 burial in Lányesók, in southern Hungary, may represent our ancestors arrival into this location before 2620 BC, or it may represent the expanding Z2105 population, which is today largely confined to small percentages of the Balkans and south-eastern Europe. The R1b-M269 burial at Kromsdorf, in central Germany, is more clear. His results suggest that R-M269 groups were partly associated with the reflux of Bell Beaker cultures that interacted with the Corded Ware culture around this region in the middle of the third millennium BC.

3000-2500 BC: M269 invades Europe

Updated: 14 July 2016
Dr. Iain McDonald
on behalf of the U106/S21 group



DESCRIPTION

This page details the archaeological DNA obtained from burials between 2500 and 2000 BC, which show Europe during the initial R-M269 expansion, before U106 formed. Below are the symbols used in this page:

- Haplogroup I:**
- Haplogroup C:**
- Haplogroup E:**
- Haplogroup G:**
- Haplogroup R:**
- Other haplogroup:**

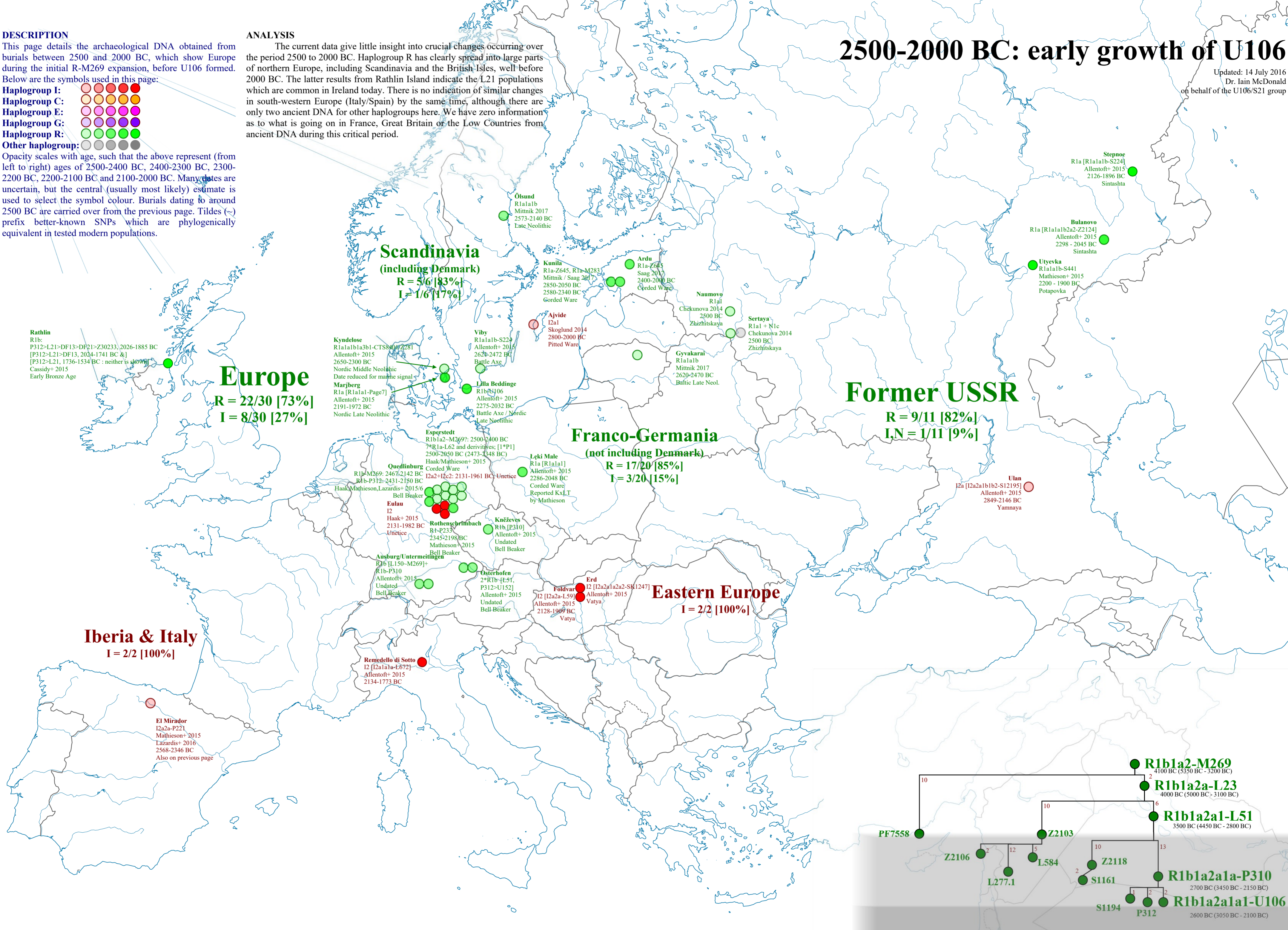
Opacity scales with age, such that the above represent (from left to right) ages of 2500-2400 BC, 2400-2300 BC, 2300-2200 BC, 2200-2100 BC and 2100-2000 BC. Many dates are uncertain, but the central (usually most likely) estimate is used to select the symbol colour. Burials dating to around 2500 BC are carried over from the previous page. Tildes (~) prefix better-known SNPs which are phylogenetically equivalent in tested modern populations.

ANALYSIS

The current data give little insight into crucial changes occurring over the period 2500 to 2000 BC. Haplogroup R has clearly spread into large parts of northern Europe, including Scandinavia and the British Isles, well before 2000 BC. The latter results from Rathlin Island indicate the L21 populations which are common in Ireland today. There is no indication of similar changes in south-western Europe (Italy/Spain) by the same time, although there are only two ancient DNA for other haplogroups here. We have zero information as to what is going on in France, Great Britain or the Low Countries from ancient DNA during this critical period.

2500-2000 BC: early growth of U106

Updated: 14 July 2016
Dr. Iain McDonald
on behalf of the U106/S21 group



**Scandinavia
(including Denmark)**
R = 5/6 [83%]
I = 1/6 [17%]

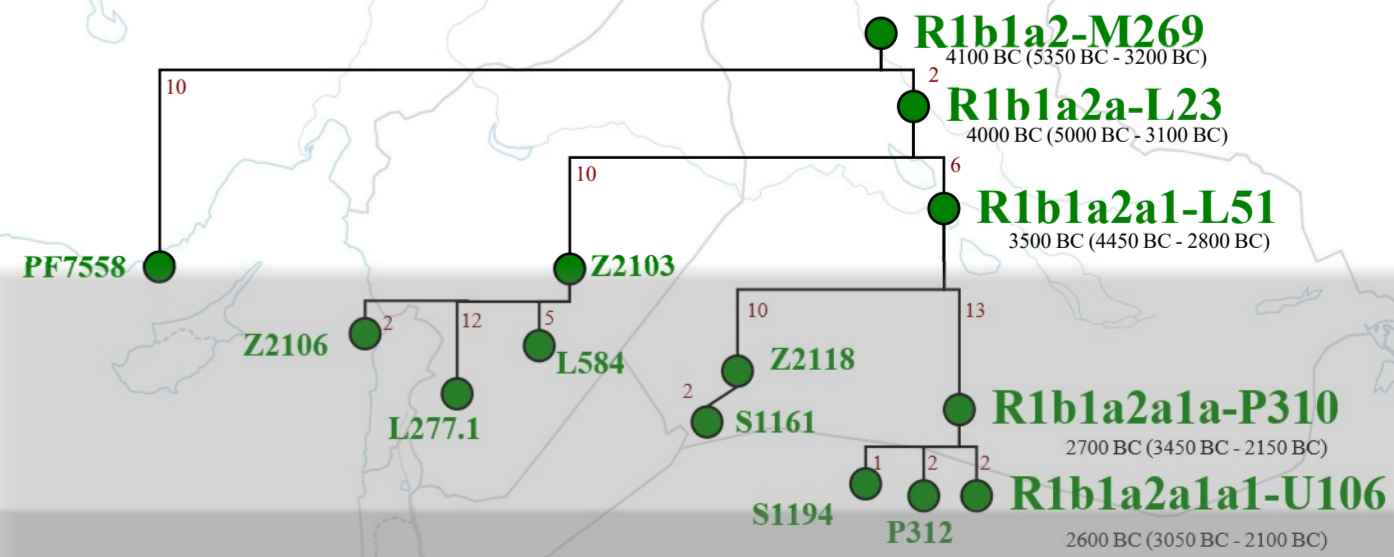
Europe
R = 22/30 [73%]
I = 8/30 [27%]

Former USSR
R = 9/11 [82%]
I,N = 1/11 [9%]

**Franco-Germania
(not including Denmark)**
R = 17/20 [85%]
I = 3/20 [15%]

Eastern Europe
I = 2/2 [100%]

Iberia & Italy
I = 2/2 [100%]



Analysis

There are a variety of options, centred around a theme, regarding the arrival of our ancestors into western Europe. Based on the amalgam of ages derived for our clades, the ancient DNA and the corresponding archaeological cultures, we can be fairly confident our ancestors arrived in western Europe in the centuries surrounding 3000 BC, having come from what later became the European part of the Soviet Union. Deriving anything further than that becomes probabilistic in nature: we cannot be conclusive about what did or did not happen, but must make informed speculation based on the few surviving details we can see.

We know that our ancestors came from easternmost Europe. We don't know that they came directly from the Yamnaya. The Yamnaya ancient DNA belonged to individuals closely related to us, but who were not our direct ancestors. They split from us as the L23 level, around 4200 BC, and around 1000 years before the first Yamnaya ancient DNA in the Samara region of the middle Volga basin. The previous 1000 years is unaccounted for in ancient DNA studies in general, hence our ancestry is not tied necessarily to the Yamnaya themselves, but to one of the cultures from which they descend. This is overlooked in both the lay and academic communities. We can say with good confidence that our ancestors were not in Europe in 5200 BC, but the ancient DNA evidence becomes limited after 4240 BC.

My description of the European part of the former Soviet Union is intentionally specific. The *Urheimat* (ancient homeland) of R-M269 and later R-L23 is theorised to have come from a variety of areas in this region, including the Ural mountains, Caucasus and the Ukraine. The finding of R1b1 remains in Latvia shows that we have a lot to find out about the distribution of our ancestors before circa 3000 BC. The ancient DNA simply isn't sampled sufficiently well across the western former USSR to identify the homeland of M269 conclusively.

The transition to Europe presents different problems. We know that all of R-P311 is descended from one man, but one man alone cannot conquer Europe, so why was he so successful? We know he was not alone. There are now a number of European burials sequenced from the start of the 3rd millennium BC, which evidence the arrival of various R1a1 clades into western Europe at the same time (R1a-S224, R1a-Z281). The presence of other R-M269 basal clades in western Europe, some following the same distributions, suggests there were other, less-successful lines came with them. It was something about the later history of our family that made us more successful.

We can't say explicitly that our ancestors came to Europe with the Corded Ware culture. The Corded Ware culture burials we have are mostly R1a. Oblaczkowo, in western Poland, is the only early Corded Ware R1b1 burial. Unfortunately, the sample recovery was poor, so we don't have any further insight into his haplogroup. A contemporary burial in Lánycsók in southern Hungary is R1b (R-M343) without further differentiation. It isn't clear if either of these are our ancestors, so it isn't clear if our route into Europe was north of the Carpathians or along the Danube. We can only say that R1b appears strongly present in Europe before 2578 BC, hence that our ancestors are likely to have arrived before this date. We lack a full picture of the genetic upheaval at this time, as we do not have ancient DNA from many cultures which immediately preceded (or were contemporaneous with) the Corded Ware culture, including the Globular Amphorae and Baden cultures. We cannot rule out that our ancestors came to Europe with these cultures (although the Baden culture is nominally thought to be indigenous). The few later burials we have in near-eastern Europe come from the Vátya culture of Hungary, and are haplogroup I2a2a.

The Corded Ware culture interacted and merged with the Bell Beaker culture. The migration of this culture and (possibly) its people is complex and subject to interpretation. A leading interpretation today is that it was carried from Iberia to central Europe, dispersed along the major waterways, then undergone a backflow to the regions of its origin. The majority of the ancient R-M269 results are from the Bell Beaker culture, and all Bell Beaker culture ancient DNA so far appears to be R-M269. In some cases, this can be specifically haplotyped as R-P312, the implication being that adoption of the Bell Beaker culture by the R-P312 ancestor allowed the spread of R-P312 across western Europe. The earliest dated burial (between 2600 and 2500 BC) is at Kromsdorf, in central Germany, and is M269+ but U106-. The (unproven) implication is that it is P312+. The earliest confirmed P312+ burial is in the nearby Quedlinburg, dated between 2431 and 2150 BC. However, various other P312 Bell Beaker culture burials exist along the upper Danube valley, and would likely be dated between 2500 and 2300 BC. Some of these show P312+ and even U152+.

With the sole exception of RISE98 (2275 - 2032 BC, southern Sweden), there are no truly ancient U106+ DNA. Taken alone, this would suggest that U106 was slow to propagate, possibly explaining its smaller population compared to P312. However, this is at odds with the large number of direct sub-clades of U106 (current 12), indicating that U106 expanded quickly after it formed (alongside P312). The lack of early U106 could represent many things, sociological, cultural or statistical. However, the location of RISE98 and the lack of U106 in existing Bell Beaker burials likely indicates that the modern distribution of R-U106 is indicative of its origin in north-western Europe.

The relative timing of the arrival of our ancestors into Europe compared to the rate of SNP formation can still be debated. It is still not clear cut whether the large block of SNPs which include L11, L151, P310 and P311 can be split. However, it is clear that the last common ancestor of this clade represents the beginning of a successful, major population expansion. A logical conclusion is that this population expansion co-incides with the arrival of our ancestors in Europe. We cannot state this with absolute certainty, however studies of the invading cultures (Globular Amphorae and Corded Ware) show little discernable difference in the carbon dates derived from this period, indicating a very rapid expansion (within one or a few generations) of these cultures across Europe.

The last common ancestor of P311 could therefore have been born anywhere in the migration from eastern to north-western Europe. If he was born in eastern Europe, he did not leave any surviving family behind there: all the ~40 R-P311, R-P312 and R-U106 basal clades appear western or central European in origin, not eastern European. We know that the P311 ancestor did leave this large family, so there cannot be a long consanguinity of our U106 and P312 ancestors in eastern Europe. So either we are looking at one of two scenarios. Firstly, that the entire P311 family moved, wholesale, from east to west: this becomes increasingly unlikely as generations pass and the P311 population rapidly grows; or secondly, that P311 formed somewhere on the path to western Europe, and both U106 and P312 are western European.

With these limits, if P311 arrived in Europe with the Corded Ware culture, this basically dictates that the common P311 ancestor lived some time in the period 2900 - 2700 BC. If P311 arrived with the Globular Amphorae culture, the period moves to somewhere around 3100 - 2800 BC. Of course, these are only two out of a number of possibilities.