Indo-European demic diffusion model

3rd edition, revised and updated

Carlos Quiles
Indo-European demic diffusion model

3rd edition, revised and updated

An essay by

Carlos Quiles

LL.B., B.B.A., M.D., M.Sc., Ph.D. in Medicine

Department of Anatomy, Cell Biology, and Zoology

University of Extremadura

Badajoz

September 2017
Abstract

Previous archaeological and linguistic theories have based the expansion of some Indo-European proto-languages on the spread of the Corded Ware cultures, under the influence of the Yamna horizon.

Investigation of human ancestry of ancient and modern DNA samples have been used together with recent archaeological and linguistic data to obtain a more precise general picture of the evolution of Eurasian cultures, peoples, and languages related to Indo-European languages.

The Indo-European demic diffusion model proposed advances the theory that the expansion of Indo-European languages from the steppe was linked to the expansion of R1b1a1a2-M269 lineages Eurasia. A North-West Indo-European group most likely expanded directly with western Yamna migrants into the East Bell Beaker culture, and the Corded Ware culture was probably not associated with the expansion of Indo-European languages. That challenges previous archaeological and linguistic theories concerning the dialectal evolution of Late Proto-Indo-European.
Table of Contents

Abstract ........................................................................................................................................... 5

List of Figures .................................................................................................................................. 11

I. Introduction ................................................................................................................................... 13

II. Materials and Methods ............................................................................................................. 15

III. Results ...................................................................................................................................... 19

III.1. Palaeolithic ......................................................................................................................... 19

III.1.1. R1b-M343 and Western Hunter Gatherer ancestry ...................................................... 20

III.1.2. Indo-Uralic and Afroasiatic ............................................................................................ 21

III.2. Palaeolithic-Mesolithic transition ....................................................................................... 25

III.2.1. R1a-M420 and Eastern Hunter-Gatherer ancestry ...................................................... 26

III.2.2. Uralic and Yukaghir ........................................................................................................ 28

III.2.3. I2-M438 and Scandinavian Hunter-Gatherer ancestry ............................................... 28

III.3. Mesolithic-Neolithic transition ........................................................................................... 31

III.3.1. Caucasus Hunter-Gatherer ancestry and Indo-Hittite .................................................. 37

III.3.2. N1c1-M46 ..................................................................................................................... 38

III.4. Neolithic-Chalcolithic transition ......................................................................................... 40

III.4.1. Anatolian ......................................................................................................................... 44

III.4.1.1. Etruscan ..................................................................................................................... 46

III.5. Late Indo-European .............................................................................................................. 47

III.5.1. Contacts with the Caucasus ............................................................................................... 49

III.6. Forest Zone ........................................................................................................................... 52
III.6.1. Corded Ware culture ............................... 53

III.6.2. Corded Ware migration .............................. 57

III.7. Tocharian expansion ................................. 66

III.8. Late Indo-European expansion ..................... 68

III.8.1. Yamna migration .................................. 68

III.8.2. The expansion of Balkan Indo-European ........ 73

III.8.2.1. Greek ............................................. 74

III.8.2.2. Other Balkan languages ....................... 78

III.8.3. The expansion of North-West Indo-European in Europe ......................... 79

III.8.3.1. Bell Beaker culture .............................. 79

III.8.3.1.1. Contacts between Bell Beaker and Corded Ware ............. 81

III.8.3.1.2. Bell Beaker expansion ........................ 84

III.8.3.2. Únětice culture ................................ 90

III.8.3.3. Germanic ...................................... 94

III.8.3.4. Tumulus Culture ................................. 100

III.8.3.4.1. Celtic ........................................ 104

III.8.3.4.2. Italic .......................................... 108

III.8.3.4.3. Italo-Celto-Germanic and Celto-Germanic contacts ......... 111

III.9. Indo-European in Corded Ware societies ............. 112

III.9.1. Balto-Slavic ....................................... 112

III.9.2. Indo-Iranian ....................................... 118

III.9.2.1. Ancient North Indian and Ancient South Indian ancestry ........ 125

III.9.3. A common Corded Ware substrate .................. 127

III.9.3.1. Agricultural Substrate Hypothesis .................. 129

IV. Discussion .............................................. 133

IV.1. Consequences of the Indo-European demic diffusion model ............. 133

IV.2. Demic diffusion, cultural diffusion, or founder effect ....................... 135
List of Figures

Figure 1. Example of diagram of expansion and relationships of Indo-European languages, based solely on linguistic considerations. Adapted from Adrados (1998). .. 16

Figure 2. Diachronic map of Palaeolithic migrations. ................................................................. 23

Figure 3. Modified from Günther et al. (2017). Mesolithic samples and their genetic affinities. ................................................................................................................................. 30

Figure 4. Diachronic map of Neolithic expansion in Europe ca. 6500-5000. .......................... 32

Figure 5. Image modified from Mathieson et al. (2017). Structure and population change in European populations with farmer ancestry. .......................................................... 33

Figure 6. Diachronic map of Mesolithic migrations in Eurasia ca. 6500-5000. ..................... 36

Figure 7. Diachronic map of Early Neolithic migrations in Europe ca. 5000-4000 BC. .......................... 42

Figure 8. Modern distribution of R1b1a1a2-M269 (xL23) lineages. ................................. 45

Figure 9. Diachronic map of Eneolithic migrations in Europe ca. 4000-3100 BC. .... 50

Figure 10. Image modified from Mathieson et al. (2017). Structure and population change in European populations with mainly hunter gatherer ancestry. ......................... 59

Figure 11. Image modified from Monroy Kuhn, Jakobsson, and Günther (2017). Kin-relationship among males at the Corded Ware site in Esperstedt, Germany. .................. 61

Figure 12. Detail of PCA analysis of free datasets including Minoans and Mycenaens. ........................................................................................................................................ 62

Figure 13. Diachronic map of Copper Age migrations ca. 3100-2600 BC ................... 63

Figure 14. Diachronic map of Copper Age migrations in Asia ca. 3100-2600 BC .... 67

Figure 15. Modified from Mathieson et al (2017). Individuals projected onto axes defined by the principal components of 799 present-day West Eurasians. .................. 70

Figure 16. Modified from Mathieson et al (2017). Supervised ADMIXTURE plot. .... 71

Figure 17. Modern distribution of haplogroup R1b1a1a2a1-L51* ......................... 72

Figure 18. Modern distribution of haplogroup R1b1a1a2a2-Z2103. .......................... 74

Figure 19. PCA analysis of free datasets including Minoans and Mycenaens (Lazaridis et al. 2017), and Scythian and Sarmatian (Unterländer et al. 2017) samples. ............... 76

Figure 20. Diachronic map of migrations in Europe ca. 2600-2250 BC. .......... 83

Figure 21. Image from Olalde et al. (2017). Proportion of Steppe-related ancestry (shown in black) in Beaker Complex groups ................................................................. 85
Figure 22. Image from Olalde et al. (2017). Principal component analysis of 990 present-day West Eurasian individuals.

Figure 23. Modern distribution of haplogroup R1b1a1a2a1a2c1-L21

Figure 24. Modern distribution of haplogroup R1b1a1a2a1a2-DF27

Figure 25. Diachronic map of migrations in Europe ca. 2250-1750 BC.

Figure 26. Diachronic map of migrations in Europe ca. 1750-1250 BC.

Figure 27. Modern distribution of haplogroup R1b1a1a2a1a1-U106

Figure 28. Diachronic map of migrations in Europe ca. 1250-750 BC

Figure 29. Modern distribution of haplogroup R1b1a1a2a1a2b-U152

Figure 30. Diachronic map of migrations in Europe ca. 750-250 BC

Figure 31. Diachronic map of migrations in Europe ca. 250 BC – 250 AD

Figure 32. Modern distribution of haplogroup R1a1a1b1a-Z282

Figure 33. Diachronic map of migrations in Europe ca. 250-750 AD

Figure 34. Diachronic map of migrations in Asia ca. 2600-2250 BC

Figure 35. Diachronic map of migrations in Asia ca. 2250-1750 BC

Figure 36. Diachronic map of migrations in Asia ca. 1750-1250 BC

Figure 37. Diachronic map of migrations in Asia ca. 1250-750 BC

Figure 38. Diachronic map of migrations in Asia ca. 750-250 BC

Figure 39. Diachronic map of migrations in Asia ca. 250 BC - 250 AD

Figure 40. Diachronic map of migrations in Asia ca. 250-750 AD

Figure 41. Modern distribution of haplogroup R1a1a1b2-Z93

Figure 42. Diachronic map of migrations ca. 750 – 1300 AD

Figure 43. Modified file from recent papers on ancient samples from Eastern European, Southeastern European, Western European, and Bell Beaker cultures

Figure 44. PCA analysis of dataset including Minoans and Mycenaean.

Figure 45. Detail of image, modified from Mathieson et al. (2017). Unsupervised ADMIXTURE plot from k=3 to 11, on a dataset consisting of 799 present-day individuals and 405 ancient individuals.

Figure 46. Detail of unsupervised ADMIXTURE plot from k=3-6, 8, 10, and 12, on a dataset consisting of 318 ancient individuals.

Figure 47. Cross-validation error as a function of the number of ancestral components K
I. Introduction

Language and culture expansion is explained by two main alternative models: the demic diffusion model, which involves mass movement of people; and the cultural diffusion model, which refers to cultural impact between populations, and involves limited genetic exchange between them. Language transfer since ancient times seems to be associated with an expansion of people (Mikhailova 2015), which is demonstrated, in most cases, by a significant replacement of patrilineal Y-DNA. Investigation of Y-DNA haplogroups help demonstrate e.g. the expansion of Han people in Northern and Southern China (Wen et al. 2004; Zhao et al. 2015), and the expansion of Arabs in the Arab peninsula (Chiaroni et al. 2010), and into Southern Levant and North Africa (Nebel et al. 2002). Recently, the genetic history of Europe – including the expansion of hunter-gatherers and farmers – has been more precisely shaped thanks to ancient DNA research (Fu et al. 2016).

The recent expansion into Europe and Asia of Eurasian pastoralists, commonly identified with Indo-European speakers in mainstream diffusion models (Gimbutas 1993; Mallory 2014), was linked to haplogroup R1a (Semino 2000; Wells et al. 2001; Zerjal et al. 1999) due to the correlation of its modern geographic distribution with the ancient Corded Ware culture, and modern Balto-Slavic, Germanic, and Indo-Iranian speaking areas (Mirabal et al. 2009; Underhill et al. 2010).

Haplogroup R1b, which shows a modern western European distribution peaking in the British Isles and around historically Basque-speaking regions (Myres et al. 2011; Lucotte 2015), was until recently associated with a Palaeolithic western European origin (Morelli et al. 2010; Semino 2000). With decreased age estimates of haplogroup R1b in Europe, a more recent spread with farming has been suggested (Myres et al. 2011; Chiaroni, Underhill, and Cavalli-Sforza 2009; Cruciani et al. 2011; Balaresque et al. 2010).
Following these genetic frameworks, Indo-European languages would have spread with an Indo-European-speaking, R1a-dominated, invasive, eastern (Corded Ware culture) population into a non-Indo-European-speaking, R1b-dominated, western Atlantic (Bell Beaker culture) population. This connection was the weakest link between the supposed archaeological and the attested historical European linguistic landscapes, needing explanatory models that included some kind of cultural diffusion model – e.g. technologically- or economically-based (Brandt et al. 2015).

Ancient DNA (aDNA) investigation allows us to disentangle complex human history (Slatkin and Racimo 2016). The most recent research of ancient genetics (Haak et al. 2015; Allentoft et al. 2015; Mathieson et al. 2015), concerned with general population movements of Eurasians westwards from the steppe, has shown with their published data that haplogroup R1b was almost absent from Western Europe until after the expansion of Eurasian pastoralists. It has also shown that the origin of most of its modern descendants in western Europe is probably to be traced to the Pontic-Caspian steppes, and therefore that its expansion into central Europe happened at nearly the same time as haplogroup R1a, i.e. from the east and after ca. 3000 BC (Haak et al. 2015). In these studies, R1a was almost absent from samples of the Yamna horizon, most of which belonged to haplogroup R1b-M269.

The earliest linguistic link between haplogroups R1b and R1a, deemed until recently a cultural diffusion along the Corded Ware – Bell Beaker contact area (and later among Bell Beaker groups), seems thus to be contested by the latest genetic research. However, alternative explanations are being sought to adapt older paradigms to the newest research, suggesting a direct connection of the expansion of Indo-European languages to the Corded Ware culture (Allentoft et al. 2015), and thus R1a as the genetic marker of the expansion of Proto-Indo-European speakers in Europe (Horvath 2015).
II. Materials and Methods

The theory presented here offers an alternative population expansion model that seems to better fit the recent genetic research (involving ancient as well as modern DNA investigations) with mainstream archaeological and linguistic models.

Linguistic models of Indo-European (IE) dialectal differentiation based on comparative grammar and internal reconstruction (Figure 1) will be used to illustrate this theory, most of which currently follow the mainstream three-stage migration model (Meid 1975). The most common nomenclature of Early, Middle, and Late Proto-Indo-European periods is used (Dunkel 1997).

Such linguistic models consider Proto-Indo-European (PIE) as the product of a long historical development, formed gradually – like most natural languages –, and having thus stages of development (Lehmann 1992). This theory is therefore in contrast with the ‘constellation analogy’ (Clackson 2007, 2013) and similar negations of a concrete community of speakers – defined in time and space – of PIE or any of its later dialects. Historical linguistics can only provide a relative historical framework for individual Indo-European languages and proto-languages, though (Mallory and Adams 2007).

Archaeology works with the concept of culture, and as such it is able to determine timelines. When these timelines complement linguistics beautifully both are able to provide a contextualized historical explanation of linguistic frameworks (Vander Linden 2015; Hänsel and Zimmer 1994). The model set forth by Marija Gimbutas (1963), impressively corrected and expanded recently by Anthony (Anthony 2007; Anthony and Brown 2011; Anthony 2013), of potential cultures where Indo-European was spoken, is used in this paper as the basic framework for the potential expansion of Indo-European peoples.
Even though Anthony links his theory to a linguistic model developed by phylogenetics (Anthony and Ringe 2015; Ringe, Warnow, and Taylor 2002), it seems more reasonable to avoid such methods, due to their controversial nature and labile results (Pereltsvaig and Lewis 2015).

The theory laid in this paper takes dialectal evolution – lying at the core of any IE expansion model – as its stable framework, and uses genetic investigation (of ancient and modern DNA samples) and its potential relationship with archaeological cultures to establish an expansion model step by step. It also takes into account that there are complex problems found in correlations of languages with archaeological cultures (Meier-Brügger 2003) and human genetics (Campbell 2015).

Ancestry of any selected population is likely to be a mixture of several ancient groups, which is reflected on the genetic structure (Haak et al. 2010; Skoglund et al. 2012; Malmström et al. 2009; Lazaridis et al. 2014). However, the genetic landscape for ancient
populations is limited by the number of ancient DNA samples and ancient populations studied (Hellenthal et al. 2014). For simplicity purposes, results of published papers will be taken into account in this study, including admixture analyses and SNPs of ancient and modern Y-DNA samples, since mtDNA samples involve a more complex analysis in demic diffusion models – where the paternal lineage of the invaded territory is believed to be replaced or displaced to a certain extent.

Y-DNA haplogroups and subclades will also be referred to as *lineage*, whereas common admixture components defined in recent papers will be referred to as *ancestry*.

For the sake of consistency, [YFull](#) estimates for year formed and time to most recent common ancestor (TMRCA) of each subclade have been used unless other sources are expressly stated. Also for the sake of consistency, YBP dates have been approximated to BC.

Public data from [FTDNA](#)-associated groups R1b, R1b-U106, R1b-P312, R1b-DF27, and R1b-U152 were used in assessing haplogroup distribution in modern populations.

Modern physical maps are used to illustrate potential expansion routes of ancient cultures, peoples, and languages, even though they pose a significant danger to the development of a sound model, since they almost invariably involve “a concatenation of weakly supported links that corporately form an ‘arrow’ of dispersion” (Mallory 2014). Map routes are only depicted as a visual help to add movement to the otherwise stationary maps of ancient cultures, peoples, languages, and ancient DNA obtained from scattered burials.

---

1 Dates were obtained from the website during May 2017.
III. Results

III.1. Palaeolithic

A sample from the Upper Palaeolithic hunter-gatherer Mal’ta boy dated ca. 22350 BC shows that his paternal lineage diverged from haplogroup R-M207* shortly before its split into R1-M173 and R2-M479 subclades (Raghavan et al. 2014). His so-called Ancient North Eurasian (ANE) ancestry contributed substantially to the genetic ancestry of Siberians, Native Americans, and Bronze Age Yamna individuals (Lazaridis et al. 2014), being close to modern-day Native Americans, Kets, Mansi, Nganasans, and Yukaghirs (Flegontov et al. 2016).

Haplogroup R-M207 is itself descended from a common ancient lineage shared with the Ust’-Ishim man (ca. 43070 BC), probably belonging to the first wave of humans to migrate out of Africa into Eurasia (Fu et al. 2014).

Samples from the earliest modern humans (arrived ca. 43000 BC), who probably displaced Neanderthals (and maybe drove them to extinction), and are related to the Ust’-Ishim and Oase1 samples, seem not to have contributed substantially to the ancestry of modern Europeans.

Palaeolithic samples from ca. 35000 (Kostenki14) to ca 12000 BC (Villabruna) seem to have descended from a single founder population, do not share ancestry with the Mal’ta cluster, and form part of the ancestry of present-day Europeans. Four population turnovers (see Figure 2) are distinguished (Fu et al. 2016):

- First, the Goyet cluster appears associated with the Aurignacian cultural complex.
• Second, the Věstonice cluster is associated with the Gravettian cultural complex, which shows that its culture may have spread at least in part by population movements.
• Third, the El Mirón cluster shows the re-emergence of a deep branch of the Goyet cluster in Iberia, associated with the Magdalenian culture, potentially representing a post-Ice Age expansion from south-western European refugia.
• Fourth, the Villabruna cluster, which shows affinity to the Near East, contributes to the ancestry of Mesolithic hunter-gatherers of south-eastern Europe, coinciding with the Bølling-Allerød interstadial – the first significant warming period after the Ice Age – and the transition within the Epigravettian in southern Europe and the Magdalenian-to-Azilian transition in western Europe.

The Villabruna cluster may therefore reflect migrations or population shifts within Europe at the end of the Ice Age, consistent with the replacement of mitochondrial DNA sequences found. This may be explained by a population expansion from south-eastern European or west Asian refugia after the Ice Age.

Within the Villabruna cluster, some individuals have affinity to East Asians, which is not driven by Basal Eurasian ancestry.

III.1.1. R1b-M343 and Western Hunter Gatherer ancestry
Haplogroup R1b-M343 was deemed to have originated ca. 16400 BC in Western Asia (Karafet et al. 2008), and it has been proposed that these lineages survived the Last Glacial Maximum in refugia near the southern Ural Mountains and the Aegean Sea (Lobov 2009). It has been proposed that its latest westward migration happened during the Late Neolithic (Myres et al. 2011).

However, an Epigravettian individual of haplogroup R1b1a-L754, dated ca. 12030 BC, was found in Villabruna (Fu et al. 2016). Based on the most recent data of modern populations, an origin of R1b1a-L754 ca. 16900 BC is suggested, with a time to MRCA ca. 15100 BC. Another old European sample classified as R1b-M343 comes from a western hunter-gatherer in Ioussierres.

Hunter-gatherers from the Iron Gates prove the regional continuity of haplogroup R1b1a-L754 (xR1b1a1a-P297, xR1b1a1a2-M269). These samples were probably from branches that have not survived in modern populations, and they cover an extensive period spanning from the first half of the 10th millennium to the first half of the 6th millennium.
BC, with the latest samples showing already Middle East farmer ancestry (Mathieson et al. 2017; González-Fortes et al. 2017).

More samples possibly related to these ancient branches are found later in Ukraine, Iberia (see below), and central European Neolithic in Quedlinburg as R1b1a-L754 (xR1b1a1a2-M269) ca. 3590 BC (Haak et al. 2015). These samples, coupled with R-M207 samples found in Ganj Dareh (Iranian Neolithic) in the first half of the 9th millennium might suggest a southern Eurasian migration route for R1b1-L278 lineages, through the Iranian plateau.

The samples of basal R1b-M343* lineage in modern populations of southern Kazakhstan (Myres et al. 2011) and Iran (Grugni et al. 2012) give further support to the southern migration route into Europe. Basal R1b1-L278* lineage was found in five cases out of 5,326 cases studied – 3 Italians, 1 West Asian, 1 East Asian (Cruciani et al. 2010) –, which also point to a potential ancestral migration into Europe.

Mesolithic R1b-M343 lineages from Villabruna (Italy ca. 12000 BC) to the latest south-eastern European hunter-gatherers cluster closely together. Their ancestry is defined as of Western Hunter-Gatherers (WHG), which includes hunter-gatherer individuals from Bichon (Switzerland ca. 11700 BC), Loschbour (Luxembourg ca. 6100 BC), as well as samples from La Braña (Iberia ca. 5865 BC), and Koros (Hungary ca. 5710 BC). Ancient individuals from France, Sicily, Croatia, France, and Germany share this ancestry, which suggests that WHG was distributed widely in Europe for at least six thousand years, and probably expanded from a south-eastern European refugium following the last Ice Age ca. 13000 BC (Mathieson et al. 2017).

III.1.2. Indo-Uralic and Afroasiatic

Indo-European has been described as “a branch of Indo-Uralic which was transformed under the influence of a Caucasian substratum” (Kortlandt 2002), which would imply an invasion of Indo-Uralic-speaking peoples to a territory of previous Caucasian hunter-gatherers. Such Caucasian influence has been supported recently by the finding of a genetic contribution (probably during the Mesolithic-Neolithic transition) of a pocket of Caucasus hunter-gatherers, who seem to have weathered much of the last Ice Age in apparent isolation (Jones et al. 2015).

Long-ranging language relationships are difficult to prove. If Uralic and Indo-European shared a common ancestor – Indo-Uralic (Kloekhorst 2008) –, it should be associated to
the post-Swiderian east European communities with a majority of WHG ancestry and R1b1a1a-P297 lineages, similar to the Balkan hunter-gatherers from the Iron Gates. If these people and their language expanded from central and south-east European communities of the Villabruna cluster with a majority of R1b1a-L754 lineages, certain hypothetic linguistic communities can be proposed:

It would be conceivable but controversial (Prósper 2013), for example, to give credit to the nature of Proto-Basque as of Pre-Indo-European substratum (Forni 2013; Blevins 2015) – beyond pre- and post-Roman IE superstrata (Koch 2013). This is supported by the presence of an Iberian Eneolithic sample of R1b1a-L754 (xR1b1a1a2-M269, V88-equivalent) at Els Trocs ca. 5180 BC (Haak et al. 2015), and its potential continuity in north Iberia at least until the Bell Beaker expansion (Mathieson et al. 2017).

Proto-Afroasiatic has been proposed to have emerged in the southern fringe of the Sahara in an “upside-down” view (Bender 2007), while R1b1a2-V88 lineages (and specifically its subclade R1b1a2b1b1a-V69) have been found in north and central Africa, mainly in Chadic-speaking populations, but also in west Egypt and in the Middle East (Cruciani et al. 2010).

That lineage, probably related to the Villabruna cluster, could have crossed the Mediterranean into northern Africa quite easily before the end of the Ice Age, possibly through the southern Italian Peninsula. The DNA coming from outside of Africa, related to Eurasian herders (Schlebusch et al. 2017) is potentially linked to this migration, since a good proxy for this ancestry (before the recent study of ancient Levantine ancestry) were present-day Sardinians (Pickrell et al. 2014). From northern Africa they could have travelled south and then east, since the Sahara was an important site of occupation and crossing of hominids during the Holocene, with Fezzan-Chad-Chotts, and Chad-Chotts-Ahnet-Moyer megalake green corridors connecting northern and central Africa – with gradual desiccation of the desert, until ca. 4000 BC (Drake et al. 2011).

This old environment could have allowed the for a west-east migration, and for a sizeable population expansion in central Saharan territory. This model would agree with Chadic languages being the most divergent of the Afroasiatic group, excluding Omotic – whose population has been shown to be mainly of sub-Saharan ancestry, in contrast with other Afroasiatic peoples (Baker, Rotimi, and Shriner 2017).
Haplogroup R2a-M124 seems to be prevalent among (ancient and modern) Dravidians, and is also found in the Caucasus, while haplogroup Q-M242 has links to Asian and Native American populations.

There are thus potential links of linguistic macro-groups to the expansion of certain lineages: one could thus make a simplistic association of Indo-Uralic (and Paleo-Siberian) with R1a-M420 lineages, Dravidic (and potentially Kartvelian and Altaic) with R2-M479 lineages, and Afroasiatic with R1b-M343 lineages, all departing from an older Nostratic language (Bomhard 2008) associated then with R-M207.

However, macro-languages are speculative, and their relationships highly controversial, with such ancient archaeological evolutions – and their relationship to population movements – quite difficult to ascertain. A still more speculative relation with an older Borean macro-family (Gell-Mann, Peiros, and Starostin 2009), for example, could be explained by certain expansions of P1-M45 lineages, which should in turn help determine dialectal evolutions.
III.2. Palaeolithic-Mesolithic transition

The end of the last Ice Age ca. 14000-12000 BC brought instability to the Pontic-Caspian zone: meltwater flew torrentially from the northern glaciers and the permafrost into the Khvalynian Sea (the Caspian Sea is a small remaining part of it). A shoreline between the middle Volga and the Ural River restricted east-west movements south of the Ural Mountains (Anthony 2007).

By 11000-9000 BC water may have poured into the Black Sea (Major et al. 2006; Ryan 2007), enlarging it and creating the Sea of Azov. Although the magnitude and rapidity of this flow remains controversial (Yanko-Hombach, Gilbert, and Dolukhanov 2007), it is agreed that meltwater created unstable shores in north-eastern Europe (Patton et al. 2017).

Deglaciation and palaeoclimatic changes were probably more important in their potential for environmental, cultural, social and historical changes of this region, though. A significant deterioration is found during the Younger Dryas – Pre-Boreal period caused by climate aridization and reduction of overall biomass density in the region, with large group segmentation, local population dispersion, increase in population mobility, and decrease in population density (Smyntina 2016).

In the Boreal period, the Pontic-Caspian steppe became stable with an increase in climatic humidity, and a growth of biomass density. Hunters – probably from eastern and western regions – settled there and population density increased. The different migration times and paths of R1a-M420 and R1b-M343 lineages puts the Ural-Caspian frontier as a linguistic and cultural barrier that might have been crossed during this period.

R1b1a1a-P297 formation (ca. 14800 BC) and TMRCA (ca. 11300 BC), and the formation of R1b1a1a2-M269 (ca. 11300 BC) compared to its quite late TMRCA (ca. 4300 BC) point to a slow spread of hunter-gatherer groups of R1b1a1a-P297 lineages in east Europe coinciding with the geographic changes associated with the last deglaciation.

To the north, the population of the final Palaeolithic Swiderian culture, which developed in Poland on the sand dunes left behind by retreating glaciers, migrated during the Palaeolithic-Mesolithic transition (ca. 9500 BC) to the north-east following the retreating tundra, which is evidenced by a 300-year-long settlement break before a new population arrived (Kobusiewicz 2002). Post-Swiderian cultures developed in the Baltic and in the Forest Zone north of the unstable Pontic-Caspian zone, and it seems reasonable to assume an eastern and south-eastern migration of hunters into the now open areas.
Old traditions from the northern Black Sea region that survived the Younger Dryas migrated into new areas: the Anetivka tradition, from the west, expanded to the north of the Dniester, to the east in the Dnieper, and to the south penetrating the Crimean Peninsula; and the Grebenniki culture (derived from the Tsarinka flint knapping tradition) expanded west from the Lower Volga, sharing the same areas as Anetivka settlements. To the east, new traditions appeared (Smyntina 2016).

III.2.1. R1a-M420 and Eastern Hunter-Gatherer ancestry

Haplogroup R1a-M420 originated ca. 25000 BC, and was proposed to diverge initially in the vicinity of present-day Iran based on a study of modern populations (Underhill et al. 2015).

Eastern Hunter-Gatherer (EHG) ancestry is represented by two individuals from Karelia – one of haplogroup R1a1a1-M417 (ca. 6425 BC) and the other of haplogroup J (ca. 5250 BC) – and one individual from Samara (ca. 5600 BC), of haplogroup R1b1a1a-P297 (Mathieson et al. 2017). It has more ANE ancestry than any other ancient or modern population (Haak et al. 2015), being close to the sample from Afontova Gora (ca. 15980 BC), to the west of Lake Baikal.

The oldest aDNA sample of R1a-M420 lineage found in east Europe was at Vasylivka, dated ca. 8690 BC, at the same site of a later sample of haplogroup R1b1a2-V88, dated ca. 7250 BC (Mathieson et al. 2017). These and other three Mesolithic samples from Ukraine show an intermediate situation between EHG and Scandinavian Hunter-Gatherer ancestry. Later, during the transition of the Mesolithic to Neolithic, a decrease in ANE ancestry and an increase in WHG ancestry is observed.

We have seen that the Villabruna cluster (and associated Mesolithic R1b-M343 lineages found from western to eastern Europe) might be the representatives of the expansion of West Hunter-Gatherer ancestry, displacing or admixing with the existing population of western Europe. On the other hand, hunter-gatherers from the Iron Gates show WHG (87%) and EHG (13%), and mtDNA haplogroups that contrast with the prevalent haplogroups U5 or U2 of WHG, EHG, and Scandinavian hunter-gatherers (Mathieson et al. 2017).

R1b-M343 lineages may have thus brought WHG ancestry to the north Pontic steppe during the Mesolithic, either from south-eastern Europe or from the Swiderian culture (see above). Supporting this eastern migration, samples of R1b1a1a-P297 (xR1b1a1a2-
M269) lineage have been found in Latvian hunter-gatherers continuously in different periods, dated from the end of the 9th millennium BC to the end of the 4th millennium BC, including Kunda and Narva cultures (Jones et al. 2017; Mathieson et al. 2017).

These Baltic samples show an intermediate ancestry between western and eastern hunter-gatherers, of ca. 70% WHG and 30% EHG, before an increase in EHG is seen during the Neolithic Comb Ware culture expansion (see below Forest Zone).

The sample from Afontova Gora 3 (ca. 15980 BC) of the Late Palaeolithic Siberian complex (with cultural and genetic links to the people from Mal’ta Buret’), shows a common cluster in PCA analysis with samples from the eastern steppe of the Mesolithic, Neolithic, and Chalcolithic (Mathieson et al. 2017).\(^2\)

Haplogroup R1a1a1-M417 (formed ca. 6500 BC, TMRCA ca. 3500 BC) is first found in the Karelian hunter-gatherer dated ca. 6425 BC. Two samples of R1a1a1-M417, dated ca. 5250 BC, have also been found in Early Neolithic Baikalic cultures near the zone where the ancient Mal’ta-Buret’ culture was located. Strontium isotope ratios confirm their local origin, with high paternal heterogeneity, with a trend to replacement by east Eurasian lineages during the Late Neolithic (Moussa et al. 2016).

A migration of R1a-M420 lineages from the Iranian area to the forests of Eastern Europe has been previously proposed (Horvath 2015).

Given the Eurasian origin of the eastern European pottery (see below Mesolithic-Neolithic transition) and its westward expansion into Europe, the likely eastern origin of EHG ancestry and R1a1a1-M417 lineages, it seems logical to find a common origin of both populations (from eastern Europe and the Baikal region) in an expansion from Eurasian territory, dated around the subclade’s formation date.

The traditional association of Forest Zone hunter-gatherers’ expansion with hunters of the Kelteminar culture, would imply a date ca. 5500 BC, which is too late for the attested samples. Ancestors of this population are supposed to have originally migrated from the Hissar range ca. 6000 BC, though, with an earlier expansion from a neighbouring area potentially fitting the available data.

---

\(^2\) See Extended Data Figure 1, p. 23.
Because of the early sample of R1a-M420 found in the Mesolithic north Pontic area, and maybe also the rare subclade R1a5-Z645 found in Estonia (Saag et al. 2017), it could be proposed that the migration of R1a-M420 subclades (including R1a1a1-M417) happened from the north-west Pontic area, with a back-migration of these lineages to the Baikal region. However, given the available archaeological data, it does not seem reasonable today to propose that only pottery was adopted from Eurasia, while population was exported.

III.2.2. Uralic and Yukaghir

The expansion of R1a1a1-M417 lineages bringing EHG ancestry from central Eurasia may have disrupted the Early Proto-Indo-European R1b1a1a-P297 community thriving in east Europe.

In this context, if the idea of an Indo-Uralic group is rejected, R1a1a1-M417 lineages might have spoken Uralo-Yukaghir languages when arriving in the Forest Zone from the east, and Uralic could therefore be considered a superstratum over a Pre-Indo-European substratum.

A Uralo-Yukaghir community spread over Eurasia is supported by the east-west direction of cultural innovations in the region, and by the finding of maximum Ancient North Eurasian ancestry in modern-day Kets, Mansi, Native Americans, Nganasans and Yukaghirs (Flegontov et al. 2016).

On the other hand, R1a1a1-M417 lineages may have brought a Yukaghir superstratum to the Indo-Uralic (an equivalent to Early Proto-Indo-European in this macro-family proposal) spoken in the Forest Zone by R1b1a1a-P297 communities similar to the Narva samples, developing a Proto-Uralic-speaking community.

Following the hypothesis of R1b1a2-V88 lineages bringing Afroasiatic to Africa, and macro-family proposals including Nostratic associated with haplogroup R-M207 (see above), it would be possible to propose an association of Eurasian R1a-M420 lineages and EHG with Indo-Uralic, which would have then been adopted by western European R1b1a1a-P297 communities.

III.2.3. I2-M438 and Scandinavian Hunter-Gatherer ancestry

Haplogroup I2-M438 formed ca. 25500 BC, and the modern European population has a TMRCA ca. 19900 BC. Individuals from Mesolithic Scandinavia, which define the so-
called Scandinavian Hunter-Gatherer (SHG) ancestry, show exclusively I2-M438 lineages. In samples of Mesolithic SHG, together with those of WHG and EHG ancestry (but for those of the Balkans), mtDNA haplogroups U5 and U2 are prevalent (Mathieson et al. 2017).

The ice sheet retracted from northern Europe allowed for the colonisation of the Scandinavian Peninsula from about 9700 BC, according to the archaeological record, both in southern and northern Scandinavia, while ice still dominated the interior.

The peninsula seems to have been colonised first from the south by peoples from central Europe, related to late-glacial lithic technology (direct blade percussion technique), which brought WHG ancestry with them.

Genomic studies show an invasion from the north-east by peoples from east Europe, who brought EHG ancestry. Their technology is probably represented in the ‘pressure blade’ technique found in northern parts of Scandinavia.

This left a paradoxical pattern of increased EHG ancestry in north and west Scandinavia, and WHG ancestry in east and central Scandinavia, which correlates with Baltic samples (Figure 3). It has also been shown that selection drove the unique combination of light skin and hair and varied blue to light-brown eye colour, as part of the adaptation to a different environment, in contrast with WHG who had the specific combination of blue eyes and dark skin (Günther et al. 2017).

Scandinavian Hunter-Gatherer (SHG) ancestry, therefore, represents a mixed group tracing parts of their ancestry to both WHG and EHG.

Many ancient DNA samples of I2-M438 haplogroups are found since the Palaeolithic, and two main branches seem to have divided early: I2a1b2-L621 lineages are found mainly in the Balkans, and I2a2a-M223 – distributed through central Europe – seems to have followed the expansion of Italo-Celtic and Germanic, and were therefore possibly integrated with R1b1a1a2a1a-L151 lineages since the Bell Beaker complex.
Figure 3. Modified from Günther et al. (2017). “Mesolithic samples and their genetic affinities – (A) Map of the Mesolithic European samples used in [https://doi.org/10.1101/164400 Günther et al. (2017)]. The pie charts show the model-based estimates of genetic ancestry for each SHG individual. The map also displays the ice sheet covering Scandinavia 10,000 BP (most credible (solid line) and maximum extend (dashed line)). Newly sequenced sites are shown in bold and italics. Additional European EHG and WHG individuals used in this study derive from sites outside this map. (B) Magnified section of genetic similarity among ancient and modern-day individuals using PCA featuring only the Mesolithic European samples”. Original under a CC-BY-NC 4.0 International license.
III.3. Mesolithic-Neolithic transition

Hunters from the Pontic-Caspian steppe – as European Mesolithic hunter-gatherers in general – possessed no domesticated animals before the arrival of stockbreeding, spread with Neolithic farmers from Anatolia after about 6100 BC, probably through different colonization routes through the Aegean, that involved diverse Neolithisation processes (Horejs et al. 2015).

While economic practices changed, some cultural traits like funerary practices did not accompany the ‘Neolithic package’ acquired in the Balkans by local fisher-hunter-gatherers. It seems that farmers and their domestic animals spread fast, in ca. 10 human generations, from sub-Mediterranean Macedonia to the northern limits of the temperate Balkan Peninsula and the adjacent Carpathian Plain, which may have put serious difficulties for the spread of cattle until selective pressure could drive genetically-driven adaptations to harsh environments (Ethier et al. 2017).

It has been supported with human ancestry studies that Middle East farmers arrived into central and western Europe with the Neolithic expansion (Brandt et al. 2013; Olalde et al. 2015; Szecsenyi-Nagy et al. 2017).

Ancient Middle Easterners show basal Eurasian ancestry with significantly less Neanderthal inheritance than East Asians, which suggests an affinity between Natufians and populations of north or sub-Saharan Africa. This is supported by the Y-DNA haplogroup E found in Levantine Neolithic populations (Lazaridis et al. 2016).

Together with WHG and EHG, samples from Neolithic Levant and those from the Neolithic of western Iran form the four streams of ancestry seen in the Middle East. There is continuity of Natufians (Levantine hunter-gatherers) with Levantine farmers, and Caucasus hunter-gatherers with farmers from the Zagros Mountains (east Iran).

While a population related to north-western Anatolian Neolithic farmers spread westward into Europe, farmers related to those of the Levant spread southward into north-east Africa (Lazaridis et al. 2016; Schuenemann et al. 2017).

Neolithic farmers arrived in Europe from a single Balkan population in two routes (see Figure 4), one to the Linearbandkeramik (LBK) culture from central Europe, and another to the Impressa complex of Croatia and Epicardial Early Neolithic from Spain (Mathieson et al. 2017). However, there are also traces of an outgroup related to the north-western
Anatolian Neolithic population in the Peloponnese Neolithic, maybe linked to the pre-pottery Neolithic of Cyprus and the Levant.

While farmers and hunter-gatherers lived in settlements in close proximity during the Neolithic, (in the Balkans, in western, central and northern Europe), there are signs of long periods with minimal admixture (Mathieson et al. 2017).
Figure 5. Image modified from Mathieson et al. (2017). “Structure and population change in European populations with farmer ancestry. A: each population is modeled as a mixture of Anatolia Neolithic, WHG, and EHG. Dashed lines show temporal relationships between populations from the same geographic areas with similar ancestries. B: Z-scores for the difference in hunter-gatherer ancestry on the autosomes compared to the X chromosome when populations are modeled as a mixture of Anatolia Neolithic and WHG. A positive score indicates that there is more hunter-gatherer ancestry on the autosomes and therefore the hunter-gatherer ancestry is male-biased”. Original under a CC-BY-NC 4.0 International license.
During the Middle Neolithic, a resurgence of male-biased hunter-gatherer ancestry is seen in central Europe and Iberia (Figure 5), while persistent frontiers between hunter-gatherers and farmers are found in central and northern Europe, coincident with the loess belt of the northern European plain, to the north of which early farming techniques were probably not suitable. It is likely that new climates and environments led to the eventual breakdown of demic diffusion, and the spread of Neolithic traits by cultural diffusion (González-Fortes et al. 2017).

That resurgence of hunter-gatherer ancestry, with a ca. 4:1 WHG:EHG contribution, is found in the Balkan Neolithic in the territory of present-day Bulgaria, close to the Danube river. This suggests a heterogeneous landscape of farmer populations with different proportions of hunter-gatherer ancestry during the early Neolithic, probably due to pockets of hunter-gatherers surviving close to the coast and major rivers (Mathieson et al. 2017).

Before the arrival of farmers to the western frontier of the Pontic-Caspian steppe, pottery was produced in the first half of the 7th millennium BC by hunter-gatherer groups first in the Volga steppes (with the earliest pottery found to date in the Elshanian culture). This culture was probably derived from the Eastern Asian tradition of the Late Pleistocene through Siberia and the Transurals (Piezonka 2015).

The first Neolithisation of the Lower Volga region (see Figure 6), with the oldest pottery of ca. 6200 BC, can be attributed to the influence of the Kairshak culture in the northern Caspian region, where the first sites with the oldest pottery appeared ca. 6500 BC (Vybornov 2016). From the north-western Caspian region pottery spread south- and westward into north Pontic societies ca. 6200-6000 BC (Zaitseva et al. 2009).

Sparsely decorated pottery dispersed north into the Forest Zone ca. 6000 BC or slightly earlier, from the upper Volga and Dvina-Lovat’ regions to the east (into the Dvina-Pechora region) and west (into the eastern Baltic), reaching the Upper Volga, Serteya, and Valday cultures, and later the Narva culture.

Contacts of north Pontic cultures with Criş settlers from the Starčevo–Kőrösi–Criş culture about 5800 BC introduced domesticated cattle to the Bug-Dniester culture, but no signs of cultural assimilation has been found, with the later invasion of Linear Pottery sites ca. 5500-5200 BC respecting a similar cultural frontier, geographically coincident with the Dniester (Anthony 2007). Hence the language of western Neolithic settlers – assumed to
come from the Middle East, if language accompanied the spread of Neolithic technology – was probably not transferred to north Pontic herders.

A second expansion of eastern pottery reached the eastern Baltic region ca. 5500 BC, expanding from the Dnieper region to the north-west, generating the sparsely decorated Dubičiai pottery (later evolving into the Neman culture), and influencing the north European regions from the Narva to the Ertebølle cultures (Piezonka 2015).
From the Bug-Dniester culture domesticated cattle, sheep, and goats spread quickly from about 5200 BC east- and northward into Pontic-Caspian sites, reaching Khvalynsk and the Samara region about 5100 BC.

A third expansion of eastern pottery spread from the Volga-Kama region to the east ca. 5000 BC, connected to influences from beyond the Urals, showing a more elaborately decorated ware (with bands of pits and impressions made from comb stamps), spreading north and west in the Sperrings and Säräisniemi 1 cultures (Piezonka 2015).

A sample of R1b1a1a-P297 reported as possibly an intermediate stage of its formation (positive and negative markers in the M478 node) was found in a Mesolithic hunter-gatherer at Lebyanzhinka in the Samara region, dated ca. 5600 BC (Mathieson et al. 2015). Later samples from the same region show continuity of R1b1a1a2-M269 lineages, which seem to have expanded from east to west in the Pontic-Caspian steppes.

In the north Pontic steppe – apart from the earlier R1a-M420 and R1b-M343 samples from Mesolithic Vasylivka (see above) – there are four samples (dated ca. 6500-4000 BC) of the Mariupol culture from Volniensky, and one sample of the Azov-Dnieper culture from Vovnihi (ca. 5400 BC), of I-M170 (and one IJ) lineage.

Samples attributed to the early Sredni Stog culture in Deriivka (dated ca. 5500-4800 BC) include nine of R1b-M343 lineage, probably from an extinct branch of R1b1a-L754...
Indo-European demic diffusion model

(xR1b1a1a-P297, xR1b1a1a2-M269); one of R1a-M420 lineage; and four samples of haplogroup I-M170, probably I2a2a1b-L701 (Mathieson et al. 2017).

This diversity of lineages points to a mix in the different groups that emerged during the Mesolithic and early Neolithic periods, before the mass expansions that occurred later.

III.3.1. Caucasus Hunter-Gatherer ancestry and Indo-Hittite

Caucasus hunter-gatherer (CHG) ancestry is defined by a Late Palaeolithic individual from Satsurblia cave (ca. 11000 BC), and a Mesolithic individual from Kotias Klde cave, in western Georgia (ca. 6000 BC) – both of haplogroup J-M304 (the Kotias sample, and possibly both, from J2-M172).

Caucasus hunter-gatherers seem to have weathered much of the last Ice Age in apparent isolation, with the individual from Satsurblia showing also signs of recent consanguinity. That isolation has continued partially into the modern population of the southern Caucasus, in terms of ancestry, as well as Y-DNA and mtDNA haplogroups (Jones et al. 2015).

CHG ancestry was believed until recently to have contributed only late to the population of the Pontic-Caspian steppe, coinciding with the formation of the Yamna culture, since Samara hunter-gatherers showed only EHG ancestry and no CHG ancestry, while Yamna samples had up to 43% of CHG ancestry (Jones et al. 2015; Lazaridis et al. 2016). However, the so-called ‘Yamna’ or ‘steppe’ ancestry has been found in individuals from the Balkans at Varna I (ca. 4630 BC), Smyadovo (ca. 4500 BC), and Trypillian culture (ca. 3780 BC), some 2,000-1,000 years before the main Chalcolithic expansions associated with the steppes (Mathieson et al. 2017).

This suggests a more likely gradual contribution from intermittent contacts with cultures from the Caucasus during the Mesolithic-Neolithic and Neolithic-Chalcolithic transitions, joint with the westward expansions (and probably inner west-east movements) of Middle Indo-European speakers in the Pontic-Caspian steppe and beyond its natural frontiers to the north.

This CHG ancestry might be involved in what is believed to be the Caucasian substratum of Indo-European (Bomhard 2017), potentially differentiating it from Uralic within a hypothetic ancient Indo-Uralic group (Kortlandt 2002), which could then be tentatively identified with the older EHG ancestry.
In any case, a conservative view will be observed in this paper, assuming Proto-Uralic to be the language of hunter-gatherers in the Forest Zone, north of the Pontic Caspian steppe, at the same time as Middle Indo-European was spoken in the steppes in the transition to the Neolithic (Parpola 2012).

III.3.2. N1c1-M46
The arrival of N1a1-M46 (previously called N1c1) lineages into Northern Europe has been dated after 5000 BC (Ilumae et al. 2016), coinciding with the TMRCA of N1a1a1a-L708. However, the more recent formation (ca. 3800 BC) and TMRCA (ca. 2300 BC) of common European lineage N1a1a1a1a-L392 point to a late and stepped spread of these hunter-gatherer groups into the Forest Zone, that cannot be associated with the expansion of Pit-Comb Ware cultures.

Cultural assimilation remains therefore the best explanation for the shared Uralic languages of modern communities with a majority of R1a1a1-M417 and N1a1-M46 lineages. Eastern groups with N1a1a1a1a-L392 lineages may have brought with them the Altaic traits found in Uralic languages (Kortlandt 2010).

An aDNA sample of haplogroup N1a-F1206 is found in the Forest Zone dated ca. 2500 BC at Serteya (Chekunova et al. 2014). Nevertheless, it is tempting to place the mass migration of Siberian hunter-gatherer communities around the Urals later, with the expansion of the poorly understood Seima-Turbino phenomenon (which began ca. 2000 BC in East Asia), since it connected cultures from Mongolia to Finland.

Three samples of haplogroup NO (xO) found in the Middle Bronze Age Okunev culture, and two samples later in the Chermuchek culture area (Hollard et al. 2014) may give support to this assumption.

Modern Estonians have R1a-M417 lineages in more than a third of their population (Laitinen et al. 2002), similar to the proportion of haplogroup N1a1-M46 (Rosser et al. 2000). However, Finns show almost two thirds of haplogroup N1a1-M46 in the modern population, and only about 10% of R1a-M417.

Estonians are close genetic relatives to Finns, as well as to Baltic peoples and Russians from the Tver region (Nelis et al. 2010). Investigation of mtDNA in modern Finns has shown that there was probably a population decline ca. 1500 BC, and later an Iron Age bottleneck with a population peak ca. 500 AD (Översti et al. 2017).
There was probably, then, a long-term, gradual replacement of previously prevalent Y-DNA R1a-M417 subclades in the region, as supported by the increased ‘steppe’ ancestry in genome-wide ancestry of modern Finns. A sudden, strong population (and cultural) change associated with the arrival of N1a1-M46 lineages – like the ones seen with R1a-M417 (Corded Ware) and R1b-M269 (Yamna) in eastern Europe, has to be rejected. A founder effect of N1a1-M46 lineages is therefore the most likely explanation for their adoption of Finnic languages, a situation that is also supported by the genetic diversity of the Saami population (Tambets et al. 2004).
III.4. Neolithic-Chalcolithic transition

The Copper Age began in Bulgaria ca. 5200-5000 BC, and Old European copper-trade network included the Pontic-Caspian steppe societies after ca. 4600 BC.

In the north Caspian steppe, Late Neolithic and the Caspian Sea region culture had coexisted during the Eneolithic in the mid-6th millennium, and in the Lower Volga a change is noticed ca. 5000-4800 BC among the carriers of the North Caspian culture (Vybornov et al. 2016). The early Khvalynsk culture may have been an autochthonous culture based on the previous North Caspian culture, or its genesis could be the result of a migration of tribes from the southern region of the Trans-Caspian area (Vybornov 2016). Both these possibilities may account for the introduction of CHG ancestry in the population of the steppe.

During the 5th millennium, a strong, long-lasting, east-west oriented exchange network can be observed in the north Pontic area between the Cucuteni-Trypillian culture in the forest-steppe and Skelja (including the site at Deriivka) in the coastal steppe (Reingruber and Rassamakin 2016).

The revolution of herding, travel, and raiding – and thus the change in the steppe – came with horseback riding, appearing ca. 4800 BC in early Khvalynsk, and spreading south- and eastward. The early Sredni Stog culture began about 4400 BC, probably related to the influence of people from the east Pontic Caspian steppe (Anthony 2007).

Within this new culture, a new elite group associated with the Suvorovo-Novodanilovka complex (Anthony 2007), probably from the eastern steppe (and thus probably speaking Middle Proto-Indo-European), was involved in raiding and trading with the lower Danube valley during the Trypillian B1 period, before and during the collapse of Old Europe.

Settlements of Suvorovo-Novodanilovka chiefs have been found along the lower Danube (Figure 7), and sites of the posterior Cernavodă I culture (ca. 3600 BC) seem to represent the assimilation of migrants from the steppes, therefore linked to Anthony’s first expansion from the Pontic-Caspian steppes into the Balkans ca. 4200-4000 BC (Anthony 2007, 2013).

Three samples dated ca. 5150 BC are found in early Khvalynsk, one of R1b1a-L754 (probably M269), one of R1a1-M459, and one of Q1a-F903 lineage. While the R1b1a-L754 sample was reported as from a high-status burial, similar to high-status individuals
buried under kurgans in later Yamna graves, and therefore founder of an elite group of patrilineally-related families, the R1a1-M459 individual shows scarce decoration and his lineage is not found in later high-status Yamna graves (Mathieson 2015, Supplementary materials). A sample attributed to the Dnieper-Donets culture (predating Sredni Stog in the same region), dated to a similar time (ca. 4380 BC) belong to haplogroup R1a-M420 (Jones et al. 2017), which points – together with the diversity found in the Khvalynsk II cemetery – to a time preceding or coinciding with the successful expansion of R1b1a1a2-M269 lineages.

Steppe-related ancestry (defined by posterior samples from east Yamna) are found in an individual from Varna I (ca. 4630 BC) and in both samples from Smyadovo, in contrast with the resurgence of WHG ancestry in central Europe and Iberia. Also, two individuals of haplogroups R-M207 and R1b1a-L754 (dated ca. 4500 BC) are found in Smyadovo, and one of haplogroup R1-M173 (dated ca. 4460 BC) in Varna I cemetery (Mathieson et al. 2017).

These samples are not proven to correspond to R1b1a1a2-M269 lineages, and dates are slightly earlier for the mass migration proposed by Anthony (2007). Nevertheless, the region shows a discontinuity in R1b1a1a-P297 lineages in the Balkans after the arrival of Middle East Neolithic farmers (of G2-P15 lineages), which represent around half of more than 30 Y-DNA samples in the period from 6000 BC to 4500 BC (Mathieson et al. 2017). The fact that haplogroup R-M207 is not found in later Balkan samples either (until the second Yamna expansion) is also significant, potentially pointing to a transitory presence of this haplogroup from the steppe. Also, steppe imports are already found in Gumelnita, in the Lower Danube region, from about 4400 BC (Reingruber and Rassamakin 2016).

The older origin of haplogroup R1b1a1a2-M269 (ca. 11300 BC) compared to a later TMRCA (ca. 4300 BC) for the subclades survived in the modern population, coinciding with the successful spread of basal R1b1a1a2a-L23* (formed ca. 4300 BC, TMRCA ca. 4200 BC), point to an expansion that occurred around this time period. This population expansion came probably from some eastern clans of Pontic-Caspian herders that shaped the Sredni Stog culture in the west, and turned into Suvorovo-Novodanilovka chiefs and south-eastern European settlers.

Both lineages are found in the Balkans, Central Europe, and Armenia (Myres et al. 2011; Herrera et al. 2012), and their expansion is therefore to be associated with the split of

III.4.1. Anatolian

Anatolian has long been considered the first to branch out of Proto-Indo-European, due to its peculiar archaisms (Trager and Smith 1950), even before the proposal of a Late Indo-European community from which all other known IE languages branched out (Meid 1975; Kortlandt 1990; Lehmann 1992; Dunkel 1997; Melchert 1998; Adrados 1998; Ringe 2006; Mallory and Adams 2007; Beekes 2011).

In the Kurgan model, Anatolian was originally associated with the expansion of the Kurgan culture of the lower Volga and Kazakhstan into the Transcaucasian Copper Age culture ca. 2400-2300 BC (Gimbutas 1963). With time, the Maykop culture has been demonstrated to be much older than previously thought, coinciding with the Uruk expansion in Mesopotamia after about 3700 BC, and especially ca. 3350 BC (Anthony 2007).

Even this new chronology does not fit well with the much older guesstimates attributed to the split from a common Indo-Hittite stem. Also, there is a strong genetic continuity in the Armenian highlands during the Neolithic and Chalcolithic, and partially also during the Bronze Age and the Iron Age (Lazaridis et al. 2016; Margaryan et al. 2017), explained by a history of genetic isolation from their surroundings (Haber et al. 2016). These data contradict an expansion of peoples from the steppe through the Caucasus.

In Anatolia, the low genetic diversity of early Middle Eastern farmers, which migrated into south-eastern Europe from north-western Anatolia during the early Neolithic, was broken by another wave of ‘eastern’ ancestry that reached south-eastern Europe before at least ca. 3800 BC. These migrants brought CHG ancestry and J-M304 lineages – typical of Caucasus and eastern Iranian populations – to the late Neolithic central and western Anatolia (Lazaridis et al. 2016; Kilinc et al. 2016).

This ‘eastern’ ancestry may have been caused by interactions between central Anatolia and the Fertile Crescent in the late Pre-Pottery Neolithic B (Özdoğan 2008), a migration related to other inter-regional exchanges, or admixture among local populations. The Tepecik-Çiftlik site’s presumed role as an obsidian hub, and its cultural links with the Levant, might have started already before the Pottery Neolithic (Kilinc et al. 2016).

Anthony’s proposal of a western migration route of Anatolian-speaking peoples through the Balkans ca. 4200 BC (Anthony 2007; Anthony and Ringe 2015) is supported by the resurge of haplogroup R1b-M343 (potentially R1b1a1a2-M269 subclades) and the
presence of steppe ancestry in the Balkans, although at a slightly earlier date, in the mid-fifth millennium (Mathieson et al. 2017).

Anatolian Middle Bronze Age migrations ca. 1900 BC saw the destruction of cities, from the Ezero culture in south-eastern Europe to north-west and north-central Anatolia (Mellaart 1958). Although traditionally associated with an east-west movement of peoples, it could well represent the opposite direction, thus including expanding Anatolian-speaking peoples through northern Anatolia, from the west to the central part. Samples from Bronze Age south-western Anatolia (ca. 2800-1800 BC) show the ‘eastern’ contribution of CHG, but lacking steppe-related EHG and WHG ancestry (Lazaridis et al. 2016).

The modern distribution of R1b1a1a2-M269 haplogroup in the Balkans and Anatolia (not reaching the Armenian highlands) points to the posterior migration of R1b1a1a2-M269 lineages with Anatolian languages (Figure 8). Its modern peak around Kosovo can be explained by posterior founder effects, which might have happened during any expansion of peoples in the region in the past four thousand years. We can tentatively assign one of these founder effects to a recent Albanian expansion.

Figure 8. Modern distribution of R1b1a1a2-M269 (xL23) lineages, adapted from Richard Rocca (2012).
III.4.1.1. Etruscan

Samples from the Remedello culture, of haplogroup I2-M438 (dated ca. 3290 BC, ca. 2745 BC, and ca. 1955 BC), and from Ötzi the Iceman, of haplogroup G2a-P15 (ca. 3225 BC), all of northern Italy, show a high affinity with Chalcolithic samples from central Anatolia. This affinity is higher between them than with earlier Anatolian Neolithic populations, which is against the interpretation of Remedello’s ancestry representing a relict population stemming from Neolithic farmers (Hofmanova et al. 2016).

Because of their shared drift with CHG ancestry independent of steppe expansions, and the fact that Kumtepe predates the northern Italian group by some 1,000 years, it has been proposed that they represent a more recent, yet undescribed, gene flow process from Anatolia into Europe. This Anatolian population shows a continued ‘eastern’ migration (Kilinc et al. 2016; Lazaridis et al. 2017).

Nevertheless, the modern distribution of R1b1a1a2-M269 in the Alps and in ancient Tyrrenia might point to an eastern route of the Suvorovo-Novodanilovka settlers of eastern Hungary, who may have mixed with a Balkan population related to such an Anatolian expansion, hence giving support to the theories describing Etruscan as an Anatolian branch (Adrados 1989, 1994). A recent Anatolian connection has also been found by examining mtDNA in modern populations of present day Tuscany (Brisighelli et al. 2009).

On the other hand, all this could well be a sign of independent back and forth migrations through the Adriatic Sea, or just repeated migrations from Anatolia to the Italian Peninsula through southern Europe (Kilinc et al. 2016).
III.5. Late Indo-European

After 4000 BC, different groups were formed in the steppes. In the west, late Sredni Stog and “Post-Mariupol” (“Extended-Position-Grave”) communities, the heirs of the western early Sredni Stog clans, remained in contact with Trypillian villagers, and some assimilation seems to have happened east of the Dnieper ca. 3700-3500 BC. These contacts are supported by the steppe-related ancestry found in a Trypillian individual ca. 3700 BC from the Verteba Cave, of G2a-P15 lineage (Mathieson et al. 2017).

In the east, early Khvalynsk gave way to late Khvalynsk and Repin societies in the Volga-Don region, whose language is to be associated with a common Late Proto-Indo-European (Anthony 2007). The split of R1b1a1a2-M269 into the eastern R1b1a1a2a2-Z2103 subclade must have happened early – possibly during the previous westward expansion of early Khvalynsk clans (of R1b1a1a2-M269 and R1b1a1a2a-L23* lineages) in and outside of the Pontic-Caspian steppes, given the similar forming date (ca. 4200 BC) and TMRCA (ca. 4100 BC).

The earliest aDNA samples of haplogroup R1b1a1a2a2-Z2103 are three individuals found in the late Khvalynsk area in Lopatino I ca. 3000 BC, Ishkinovka I ca. 3000 BC, and Peshany V3 ca. 2985 BC (Haak et al. 2015). All samples from the Samara region are either R1b1a1a2a2-Z2103 or older lineages, except for one R1b1a1a2a-L23 (xR1b1a1a2a2-Z2103, xR1b1a2a-L51) at Lopatino II dated ca. 3000 BC (Haak et al. 2015), which suggests a differentiation of R1b1a1a2a-L23 into its subclades near this region.

Haplogroup R1b1a1a2a1-L51 (formed ca. 4200 BC, TMRCA ca. 3900), given its distribution into west Europe, is hypothesized to have expanded successfully to a certain extent during the common Yamna (“Pit Grave”) period of the Pontic-Caspian steppe cultures, but later and more marginally than R1b1a1a2a2-Z2103 groups.

Given the lack of aDNA from the Western Yamna horizon, and the later westward expansion of R1b1a1a2a2-Z2103 lineages, it is probably safest to assume a western location of R1b1a1a2a1-L51 lineages within Yamna. It would have formed a community with R1b1a1a2a2-Z2103, but somehow separated culturally from it, and thus the two main dialects of Late Proto-Indo-European may have developed separately.

3 Additional information from Sergey Malyshev at FTDNA R1 Basal Subclades project reads Z8129/Y12537 (equivalent Z2103)
Graeco-Aryan (probably including at least Greek, Armenian, and Indo-Iranian) has been argued as a dialect *continuum* or a linguistic community where a number of common innovations were shared at an early time (Mallory and Adams 2007; West 2007). North-West Indo-European – including Italic, Celtic, Germanic, and Balto-Slavic – has been proposed as a group of closely related dialects with some form of shared linguistic history, presumably about the 3rd to 2nd millennium BC, after the initial dispersal of the Indo-European languages but before the emergence of the individual language groups in Europe (Oettinger 1997, 2003; Adrados 1998; Mallory and Adams 2007; Mallory 2013; Beekes 2011).

Tocharian would have been part of this group at an earlier stage, forming a ‘Northern’ Indo-European group – so called because of their later migrations, contrasting with the ‘Southern’ or Graeco-Arian Indo-European dialects (Adrados 1998; Mallory and Adams 2007; West 2007).

Both linguistic communities remained thus in close contact, and are probably to be located to the eastern Don-Volga steppes, spreading across the Pontic-Caspian steppes after about 3300 BC (Anthony 2007). Because of their later expansion, their division could be speculatively traced back to the early division of Volga-Don groups: the western, Don-based Repin culture, and the eastern, Volga-based late Khvalynsk culture (Figure 9).

In light of the most likely original distribution of both dialects in the Pontic-Caspian steppe, the names ‘Eastern’ (for Southern or Graeco-Aryan) and ‘Western’ (for Northern) Late Proto-Indo-European would probably be more appropriate.

The westward and eastward expansion of the Repin culture about 3300 BC is associated to the rapid diffusion of the Yamna horizon across the Pontic-Caspian steppes, and a common, “disintegrating Indo-European” (Bomhard 2015) must have been spoken in this common period, where laryngeals were already unstable. Late Indo-European had possibly already undergone the first common phase of laryngeal loss to leave the traditionally reconstructed long and short vowels (Szemerényi 1967), with a single laryngeal (Polomé 1987) remaining mainly in compounds with sonorants, whose later dialectal evolution is controversial (Adrados, Bernabé, and Mendoza 2010; Clackson 2007).

Radiocarbon dates for the arrival of the Yamna culture in the territory of Ukraine begin ca. 3350 BC, including the early graves with pottery of the Repin type, with the majority
of dates lying in the span 3050-2300 BC, although chronologies vary widely in specific regions (Rassamakin and Nikolova 2008).

III.5.1. Contacts with the Caucasus

The Uruk expansion in Mesopotamia after about 3700 BC intensified during the late Uruk period ca. 3350-3100 BC, and its expansion reached toward the gold, silver, and copper sources in the Caucasus Mountains. The Maykop culture of rich chieftains’ graves with Mesopotamian ornaments developed from this trade network in the North Caucasus Piedmont. A western and probably also a later eastern southern routes, through the shores of the Black and Caspian seas respectively, have been proposed (Anthony 2007).

Steppe-Caucasian trade is supported by Maykop imports found in the north Pontic steppe from the Dniester to the lower Volga in the east, but no Caucasian imports have been found in the Volga-Ural region. Late Maykop peoples – most likely speaking languages ancestral to modern Caucasian languages – probably interacted with individuals from Repin and late Khvalynsk cultures, and the contact was most direct on the lower Don.

Late Maykop graves incorporated carved stone stelae like those of western Yamna. The trading of drugs, wool, and horses has been proposed as main steppe imports into Maykop (Anthony 2007).

It is probably during these times of expansion that admixture in the Middle East was levelled: western Iran had wide impact on Levantine populations in the south and on north-west Anatolians in the west; and the population of eastern Iran and the Caucasus was also composed of the same ancestral components, albeit in slightly different proportions.

Horse trade – including wheels, carts, and the possibility of a quicker transport of metals into Uruk – is proof of an indirect contact between steppe herders and Mesopotamia. The association of exported domesticated horses with experienced breeders and riders of the lower Don offers a solid framework to support the hypothesis of the presence of Late-Indo-European-speaking peoples in Mesopotamia, and thus allow for IE borrowings in Sumerian (Sahala 2009-2013).

The condition of North-West Indo-European as an Euphratic superstratum of Sumerian (Whittaker 2008, 2012) would require a more detailed explanation of internal and external influence, and reasons for potential language replacement and expansion in Mesopotamia.
III.6. Forest Zone

The oldest remains of R1a1a1-M417 lineages are found in the forests north of the Pontic-Caspian steppe: ca. 6425 BC in Yzhnyy Oleni Ostrov (Haak et al. 2015), and ca. 4000 BC in Serteya VIII (Chekunova et al. 2014). During this stage of Rudnyayan culture there is continuity in relation to the previous stage, and contacts are made with eastern Baltic area and through the Western Dvina (Mazurkevich et al. 2009).

The introduction of Typical Cord Ware, which heralded the appearance of Neolithic traits in the Forest Zone, is dated to around 3900 BC, and it was discontinued ca. 3400 BC. It was a relatively uniform culture that covered a vast area ranging from the Urals to the Baltic Sea, and from Northern Ukraine to the Arctic Ocean, although in southern Finland and Karelia variants of the older types remain still in use (Nordqvist and Mökkönen 2016).

Samples from Zvejnieki in Latvia, which had central European ancestry (70% WHG, 30% EHG) during the Mesolithic and Early Neolithic, show a dramatic shift with the introduction of the Comb Pit Ware culture in samples from Zvejnieki (73-100% EHG), which proves that a westward migration of peoples accompanied cultural changes in the region (Mathieson et al. 2017). The last sample obtained, of ca. 2885 BC, reveals another marked transition to a maximum steppe admixture (see below).

The disintegration of the Comb Ware phase began ca. 3500 BC, coinciding with the influence of the Volga-Kama region and the birth of several variants of Asbestos- and Organic-tempered Wares, although no break has been observed in cultural development (Nordqvist et al. 2012). These groups also maintained vast and varying intra- and inter-regional contact networks.

During this period of 3500-3000 BC a shift to drier and cooler conditions is found in the steppes, with steppes expanding, and therefore also Yamna pastoralists and their cattle following them. The emergence of Volosovo and Garino-Vor metallurgy in the fourth millennium has been attributed to external influences from Yamna.

R1a1a1-M417 formation based on modern populations is dated ca. 6500 BC, with a TMRCA ca. 3500 BC, and published research pointing to a slightly earlier date ca. 3800 BC (Underhill et al. 2015), dates that are coincident with the aforementioned cultural and climatic changes.
Individuals from the Forest Zone were not found to have received genetic influx from Anatolian-farmer-related genes during the Mesolithic or Neolithic, and therefore an inner cultural diffusion of pottery, farming and metallurgy is assumed for the population of the Baltic and Dnieper Rapids (Jones et al. 2017).

Between 3500-2000 BC an interruption in cultural continuity in the Forest Zone is found, coinciding with a major change in the environment, with selective felling and subsequent regeneration of forests in the Pit-Comb Ware area (Mazurkevich et al. 2009; Poska and Saarse 2002). This could have been caused by the complex movement of peoples in this period, as reflected by the interaction or “checkerboard of regional cultures covering the rolling hills and valleys of the forest steppe zone” (Anthony 2007), and a complex set of cultures is found in the East European Forest Zone, different from Central European cultures (Czebreszuk and Szmyt 2004).

III.6.1. Corded Ware culture

The first horizon of Corded Ware culture appears in the Early Eneolithic Bubanj-Saluță-Krivodol cultural complex and other Old European cultures in the eastern Balkans only sporadically, possibly from influence of the Sredni Stog culture, at the end of the 5th millennium BC (ca. 4200 BC), in territories of autochthonous ceramic forms not associated with the Suvorovo-Novodanilovka expansion. This pottery developed in a western Pontic territory where corded ornament seems correlated with sheep herding. The spread of this pottery is clearly (and almost exclusively) identified with the Coțofeni group in the 2nd Corded Ware horizon, as part of the cultures of the Lower Danube and northern Bulgaria in the 4th millennium and the Cucuteni-Trypillian culture. These cultures are connected with movements of steppe-related Cernavodă I society in the Danube delta, to the north into Coțofeni, and south into Ezerovo (Bulatović 2014).

Samples from the Balkans at Varna I (ca. 4630 BC), and Smyadovo (ca. 4500 BC), before or possibly coincident with the Suvorovo-Novodanilovka expansion, and 1,500 years before the Yamna expansion, already show the so-called ‘Yamna component’ (Mathieson et al. 2017). Before that, more than 20 samples from the Ukraine Eneolithic at Deriivka and Volynia (ca. 5150 BC) showed mostly EHG ancestry, like Ukraine Mesolithic samples, but with a contribution from WHG ancestry (see above), so the contribution of CHG to both Ukraine Middle Neolithic and Old European samples must have happened later.
The Corded Ware culture territory expanded from the Coțofeni territory to the south during the Eneolithic period, except for the central Balkans, where new steppe elements are noticed during this period. The Usatovo culture, settled in the territory of the Trypillian culture, replaced the Coțofeni culture at the time of the expansion of the third horizon of the Corded Ware culture into Central Europe.

There was a long-ranging connection between the north-west Pontic steppe area and the border of the Forest Zone up to the eastern Baltic area, centred on the Dniester-Bug limes (encompassing the Dniester, Dnieper, and Bug rivers). It also included the areas between the Vistula and the Dnieper (with the Lesser Poland area) – which topographically form a natural continuum. The origin of this expanding third horizon has been placed, as one of the best defined archaeological ideas in Archaeology, in this region between Lesser Poland and adjacent regions of Ukraine and Slovakia, confirmed by radiocarbon analysis to ca. 3000-2900 BC (Kristiansen 1989; Włodarczak 2008; Kristiansen et al. 2017; Anthony and Brown 2017; Kadrow 2008).

Different connecting routes to the north were used by Old European (and especially Trypillian culture) societies from the steppe, forest-steppe and Forest Zone, influencing the Funnelbeaker and Baltic cultures for millennia (Klochko and Kośko 2009; Szmyt 2013; Czebreszuk and Szmyt 2004; Kadrow and Zakościelna 1999). This natural continuum in eastern Europe saw large scale economic and social changes, with Baden and Globular Amphorae Cultures playing a major role (Kristiansen 1989; Kristiansen et al. 2017).

The connection between pre-Caucasian (Maykop) and Late Trypillian cultures that had moved to the left bank of the Dnieper points not only to Caucasian imports, but to a likely Caucasian immigration in a series of small shifts or ‘shuttle’ movements, possibly with the aim of exchange, trade, spoils of war, borrowing of technological devices, etc. This migration is linked to the creation of “bridge” communities, like the Zhyvotylivska-Volchans’k cultural group, and the Late Trypillian Gordionesti group (Ivanova and Toschev 2015).

A sample identified as an Ukraine Eneolithic outlier from Deriivka (ca. 3500 BC), and a Trypillian outlier from the Verteba cave (ca. 3325 BC), show contributions from both CHG and Middle Eastern farmer ancestry (Mathieson et al. 2017). The new farmer ancestry found further illustrates the complexity of human interaction in this western
The roots of the third horizon of the Corded Ware culture is to be found at the end of the 
4th millennium in Podolia and Volynia, and scattered Corded Ware sites from these 
regions to the east of the Bug river show old and young stages of the culture (Kadrow 
2008). In this region, Funnelbeaker traits are not found, and the late Globular Amphora 
culture expansions to this region (after ca. 2950 BC) cannot account for its migrations. 

In Lesser Poland, during the first 300 years of its existence, the Corded Ware culture 
developed among the settlements of the agrarian Baden and Globular Amphora cultures, 
without mixing (Włodarczak 2001), among a complex regional picture formed during the 
4th millennium (Zastawny 2015; Wilk 2016).

Sampled individuals from Globular Amphora culture in Poland and Ukraine form a tight 
genetic cluster, showing genetic homogeneity kept over a large distance, with 25% WHG, 
which suggests a persistent frontier between east-central and eastern European groups. 
One sample from the Baden culture also clusters closely with previous and posterior 
Balkan samples from neighbouring regions (Mathieson et al. 2017).

At the end of the Trypillian culture, herding/hunting trends intensified, and the 
agricultural system collapsed, with people moving to the steppe zone, as confirmed by 
the presence of numerous graves to the south (Rassamakin 1999). At the same time, the 
Trypillian world absorbed a foreign tradition related to materials of settlement sites of the 
Dnieper steppes – such as the late Sredni Stog culture –, like cord impressions and burial 
rites similar to the later Corded Ware culture, marking also the transformation of decors 
and changes in their interpretation (Palaguta 2007).

The similarity in burial rituals between Yamna and Corded Ware made Gimbutas define 
a common “Kurgan people”, whose relationship has also been long supported by 
Kristiansen (Kristiansen 1989; Kristiansen et al. 2017). An equivalence of both burial 
rites has been, however, rejected (Häusler 1963, 1978, 1983), and it is generally agreed 
that the Yamna culture did not expand to the north of the Tisza River.

The importance of horse exploitation in Deriivka, in the forest-steppe zone of the north 
Pontic region along the Dnieper region, during the Middle Eneolithic period (probably ca. 
3700-3530 BC), suggests that horses played a significant role in the life of this Sredni
Stog community (Anthony and Brown 2003). In its late period (ca. 4000-3500 BC), this culture had adopted corded ware pottery, and stone battle-axes.

However, this western steppe peoples were mainly hunters (Rassamakin 1999), and the ‘herding skill’ essential for wild horse domestication seems absent (Kuzmina 2003). All this has been confirmed with zooarchaeological evidence and new molecular and stable isotope results, suggesting an absence of horse domestication in territories of the late Sredni Stog culture in the north Pontic steppe (Mileto et al. 2017), before the advent of migrants from the Indo-European-speaking Repin culture.

After ca. 3300 BC global climatic changes increased fluvial activities in river valleys and caused deforestation, intensified by human agents (due to agricultural needs), which favoured pastoralism and nomadisation of the settlement system, and a consequent change of the social structure (Kristiansen 1989; Kadrow 2008). These changes were stabilised by a new ideology and new symbols imprinted in material culture, a new “picture of the world” of the emerging community, consisting of new and old local elements, into a new, original Corded Ware culture (Kadrow 2008; Habermas 2002).

Samples from the Balkan Neolithic and outliers from west Ukraine and Trypillia that show Yamna ancestry (see above) cluster closely together, and close to later central European samples from Corded Ware and Bell Beaker cultures. However, they show less ‘east Yamna’ ancestry than some later Corded Ware samples, and other Trypillian samples cluster closer to Balkan samples. This suggests that the original expanding population from the third Corded Ware horizon – if directly derived from the sampled population of the Trypillian culture – received further CHG ancestry, potentially during its later adaptation to the north Pontic steppe and the mixture with its population.

There is an increased ‘Yamna component’ found later in samples from the Yamna culture in Ukraine. An original migration of the Corded Ware culture from the north-west Pontic steppe and steppe-forest zone is therefore likely, and may account for the later ‘Yamna component’ found in Yamna and Corded Ware individuals (Allentoft et al. 2015; Haak et al. 2015; Mathieson et al. 2015).

No direct cultural connection has been found in this area with Yamna migrants (Bulatović 2014). Only later, during the contemporaneous Corded Ware and Yamna migration waves were direct contacts possibly between Yamna and Corded Ware herders on the upper Dniester region (Anthony 2007; Gimbutas 1977).
The most recent direct connection of the north Pontic steppe to Central European areas came from Usatovo, which continued the previous Gordionesti group. Usatovo migrants seem to have penetrated in the first half of the 3rd millennium BC northward up the Dniester, South Bug and Dnieper valleys, as Globular Amphorae and Corded Ware cultures expanded to the east beginning ca. 2900 BC, forming the Middle Dnieper culture in the forest-steppe zone around Kiev ca. 2800-2600 BC (Anthony 2007).

Corded Ware culture research is usually discussed as a purely east-central European phenomenon. However, recently obtained dates suggest that the appearance of Corded Ware in central Russia (either of early Fatyanovo or maybe proto-stages) may had begun from 2700-2600 onwards, and earlier still in the Baltic (Lougas, Kriiska, and Maldre 2016).

Eastern influence is found in the southern Baltic and Estonia, connecting cultures previously identified as non-Corded Ware to the emergence of the new cultural expansion, with continuums proposed between late Comb Ware and Corded Ware pottery.

The communication between Forest Zone hunter-gatherers had old roots, and Corded Ware chronology needs further refinement, because Corded Ware was present in the northern Baltic Sea region since ca. 2800 BC (Nordqvist 2016). This connection may explain the prevalence of R1a1a1-M417 subclades during the Corded Ware expansion.

**III.6.2. Corded Ware migration**

The Corded Ware culture has been argued to practice exogamy – most adult women being of non-local origin – based on a recent work on diet and mobility (Sjogren, Price, and Kristiansen 2016), and mtDNA has been documented to be more varied among Corded Ware females than men (Lazaridis et al. 2014). Exogamy is described as a well-established practice over a long period of time, and the origin of females has been linked to Neolithic cultures still residing on the higher elevations in the region shared with Corded Ware cultures, which colonized rivers and valleys (Kristiansen et al. 2017).

A violent picture has been proposed for the Corded Ware culture society (Haak et al. 2008), with warrior youth bands with seasonal activities, in a way similar to that documented in the Russian steppe from the Bronze Age onwards (Kristiansen et al. 2017).

In a first stage of the CWC migrations, their nomadic way of life is represented by site-camps close to tumuli, located on routes of their long-lasting travels. In a second phase,
semi-nomadic economic allowed for settlement micro-regions to appear, and cemeteries were a kind of landmark during the population’s seasonal migrations (Machnik 2004). Necropolises were significant areas to which they returned periodically, establishing eventually site-camps, beginning their transformation into Early Bronze Age groups with a more stable economy (Witkowska 2006).

Analysis of aDNA has revealed that the plague was a prehistoric disease, associated with the Eurasian steppes, and linked to the Corded Ware culture expansion (Rasmussen et al. 2015; Andrades Valtueña et al. 2017), which connected vast areas in central-east Europe in a relatively short period. Also, the population to the north of the loess belt of the northern European plain is known to have been much lesser than in south-east Europe (Müller 2013).

This demographic disadvantage of central-east European lands might have provided part of the ‘push’ for the migration and expansion of the Corded Ware population (Anthony and Brown 2017), which show a clear genetic homogeneity – close to the ‘Yamna component’ – from western to eastern Europe, in spite of the proposed practice of exogamy. This explosive migration over a sparsely populated area accounts for the expansion of R1a1a1b-Z645 lineages, possibly originally from the Dnieper-Dniester or Middle Dnieper region.

The most recent sample from Zvejnieki, of the Late Neolithic / Corded Ware (ca. 2885 BC), clusters quite closely with Yamna samples, revealing a break of the regional population with the samples from a thousand years earlier (Figure 10), which were closer to EHG (Mathieson et al. 2017). The lack of more EHG ancestry – as expected in a sample potentially generations away from southern migrants (and thus admixed with the local population) – could point to the previous presence of the ‘Yamna component’ in neighbouring regions of east-central Europe, or to a recent contribution from the steppe.

Its mtDNA haplogroup is U5a1b. Although different U5, U4, and U2 subclades are commonly found in hunter-gatherer groups, this precise haplogroup is found previously only in four Sredni Stog samples from Deriivka (ca. 5150 BC), and later in central European Benzingerode (ca. 2275 BC), and in a sample from Przeclawice (ca. 1790 BC), of the Únětice culture (Haak et al. 2015; Mathieson et al. 2017). This may give support to the origin of the Corded Ware culture near the region of the Sredni Stog culture, and
Indo-European demic diffusion model points to the north-west Pontic population (and women in particular) as a potential origin of the steppe admixture found in the sample.

Figure 10. Image modified from Mathieson et al. (2017). “Structure and population change in European populations with mainly hunter gatherer ancestry. Background: Effective migration surface inferred using 109 Mesolithic and Early Neolithic genomes. Blue indicates regions of high migration and red regions of low migration (relative migration rate m). Pies: Each pie represents one or more individuals. Blue, red and green segments pies show, respectively, WHG, EHG and CHG ancestry – we chose a 2-population model if it fitted in qpAdm with p>0.01. Pies with 100% ancestry are those that were fixed as source populations. In Latvia, Samara and Ukraine, sets of three pies show the transition in ancestries, approximately from Mesolithic-Neolithic-Yamnaya. Populations with labels in brackets were not used to fit the migration surface. B: Ancestry transitions for hunter-gatherer populations in the 3-population
WHG+EHG+CHG model. Dashed lines show populations with geographic continuity”. Original under a CC-BY-NC 4.0 International license.

R1a1a1b-Z645, split from R1a1a1-M417 ca. 3500 BC, shows a TMRCA of ca. 3000 BC, coinciding with the formation date for mainly-European subclades R1a1a1b1-Z283 and R1a1a1b1a-Z282, and mainly-Asian subclade R1a1a1b2-Z93. The common TMRCA for R1a1a1b1-Z283 and R1a1a1b1a-Z282 suggests an expansion at nearly the same time as peoples of Corded Ware cultures are supposed to have migrated east- and westward, reaching the Middle Elbe-Saale region – where most aDNA samples analysed come from – about 2750 BC (Figure 13). The common TMRCA of 2700 BC for modern Asian lineages gives support to a later successful expansion into Asia centred on the eastern part of the Pontic-Caspian steppes (see Indo-Iranian).

Y-DNA samples of haplogroup R1a-M420 (probably R1a1a1-M417) are found in central Corded Ware culture groups (Allentoft et al. 2015; Mathieson et al. 2015; Haak et al. 2008; Mittnik et al. 2017; Saag et al. 2017): one sample in Bergreinfeld (ca. 2647), two in Eulau (ca. 2600 BC), one4 from Kyndelöse (ca. 2670 BC), seven from Esperstedt (one dated ca. 2430 BC and other six ca. 2275 BC), and two from the Battle Axe culture – one from Viby (ca. 2500 BC), and one from Ölsund (ca. 2350 BC).

Five samples from Erperstedt have been inferred to be relatives via paternal line (Figure 11), which confirms their common sharing of bottom subclade R1a1a1-M417 (Monroy Kuhn, Jakobsson, and Günther 2017). The oldest sample from Erperstedt (labelled I0104), a second-degree relative relative to the rest, has been found to cluster the closest to steppe samples, closer than any other Corded Ware sample, previous or posterior, or any samples from Sintashta, or Potapovka (Figure 12). This, connected with the exogamy prevalent among Corded Ware peoples, and the nomadic nature of its culture, precludes a proper interpretation of the ancestry found in the family, since it might have come quite recently directly from the steppe, at a time when the Catacomb and Potapovka cultures had already replaced the Yamna culture.

In terms of mtDNA the Corded Ware culture shows common hunter-gatherer haplogroups U5, U2, or U4 (Brandt et al. 2013). Out of seventeen samples from the Corded Ware culture, four (all of Erperstedt), include different J1c subclades, only found previously

4 Published as corresponding to Nordic Middle Neolithic culture, additional information by Vladimir Tagankin revealed a branch typical of modern Nordic (R1a1a1b1a3-Z284) subclades, and a new date of ca. 2475 BC, including reduction for high marine signal.
accompanying the expansion of Neolithic Middle Eastern farmers, including the Globular Amphora culture (in two out of nine samples), and later in one out of five samples of the Sintashta culture (Mathieson et al. 2017).

Two samples from Tiefbrunn of haplogroups R-M207 (ca. 2755 BC) and R1b1-L278 (ca. 2725 BC) are of dubious subclade – the first could be R1a1a1-M417 or R1b1-L278, the second might correspond to older European hunter-gatherer lineages, or they might be associated with the expansion of Corded Ware cultures from the Balkans. Other non-R-M207 samples from central Corded Ware cultures include the two oldest ones, of IJ and G2a-P15 lineages, from Jagodno ca. 2800 BC (Gworys et al. 2013); and one of haplogroup P-P295 in Esperstedt from ca. 2275 BC.

In the Baltic and the Forest zone, four samples are dated around 2500 BC: haplogroups R1a1a1b-Z645 and R1a1a1b1-Z283 in Kunila, and two samples of haplogroup R1a1a1b-
Z645 in Ardu. These samples together with its previous presence in Usvyatyan culture (ca. 2500 BC) and in Naumovo and Sertaya II (Chekunova et al. 2014), and its continuity in later times suggest that R1a1a1b-Z645 lineages almost fully replaced the previous R1b-M343 lineages in the eastern Baltic around the time of the Corded Ware culture expansion.

Figure 12. Detail of PCA analysis of free datasets including Minoans and Mycenaeans (Lazaridis et al. 2017), and Scythian and Sarmatian (Unterländer et al. 2017) datasets. PC2 vs. PC1. The graphic has been arranged so that ancestries and samples are located in geographically friendly axes similar to north-south (Y), east-west (X). Symbols are used, in a simplified manner, in accordance with symbols for Y-DNA haplogroups used in the maps. Labels have been used for simplification of important components. Areas are drawn surrounding Yamna, Poltavka, Afanasevo, Corded Ware (including samples from Estonia, Battle Axe, and Poltavka outlier), and succeeding Sintashta and Potapovka cultures, as well as Bell Beaker. Corded Ware sample I0104, from Erperstedt, has also been labelled. See full, high-resolution version at <https://indo-european.info/ie/Greek>.

The oldest R1a1a1-M417 lineages of Central Europe are found in early Corded Ware groups, while ancient DNA from Neolithic Linear Pottery (ca. 5500–3500 BC) and Globular Amphora (ca. 3400-2800 BC) cultures have been found to correspond mainly to I2-M438 and G2a-P15 lineages, with no steppe-related ancestry in admixture analyses (Haak et al. 2015; Mathieson et al. 2017). This raises the possibility of a wide and rapid spread of R1a1a1-M417, and especially R1a1a1b1-Z283 subclades in Europe (Underhill et al. 2015) associated with the expansion of the Corded Ware culture. The diversification of R1a1a1b2-Z93 in the Middle East and South Asia remains more obscure (Underhill et al. 2015).

However, in samples obtained from Estonia it was seen that admixture between Corded Ware culture farmers and Comb Ceramic culture hunter-gatherers may have been limited
in males of R1a1a1-M417 lineages. Also, the presence of a genetic component associated with Caucasus hunter-gatherers – also present in Yamna migrants, Eastern hunter-gatherers, and individuals from the Estonian Comb Ceramic culture, means that the expansion of the Corded Ware culture cannot be seen as the sole means for the spread of this genetic component, at least in eastern Europe (Saag et al. 2017).

Indo-European demic diffusion model
III.7. Tocharian expansion

The Early Bronze Age Afanasevo culture (ca. 3500-2600 BC) in the Altai-Sayan region are genetically indistinguishable from Yamna in admixture analyses using exclusively female samples (Allentoft et al. 2015).

Linguists have placed Proto-Tocharian together with the North-West Indo-European dialectal zone, but showing innovations compatible with an isolated development (Mallory and Adams 2007). The findings of Anthony (2007) regarding the early spread of a group from the Repin society into Afanasevo (ca. 3000-2800 BC) supports the eastward expansion of a western Yamna group, and therefore the position of Tocharian as part of an early ‘Northern’ Indo-European split.

The sample of haplogroup R1b1a1a2a-L23 (xR1b1a1a2a2-Z2103, xR1b1a1a2a1-L51) at Lopatino II (ca. 3000 BC) in the Samara region (Haak et al. 2015) might be speculatively interpreted as a member of one of several clans different from those sharing R1b1a1a2a2-Z2103 lineages in the Volga-Ural zone, a region that shows an overwhelming majority of R1b1a1a2a2-Z2103 lineages in previous (late Khvalynsk) and posterior (Poltavka) cultures.

In the Copper Age, one sample of the Bolshemys culture (ca. 4th millennium BC) and three individuals from the succeeding Afanasevo culture (ca. 3000 BC) belong to R1b1a1a2-M269 lineages, and three more to haplogroup R1b1-L278 (Holland 2014), which more clearly points to the expansion of male migrants from Yamna (Figure 14). A division between a Mongol Altai and a Siberian area of Indo-European influence was clearly seen in a later period, where only one sample from the Okunev culture (ca. 2300-1800 BC) was of R1b1a1a2-M269 lineage, with three samples of haplogroup N1a1-M46 (see above N1c1-M46) that suggest a change of the paternal line in the region (Holland 2014).

However, the main lineages found in Tarim Basin mummies of the Xiaohe necropolis (dated ca. 2000 BC), 11 out of 12 male remains, were of R1a1a1-M417 lineages (Li et al. 2010). If these were actually ancestors of later Tocharian speakers, this would support a cultural assimilation of Pre-Tocharian into peoples of R1a1a1-M417 lineages, and also that the migration of the Pre-Tocharian language to the Tarim Basin was coincident with

---

5 These R1b1-L278 samples were most likely R1b1a1a2-M269, according to a phylogenetic network created by Holland (2014).
the Andronovo expansion, which “had transformed the steppes from a series of isolated cultural ponds to a corridor of communication” (Anthony 2007). That change is clearly attested in aDNA samples from the final Bronze Age, where no sample of haplogroup R1b1a1a2-M269 is found, but haplogroup R1a1a1b2-Z93 is found in this area and others Eurasian sites distant from each other. This period marks also the appearance of Paleo-Siberian lineages in south Siberia (Holland 2014).

From all paternal lineages found in modern Uyghurs (Zhong et al. 2013), only that of R1b1a1a2-M269 lineages cannot be explained by earlier or later population expansions.

III.8. Late Indo-European expansion

Anthony’s third migration wave of ca. 3000-2800 BC (Anthony 2013) must include the expansion of peoples of R1b1a1a2a2-Z2103 and R1b1a1a2a1-L51 lineages into Europe.

The most obvious material division within the early Yamna horizon was between east and west (see above Figure 13). According to forming and TMRCA dates of R1b1a1a2a2-Z2103 lineages, communities carrying different R1b1a1a2a2-Z2103 subclades might have already developed differentiated groups based on clans within the Volga–Ural–North Caucasian zone, a part of the more mobile eastern Yamna pastoral economy (Anthony 2007). In a central region, a late sample at Stalingrad Quarry ca. 2675 BC – after the migrations into south-eastern Europe – shows a subclade R1b1a1a2a2c-Z2106 (Allentoft et al. 2015).

The remaining North-West Indo-European community – separated from Pre-Tocharian speakers – lived more likely around the South Bug – Lower Don steppe, and it is possible that their lineages were dominated by R1b1a1a2a1-L51 lineages, which had expanded with successful clans probably by 3900 BC according to its TMRCA, and to the same time of formation of subclade R1b1a1a2a1a-L151.

The western community expanded west possibly early within the southern stream of the third migration wave (with a TMRCA ca. 2800 BC for R1b1a1a2a1a-L151), from the Bug-Dnieper-Azov steppes into the lower Danube valley and Bulgaria. They pushed farther up the Danube to the middle Danube valley in eastern Hungary through an Old Europe in crisis – contemporary with late Baden / Cernevoă III (Anthony 2013, 2007).

III.8.1. Yamna migration

Yamna settlements spread westward into the Danube valley and to the north up the Prut River, beginning ca. 3100 and continuing up to ca. 2800 BC (see above Figure 13), which might have given another push to the migration of Corded Ware peoples to the north.

A real current of immigration is noticed ca. 2950 BC (Heyd 2012), later than the expansion of the Corded Ware horizon, with the earliest settlement appearing in Varna bay in Bulgaria. A large settlement appeared to the west in south-western Romania divided by the Danube River (Tarnava-Rast group). Migrants pushed west, appearing west of the Iron Gates in Jabuke, but the largest number of migrants ended up in the central Carpathian basin. Another settlement appeared south of the Varna bay, in the Balkan uplands (Kovachevo-Troyanovo), within the Ezero culture (Anthony 2007).
A rapid decline in human activities peaked in Central Europe between 4000–3000 BC and recovered only after 3000 BC, accelerating after 2500 BC. This decline has been related to adaptation processes during climatic changes (Kolář et al. 2016; Gardner 2002) – which might have helped the expansion of Yamna settlers into scarcely populated areas. The area recovered after 3000 BC with a more humid climate that favoured grassland productivity (Harrison and Heyd 2007), accelerating after 2500 BC, at the same time as the horse, the wheel, and pastoralist societies expanded into these areas. Their migration seems not to have been a traumatic event. There might have been local conflicts and raids, but there are signs of interaction with contemporary societies, as well as exchange of ideas, innovations and material culture (Heyd 2012).

The massive Yamna migration in south-east Europe is said to have been well organized, either in loose family alliances (the most likely scenario) or in clans, in any case with a clear leadership and structure (Heyd 2012). There were possibly more than one wave of migrations, with differences noted north and south of the Balkans, which could correspond to the different lineages expanded to the west and south. At least one migration wave seems to have come from the north Pontic region, due to the presence of wagons (or parts of wagons) and stelae – characteristic of the Kemi-Oba and neighbouring zones of the South Bug - Lower Don steppe –, in burial mound cemeteries of Yamna settlements (Kaiser and Winger 2015).

Large stone anthropomorphic stelae seem to have first appeared in the Mikhailovka I culture in the second half of the 4th millennium. Mikhailovka I areas were replaced by the Usatovo culture, but its culture continued in the Kemi-Oba culture of Crimea. Carved stone stelae appear to have expanded in frequency and elaboration in both territories, and in part of the north Pontic steppes, after about 3300 BC (Anthony 2007). Strikingly similar stone stelae appeared later in the Caucasus, Troy, and also in Central and Western Europe, and with special frequency in the Swiss Alps and in the Provence. A maritime route for such cultural expansion has been proposed, which would justify e.g. its early presence in Troy (Anthony 2007).

Mainly associated with funerary customs in the Yamna horizon, the use of other carved anthropomorphic stones seem to herald the influence of the Yamna culture in Europe, just like the building of tumuli, the enhancement of gender distinctions, and the internationalisation of special objects made of rare materials as status indicators. This influence was seen in the Corded Ware/Single Grave culture in central and eastern Europe.
in the east, Vučedol in the western Balkans, Makó/Kosihý-Čaka/Somogyvár in the Carpathian Basin and even the Bell Beaker culture in south-western Europe around 2700/2600 BC (Harrison and Heyd 2007). Stone stelae and figurines might have also been used quite differently, or for different purposes, in certain local cultures (Robb 2009; Díaz-Guardamino 2014).

Radiocarbon dates from the north Pontic steppe show the late presence of steppe materials cultures in the Carpathian EBA (ca. 2500 BC), in the Makó-Kosihý-Čaka/Somogyvár-Vinkovci/Late Vučedol, and others like Schneckenberg-Gлина III, Csepel, or Early Nagyrév. These cultures have been argued to form a cultural unity, and it is proposed that such influence may have come from Yamna settlements on the left bank of the Tisza River (Rassamakin and Nikolova 2008).

Figure 15. Modified from Mathieson et al (2017). “Individuals projected onto axes defined by the principal components of 799 present-day West Eurasians (not shown in this plot for clarity). Projected points include selected published individuals (faded colored circles, labeled) and newly reported individuals (other symbols; outliers shown by additional black circles). Colored polygons indicate the individuals that had cluster memberships fixed at 100% for the supervised admixture analysis [see below]”. Original under a CC-BY-NC 4.0 International license.
Such wide-ranging European cultural influence was not accompanied by significant steppe migrations in the case of the Corded Ware and Proto-Beaker cultures (Harrison and Heyd 2007), and were mostly the result of a strong influence by the so-called Yamna package. In fact, while the west Beaker culture arriving in Sion (ca. 2500 BC) had stelae marking individual burials, it seems to be the destruction of these stelae the event that signals the arrival of east Bell Beakers ca. 2425 BC (Heyd 2012), probably associated with a later expansion of Yamna lineages into western Europe. This destruction shows the doctrinal conflict within the Bell Beaker ideology, with two groups (a south-western and an eastern one) in conflict: another change in burials, to cists reflecting the new family-based structures, and with heads pointing to the rising sun in the east.
It seems that during the expansion of the western Beaker culture to the east, Yamna migrants reached southern Germany. The farther west that a Yamna burial has appeared is Bleckendorf in Saxony-Anhalt, dated ca. 2675 BC (Harrison and Heyd 2007).

Individuals from Yamna in the north Pontic steppe have been sampled near Pokrov, at the Dnipropetrovsk region, two at Shevchenko (ca. 3000 BC) and one at Ozera (ca. 3005 BC), all of them showing the so-called ‘Yamna component’, but the latter with a distinct contribution and a ‘southern’ drift in PCA (see Figure 15 and Figure 16), toward Iran and Levant Neolithic (Mathieson et al. 2017). ‘South-eastern’ admixture is found in a Yamna sample from Bulgaria at Mednikarovska (ca. 2955 BC).

Support for the western migrations of R1b1a1a2a-L23 lineages from Yamna is found in the sample of haplogroup R1b1a1a2a2-Z2103 from Beli Manastir, of the Vučedol culture, dated ca. 2775 BC (Mathieson et al. 2017), with an admixture similar to the sample at Mednikarovo. This eastern Yamna subclade is also found later in a Bell Beaker individual from Hungary at Szigetszentmiklós ca. 2330 BC (Olalde et al. 2017). Other early Bell Beaker samples (see below) and the modern distribution of basal R1b1a1a2a-L51* (Figure 17) in Central Europe all suggest a migration of peoples from the Yamna culture to the west – and mainly western Yamna R1b1a1a2a-L51 lineages – along the Danube.

Figure 17. Modern distribution of haplogroup R1b1a1a2a1-L51*. Adapted from Richard Rocca (2012)
III.8.2. The expansion of Balkan Indo-European

Later samples of R1b1a1a2a2-Z2103 subclades suggest a genetic continuity of early Khvalynsk population in the pastoral groups within Volga–Ural–North Caucasian steppe: early Yamna in Ekaterinovka ca. 2840 BC, late Yamna in Temrta IV ca. 2760 BC, and two early samples from the Poltavka culture in Lopatino II ca. 2770 and in Kutuluk ca. 2680 BC (Mathieson et al. 2015).

Eastern groups – if only because of their location within Yamna – may have followed the southern stream somehow later compared to western R1b1a1a2a1-L51 lineages, possibly through the settlement of Kovachevo-Troyanovo. The sample at Beli Manastir of the Vučedol culture (see above) shows how different groups of different lineages might have been involved in the Yamna migration to south-eastern Europe.

The population of Bronze Age tell settlements from the Carpathian Basin show ritual practices in common with the Mycenaean world, with an official cult practiced in specific buildings, like temples destined to serve the entire community, complemented by a family cult, represented by fireplaces and small altar pieces or miniature wagons made of clay.

There were a potential solar cult (reminiscent of the Zeus/Apollo cult) before its appearance later in the Urnfields culture and in the Nordic Bronze Age; human sacrifices potentially addressed to a deity of war (such as Ares); food offerings potentially for some deity of fertility (like the “Great Mother”); animal idols and drinking vessels; a “hero cult” with weapons and other metal objects, etc (Gogaltan 2012). All of this strengthens the idea of a common Balkan community, in contact with central European cultures, from which the Proto-Greek population could have migrated to the south.

R1b1a1a2a2-Z2103 lineages appear in modern populations (Figure 18) as:

- A Balkan group – mainly of R1b1a1a2a2c-Z2106 lineages, like the sample found in Stalingrad Quarry dated ca. 2670 –, with a central group dominated by R1b1a1a2a2c1a-Z2110 found in Greece and Central Europe.
- An Armenian group – of R1b1a1a2a2a-L584 lineages – found mainly in the Armenian highlands.
- A northern group – of R1b1a1a2a2b-L277.1 lineage, possibly related to the Vučedol sample ca. 2775 BC and the Bell Beaker sample from Szigetszentmiklós ca. 2330 BC (see above).
The modern distribution of R1b1a1a2a2-Z2103 subclades around the Balkans gives therefore support to the existence of a Paleo-Balkan Sprachbund or dialect continuum (Kortlandt 2003).

**III.8.2.1. Greek**

The southern stream of Yamna migrants showed a later expansion from the Lower Danube and Tisza rivers to the southern tip of the Balkan peninsula, and its population is probably represented by the transition of Early Helladic II to Early Helladic III period ca. 2200 BC (Gimbutas 1977), coincident with the arrival of Minyan pottery style. Others have proposed a later date, the beginning of the Middle Helladic culture ca. 2000-1900 BC (Beekes 2011).

A recent study of Minoan samples from Crete (ca. 2900-1700 BC) and Mycenaean samples from mainland southern Greece (ca. 1700-1200 BC) have shown an introduction of CHG without EHG compared to Mesolithic samples, thus independent of a steppe invasion. This eastern influence is also found in Y-DNA haplogroups, all four of them J-M304. This eastern influence may have arrived from Anatolia, since Neolithic samples from central Anatolia at Tepecik-Çiftlik already show it some millennia earlier (Kilinc et
al. 2016), and more recent samples from south-western Anatolia (ca. 2800-1800 BC) show it too.

Mycenaean samples show a ‘northern’ contribution, apart from the Neolithic farmer and ‘eastern’ ancestry shown by Minoan samples (see Figure 19). This ‘northern’ contribution may be interpreted in part as a steppe ancestry similar to that found in the Balkans in the third millennium, suggesting a rapid migration of Proto-Greek from the Balkans, although it is found only in samples from mainland Greece, and not in one sample from Crete (Lazaridis et al. 2017).

The introduction of millet and the horse in northern Greece seem to coincide, pointing to their introduction by horse breeding, millet-consuming cultures from the north or north-east, via river valleys leading to the Danube (Valamoti 2016).

The potential invasion and assimilation of R1b1a1a2a2-Z2103 settlers in Proto-Anatolian-speaking populations in Cernavodă III and Ezero cultures further confounds modern genetic studies, and more aDNA samples are needed to more clearly depict the expansion of both populations. Such a contact may have happened early during the southward migration, as suggested by Anatolian loanwords found in Greek.

The complexity of Y-DNA haplogroups found in the modern population of Greece bears witness to the thousands of years of European and Asian interaction in the formation of its peoples. While it is clear that Yamna ancestry does not represent a big part of its modern population in any study published to date, it is unclear how the ancient population was affected by the migration of peoples of R1b1a1a2a2-Z2103 lineages.

Analysis of modern Greek and Cretan lineages point to a Neolithic expansion of haplogroup R1b1a1a2-M269 in the region, which was found nearer to Italian than to Balkan lineages (King et al. 2008). Analysis of Greek-Cypriot modern populations revealed the presence of R1b1a1a2a2-Z2103 lineages in easternmost and westernmost sides of the island, with central (R1b1a1a2a2c1a-Z2110) lineages appearing only in the east (Voskarides et al. 2016). The early attestation of Mycenaean Greek in the island points to an early expansion of R1b1a1a2a2c1a-Z2110 lineages, but the early Anatolian influence over the island precludes a precise identification of their origin.
Figure 19. PCA analysis of free datasets including Minoans and Mycenaeans (Lazaridis et al. 2017), and Scythian and Sarmatian (Unterländer et al. 2017) samples. PC2 vs. PC1. The graphic has been arranged so that ancestries and samples are located in geographically friendly axes similar to north-south (Y), east-west(X). Symbols are used, in a simplified manner, in accordance with symbols for Y-DNA haplogroups used in the maps. Labels have been used for simplification of important components. Areas are drawn surrounding Yamna/Poltavka, Corded Ware (including samples from Estonia, Battle Axe, and Poltavka outlier), and succeeding Sintashta and Potapovka cultures, as well as Bell Beaker. Corded Ware sample I0104, from Erperstedt, has also been labelled. Continues on the next page. See full, high-resolution version at <https://indo-european.info/ie/Greek>.
Most R1b1a1a2-M269 lineages in Greece are probably linked to the Bronze Age expansion associated with the Minyan pottery, in turn linked to an earlier Yamna expansion into south-eastern Europe. Most R1a1a1b2-M417 lineages – whose proportion in Greece and in historically Greek Anatolia increases with latitude and in peripheral areas
– are thus to be linked to the recent southern migration of Balkanic populations of R1a1a1b1-Z283 lineages (mainly Slavs, but also Aromanians, and Albanians), and to the western spread of Iranian peoples of R1a1a1b2-Z93 lineages (Heraclides et al. 2017).

A potentially older invasion of certain R1a1a1b2-Z93 lineages during the Bronze Age could be supported by the finding of haplogroup R1a1a1b2-Z93 in an individual from Merichleri, in Bronze Age Bulgaria, ca. 1690 BC (Mathieson et al. 2017).

On the controversial ethnicity and language of the Sea Peoples – and the closely related Philistine question – (Woudhuizen 2006; Maeir, Davis, and Hitchcock 2016; Middleton 2015), genetic research points to a mixture of steppe ancestry found in the Lebanese population that occurred ca. 1740-160 BC (Haber et al. 2017). This mixture occurred probably in the earlier part of that period, possibly still earlier than the estimated range –, showing a clear haplogroup turnover in the modern Levantine population, with new subclades R1b1a1a2a2-Z2103 potentially from early (IE-speaking) Balkan individuals, and R1a1a1b2-Z93 probably from eastern (Iranian) migrants.

**III.8.2.2. Other Balkan languages**

The language ancestral to Armenian is – like Phrygian – believed to have belonged to the peoples that came from the west and overran the Hittite empire in the 12th century BC (Beekes 2011). The language ancestral to Albanian, sometimes identified with Illyrian, might have also had its origin in the Balkans early during the west migration of Balkan Indo-European.

Both the Albanian and Armenian languages are spoken by modern populations where the majority of R1b-M343 subclades are R1b1a1a2-M269 and R1b1a1a2a-L23 lineages, which point to a resurgence of a Proto-Anatolian genetic component (together with European hunter-gatherer lineages) after the southern expansion of Yamna groups of R1b1a1a2a2-Z2103 lineages.

In the case of the Armenian highlands, there is ancestry levelling and genetic continuity in the Middle East region during the Neolithic and Chalcolithic (Lazaridis et al. 2016), including ancient mtDNA lineages, also partially during the Bronze Age and Iron Age, which may point to a late and heavily male-biased migration (Margaryan et al. 2017). This continuity has traditionally been explained by a history of genetic isolation from their surroundings (Haber et al. 2016).
The oldest male sample found in the region is of haplogroup R1b1-L278 (x R1b1a1a2-M269), dated ca. 2619-2465 BC, from the Kura-Araxes culture (Lazaridis et al. 2016)\(^6\), suggesting the presence of previous R1b1-L278 lineages in the region, probably from a Mesolithic migration – either from south-eastern Europe or from the Pontic-Caspian steppes –, unrelated to the later migration of Proto-Armenian speakers.

Populations of the western part of the Armenian Highland, Van, Turkey, and Lebanon show genetic affinity with European populations, and their absence in previous studies “should be considered a consequence of the absence in their Armenian datasets of populations from the western region of the Armenian highland” (Hovhannisyan et al. 2014). Ascertaining the origin of the Armenian population is hindered by the loss of data due to the effects of the Armenian Genocide.

**III.8.3. The expansion of North-West Indo-European in Europe**

**III.8.3.1. Bell Beaker culture**

The Bell Beaker phenomenon is defined by groups that show a common know-how in technology, especially regarding pottery, copper metallurgy (Amzallag 2009), and flint. No single unified network of know-how transmission can be reconstructed, only local or regional networks (Linden 2015). Despite this a supra-local homogeneity can be observed in the whole of Europe from 2500 BC “in similar funerary rituals, in the way of interacting with territory, in the way of representing iconography and decorating pottery, and in the way of representing social differences” (Martínez and Salanova 2015). The Bell Beaker phenomenon made thus the previous regional networks of Western Europe uniform with identical social codes.

With the advent of radiocarbon dating the compilation of Bell Beaker pottery dates (Müller and VanVilligen 2001) showed that the most likely origin of the pottery style was Iberia, pointing to high quality, tall beakers of the so-called maritime style. Only later were these dates and the Bell Beaker migrations put together in a common paradigm, when it was noted that the expansion of beakers with lower profiles and a more complex decoration, from East Group beakers, were replaced in the Danube area by plain jars, cups

---

\(^6\) Lazaridis, Twitter, 18 June 2016: "I1635 (Armenia_EBA) is R1b1-M415(xM269). We'll be sure to include in the revision. Thanks to the person who noticed! #ILovePreprints."
and plates. These vessels then dominated in the later developments (Harrison and Heyd 2007).

The migration of mobile Yamna migrants into the plains of the lower Danube and the central Carpathian basin is noted in small and large groups in the Balkans, establishing pastoral societies as forerunners of Heyd’s “Yamna package”, with domesticated horses, ox-drawn wagons, and herds of cattle and sheep, and noticed as far as southern and central Germany. The so-called “proto-Bell Beaker package” arises at the same time ca. 2900-2800 in Portugal, with a new culture of large fortified settlements, megalithic tombs and collective burials (see above Figure 13). This “proto-package” is found for example in the Maritime Beaker, and it expanded ca. 2700-2500 getting enriched through some areas in Western Europe (e.g. in the corded Beaker type), but clear internal social boundaries existed in this period. This Bell Beaker idea turned into the classical “Bell Beaker Package” during its expansion to the east, reaching southern France in the 26th c. BC, then arriving in central Europe, and the Csepel group of the Carpathian basin, around 2500 BC (Harrison and Heyd 2007).

The Bell Beaker migrations (ca. 2500-2100 BC) between Central Europe, the British Isles, and the Iberian Peninsula have long been associated (Gimbutas 1993) with the expansion over central and western Europe of Yamna migrants through the Vučedol complex in the Carpathian basin. Specific correspondences were found in burial rites, armament, costume, ornaments, technology in general, and also in ranked society, funerary rites, belief in life after death, and in general symbolism. According to Anthony (2007), Bell Beaker sites of Csepel around Budapest (ca. 2800-2600), west of the East Hungarian settlement region of Yamna migrants, could have been the direct bridge between Yamna on the east and Moravia and Bavaria to the west. Through this area, Yamna dialects spread to southern Germany, where decorated cup styles, domestic pot types, and grave dagger types from the Middle Danube were adopted ca. 2600 BC (Figure 20).

This small original region along the Danube, at the crossroads of the Yamna and proto-Bell Beaker “packages”, before its expansion as a Bell Beaker folk throughout Europe, has therefore strong potential as homeland of a common North-West Indo-European language that would later evolve into the majority of the known Indo-European branches of Europe (Mallory 2013).
Contemporaneous with these changes was the evolution noted in the East Group (in southern Germany, Moravia, and the upper Danube territories), into:

- ranked family-based social structures, rooted on self-sufficient farmsteads;
- a progressive specialization in stockbreeding and plant cultivation of less demanding species;
- with burials following family units, signalled by ‘founder’ graves;
- and without defensive position, hillforts, or fortifications – unlike later chiefdoms of the Bronze Age, where families and single persons gain power.

This structure allowed for individual and social mobility, increased communication and internal exchange of information, goods, genes, and social values (Heyd 2007).

Main R1b1a1a2a1a-L151 subclades R1b1a1a2a1a1-U106/S21/M405 split ca. 2800 BC with identical TMRCA, while R1b1a1a2a1a2-P312/S116 split at the same time, but shows a TMRCA ca. 2600, which points to a comparatively later successful expansion of its surviving subclades.

R1b1a1a2a1a2-P312 split later still into common west European subclades R1b1a1a2a1a2-DF27, R1b1a1a2a1a2b-U152, and R1b1a1a2a1a2c-L21, ca. 2400 BC, with the same date for TMRCA, which suggests an expansion coincident with the spread of Bell Beaker peoples to the west.

**III.8.3.1.1. Contacts between Bell Beaker and Corded Ware**

Settlement areas of both cultures, the Bell Beaker and the Corded Ware culture, especially in the common territories of central Europe, seemed to remain separated. Available data suggest rejection and aversion, but also some form of social discourse between the groups.

Neighbouring groups of Bell Beaker, Globular Amphora and Corded Ware cultures of east-central Europe show certain similar artefacts, but made of different materials, and with different interpretations, which might signal imitation among culturally different groups. These cultural differences between Corded Ware and Bell Beaker cultures are maintained over vast distances, from east to west Europe (Czebreszuk and Szmyt 2008), potentially suggesting a strong ethnolinguistic difference among groups that were also genetically (in ancestry and haplogroups) heterogeneous.

With the interaction of both groups, Corded Ware burials adapted to Bell Beaker customs, and a decline in Corded Ware remains is found in shared areas.
The pattern observed is of spatial separation followed by partial integration (dissolution of the spatial-cultural divide), suggesting a land capture by the expanding Bell Beaker culture, and also an ethnic dimension based on cultural expressions and physical anthropology (Heyd 2007). This separation is later observed clearly in the heirs of both cultures: the Danubian Early Bronze Age of Southern German groups, with a Bell Beaker foundation; the Únětice Early Bronze Age, on a Carpathian foundation; and the Mierzanowice/Nitra Early Bronze Age, with origins in the Corded Ware culture (Bertemes and Heyd 2002). Each of them shows a different ideological resolution to these interactions in the Late Copper Age, and the creation of new social identities.

The regional substrate for many eastern and northern Bell Beaker groups is in many cases formed by late Corded Ware culture groups – with some pottery types persisting in later times, and with individual burials being also used by later settlers. However, in western and southern Bell Beaker territory previous regional substrates do not herald the Bell Beaker groups, with newer settlements using locations different to Late Neolithic sites, and collective graves being reused or substituted by individual graves (Besse 2014).

Tumulus building was identified by Gimbutas as one of the main cultural manifestations of the Kurgan culture (and “Kurgan people”) that spread Indo-European languages during the Neolithic. The practice of mound-building– and single graves – is nevertheless so widespread in time and space that it is hard to associate it with one particular ethnic group (Harding 2011).
III.8.3.1.2. Bell Beaker expansion

Steppe ancestry has been recently found widespread in central European Bell Beaker individuals, but not in Iberian samples, therefore excluding a substantial contribution to central Europe from Iberian Beaker Complex-associated individuals (Olalde et al. 2017).
This further supports Heyd’s (2007, 2012) archaeological model, and contradicts the previous model of Bell Beaker population expansion out of Iberia.

In the lower Danube clear patterns of patrilocality and female exogamy have been found, apart from a continuing kinship relation in the transition of the Chalcolithic to the Bronze Age. There is evidence of continuing traditions from the Bell Beaker cultures to Early Bronze Age cultures in the region, with female mobility as a force for regional and supraregional communication and exchange (Knipper et al. 2017).

Bell Beaker samples from Hungary at overlapping dates show very different proportions (from 0% to 74%) of steppe ancestry (see Figure 21), a heterogeneity consistent with early stages of admixture between European farmers and Yamna migrants. Also, in southern France and Britain there is a greater affinity to the Iberian Early Neolithic farmers than to central European farmers (Olalde et al. 2017), which suggests a previous spread of migrants along the Atlantic coast, consistent with the interaction of Megalithic-building cultures – and potentially the spread of the proto-Beaker package – involving movements of people (Figure 22).
The oldest samples from the Bell Beaker culture are two individuals from Kromsdorf dated ca. 2550 BC, one of R1b1a1a2-M269 (xR1b1a1a2a1a1-U106), and the other of R1b-M343 (M269 unclear) lineages (Lee et al. 2012). The oldest samples of R1b1a1a2a1a2-P312 lineages are found in Osterhofen ca. 2540 BC, and in Sierentz – Les villas d’Aurèle ca. 2430 BC. The oldest sample of haplogroup R1b1a1a2a1a2b-U152 is found in Budapest ca. 2335 BC; of haplogroup R1b1a1a2a1a2a-DF27 in Quedlinburg ca. 2290 BC (Lazaridis et al. 2016); and of haplogroup R1b1a1a2a1a2c-L21 in Amesbury ca. 2290 BC (Olalde et al. 2017).

![Image](https://via.placeholder.com/150)

**Figure 22.** Image from Olalde et al. (2017). “Principal component analysis of 990 present-day West Eurasian individuals (grey dots), with previously published (pale yellow) and new ancient samples projected onto the first two principal components”. Original under a [CC-BY-NC 4.0 International license](https://creativecommons.org/licenses/by-nc/4.0/).

Regarding the arrival of Bell Beakers into Britain – marked by the appearance of steppe ancestry in aDNA samples –, it has been found that they are more closely related to lower

---

7 ZZ11+ equivalent (ancestral to DF27 and U152); DF27+? In 390k BAM file, but short, may actually belong to chromosomes 2 or 5. Additional information from Alex Williamson.
Rhine individuals, and not to the Atlantic façade of western Europe (Olalde et al. 2017). Demic diffusion of R1b1a1a2a1a2c1-L21 lineages (Figure 23) accompanying Bell Beaker expansion in the British Isles was already supported by ancient DNA analysis (Cassidy et al. 2016), and all samples of the British Isles have been reported as of R1b1a1a2a1a2c-L21 (or upper haplogroup). These data and the modern distribution of L21 subclades associated with the British Isles point more likely to a single successful migration of clans of R1b1a1a2a1a2c-L21 lineages into Britain. The contended Pre-Celtic Irish and Pictish substrates might have been therefore of Late Indo-European nature, imported by peoples of R1b1a1a2a1a2c1-L21 lineages.

Recent research already supported a considerable degree of mobility with little difference between male and female migration in Britain (Parker Pearson et al. 2016), and not an exchange of female marriage partners (Brodie 2001) or inter-cultural contact consolidation (Vander Linden 2007), as previously proposed.

![Figure 23. Modern distribution of haplogroup R1b1a1a2a1a2c1-L21. Adapted from Myres et al. (2011)](image-url)
Studies of ancient Indo-European hydronymy (Krahe 1964; Krahe 1949; Nicolaisen 1957) have revealed a quasi-uniform name-giving system for water courses that shows Indo-European water-words and suffixes following rules of Late Proto-Indo-European word formation (Adrados 1998). This points to an ancient wave of Late Indo-European speakers spread over Western and Central Europe before the Celtic and Germanic expansions, including the British Isles, the Italian and Iberian peninsulas.

The four certain samples of R1b1a1a2a1a2b-U152 lineage have been found in Bell Beaker territories from east to west Europe. The presence of a potential R1b1a1a2a1a2a-DF27 lineage in the central Bell Beaker group, and its subclade distribution in the modern west European population might be linked to this lineage’s expansion to the west and south during the Bell Beaker phenomenon (Figure 24).

The expansion of R1b1a1a2a1a-L151 lineages could then be linked to the first introduction of Indo-European languages in Western Europe (Cassidy et al. 2016), which could have left some traces of their presence in historic times. However, the later expansion of Celtic languages, and an apparent resurgence of the probably indigenous Proto-Iberian and Proto-Basque languages – possibly the descendant of the languages of early farmers, similar to Paleo-Sardinian (Terradas et al. 2014) – over an expanding Iberian subclade (R1b1a1a2a1a2b1a1-M167/SRY2627) of the R1b1a1a2a1a2a-DF27 lineage (Gunther et al. 2015) have left scarce data on the older situation.

Samples from northern and south-western Iberia in the transition from the Chalcolithic to the Bronze Age have been found to have steppe ancestry, although in lesser proportion than in central and western Europe (Olalde et al. 2017; Martiniano et al. 2017). Samples from north-east Iberia, dated ca. 2385 BC, show haplogroups R1b1a-L754 (xR1b1a1a2a-L23), G2-P287, and I2a2-M436, which point to at least a partial persistence of old European hunter-gatherer lineages in the region.

However, there is a clear replacement of autochthonous I2-M438 and G2a-P15 subclades by R1b1a1a2a1a2-P312 lineages. The expansion of haplogroup R1b1a1a2a1a2a-DF27 from northern Iberia, based on studies of the modern population, points to a time ca. 2200 BC (Solé-Morata et al. 2017). The greater admixture with the autochthonous population probably suggests a different type of expansion than that of Bell Beakers in central Europe and the British Isles, possibly a late, renewed male-driven invasion from northern Iberia, once admixed with the local population.
The only certain Indo-European language of Iberia that can be considered of a non-Celtic nature is Lusitanian (which has been linked to a potential Galaico-Lusitanian group of the north-western Iberian Peninsula), and there has been some discussion about the pre-Celtic nature of the languages of Cantabri, Astures, Pellendones, Carpetani, and Vettones. Also, while the position of Tartessian as Indo-European (Koch 2009) is highly doubted8 – , there is some support for a borrowing of names from a “lost Indo-European language” over the course of long-term contacts (Mikhailova 2015).

The emergence of El Argar groups was preceded by a break in Chalcolithic cultural traditions in south-east Iberia, which points to an upheaval of existing social structures or an influx of groups that cannot be distinguished from the local population at the present of genetic resolution, e.g. from south-eastern Europe (Szecsenyi-Nagy et al. 2017). This could point to the time of resurge of groups associated with previous Neolithic cultures that might have conserved Pre-Iberian and Pre-Basque languages until historic times.

8 It was criticized extensively in a special section of Vol. 42 of The Journal of Indo-European Studies (No. 3 & 4, Fall/Winter 2014)
The other region where modern R1b1a1a2a1a2a-DF27 lineages peak in the modern population corresponds to the old Nordwestblock cultural region, where a non-Celtic, non-Germanic Indo-European language might have been spoken (Kuhn, Hachmann, and Kossack 1986).

To the east, in the Vistula group of Lesser Poland, Bell Beaker samples of R1b1a1a2-M269 are found from 2400 BC to 2300 BC. Later an eastern sample is found near the Oder ca. 2170 BC (see below), all of which – together with long-term and long-distance economic exchange (especially regarding amber imports) during the Bronze Age (Makarowicz 2009) – supports the presence of Old European river names in east Europe dating to this period.

III.8.3.2. Únětice culture

The Únětice culture (ca. 2300-1700 BC) has been cited as a pan-European cultural phenomenon (Kristiansen and Larsson 2005), whose influence covered large areas due to intensive exchange (Pokutta 2013), with Únětice pottery and bronze artefacts found from Ireland to Scandinavia, the Italian Peninsula, and the Balkans (Figure 25). As such, it is candidate for a late community connecting a *continuum* of already scattered North-West Indo-European languages ancestral to Italic, Celtic, and Germanic, and perhaps to Balto-Slavic, where words were frequently exchanged, sharing a common lexicon and certain regional isoglosses (Gamkrelidze and Ivanov 1995). At the same time, strong phonetic differences found early in North-West Indo-European dialects, especially in the compounds with sonorants (Adrados, Bernabé, and Mendoza 2010; Clackson 2007), signal a period of already differentiated but inter-connected communities.

Thought to have evolved from Bell Beaker cultures, the scarce ancient Y-DNA available comes from four samples: one individual from Corded Ware/Proto-Únětice culture of R1b1a-L755 lineage at Łęki Małe ca. 2170 BC (Mathieson et al. 2017), and three from Únětice proper, of typical European hunter gatherer I2-M38 subclades – one ca. 2050 BC from Eulau, and two from Erperstedt dated ca. 2055 BC and 2035 BC (Mathieson et al. 2015). On the periphery of the Únětice culture territory, haplogroup R1b1a1a2-M269 is found in Gata/Wieselburg (ca. 1765 BC), and haplogroup R1b1a1a2a1a-L151 in Untermeitingen ca. 1605 (Allentoft et al. 2015).
Ancient DNA samples suggest at least a partial resurgence of hunter-gatherer ancestry in Únětice, although only a slightly lesser genetic affinity to Yamna than in Bell Beaker groups (Haak et al. 2015). Úněticean genetic melting pot strengthens its origin as the vector of cultural diffusion of North-West Indo-European languages, essentially connecting Barbed Wire Beaker cultures from the Low Countries and the Northern Lowlands (and late Nordic Neolithic) – probably speaking languages ancestral to Germanic – with peoples of Southern German cultures, as predecessors of core regions of the Tumulus culture – possibly speaking West Indo-European, i.e. pre-Italo-Celtic (Mallory 2013).

This suggests that Únětice connected these with eastern cultures like south-eastern European cultures – heirs of Bell Beaker and Carpathian groups – and the eastern Mierzanowice/Nitra culture – heir of Bell Beaker and Corded Ware groups. Therefore, the language ancestral to Balto-Slavic (or “Temematic”, see below) was probably spoken either by the Únětice population, or by eastern cultures that were connected to western Indo-European languages through Únětice.

Bell Beakers and early Únětice represented the first prospectors and metallurgists, travelling and sharing their skills, with Adlerberg and Straubing groups of the Southern German cultures being small local centres (Kristiansen 1987).
**III.8.3.3. Germanic**

In Scandinavia, farming communities had already abandoned their subsistence strategy for the development of transhumance (Jensen 2003). With the latest Middle Neolithic phase (ca. 2800-2400), Corded Ware/Battle Axe groups and influence appear in Norway, but the extent of their influence is unclear, and seems constrained to some limited “islands” or groups in the east, probably interacting with late Funnel Beaker and hunter-gatherer groups (Prescott 2012).

A migration of Bell Beaker groups to Jutland during the mid-3rd millennium BC seems to have brought skills in mining and sailing, introducing mass production of flint daggers, as well as the first metal daggers (an imitation of copper and bronze prototypes). This Dagger Period of the late Nordic Neolithic also represents the introduction of a more ranked social organization, and a new ideology, with a farm institution (long houses and fields), and an economy based on agro-pastoralism, integrating the diverse previous cultural traditions into a single south Scandinavian cultural sphere (Kristiansen 2009).

A massive transfer of knowledge, institutions and practices – which must have been accompanied by a movement of peoples – happened ca. 2400-2350 BC in southern and south-western Norway – probably disembarking at or near the harbour discovered in Slettabø – and occupied environments similar to Jutland. Settlers reached as far north as Mjeltehaugen, possibly as ‘scouts’ (Anthony 1990), maybe even as travelling metal prospectors, establishing a new elite on the north-western coast of Norway (Sand-Eriksen 2017).

The rapid adaptation of new practices to such hostile environments indicate the active participation of people with long traditions in the region. Such migratory movements “were probably related to the inherently expansive pastoral ideology, bolstered by a male warrior ideal, wanderlust, ideologically encouraged travelling/knowledge seeking, but also resource prospecting in a world rapidly embracing metallurgy and trade in exotica” (Prescott 2012).

The Bell Beaker period is the only reasonable candidate for the spread and final entrenchment of a common Indo-European language throughout Scandinavia (and not just Corded Ware core areas of southern and eastern Scandinavia), and particularly Norway (Prescott and Walderhaug 1995). The Dagger Period integrated the diverse cultural traditions of the previous period into a single south Scandinavian cultural sphere,
which suggests the formation of “a shared Nordic language based upon the frequent interaction that followed from the distribution of flint daggers” (Kristiansen 2009).

Such a direct strong cultural connection was possible thanks to communications through the strait of Skagerrak between northern Jutland and southern Norway. This, and the superior attractiveness of the Bell Beaker culture – apart from military expansiveness – provided the necessary unifying conditions and infrastructure for the expansion and linguistic unification of Scandinavia quickly during the following Late Neolithic and early Bronze Age period (Prescott 2012).

To a certain extent – taking the example mentioned by Prescott (2012) –, the previous Corded Ware and the following Bell Beaker cultures could be compared with explorations of Sparta and Athens: Sparta and Corded Ware represent terrestrial brutal force, while Athens and Bell Beaker represent maritime mobility and trade, and an appealing ideology.

The best candidate for an original homeland of the Pre-Germanic dialect of North-West Indo-European migrating into Scandinavia is the Beaker culture of the Low Countries and Northern Lowland (Kristiansen 2009).

Cultural groups of these regions show a period of evolution (Figure 26) starting ca. 1850 BC until their complete cultural change – evident after ca. 1500 BC (Fokkens and Harding 2013) – into the Elp culture (ca. 1800-800 BC). Samples of haplogroup R1b1a1a2a1a1-U106 are found quite late, in the Nordic Late Neolithic at Lilla Beddinge ca. 2150 BC (Allentoft et al. 2015), and in Oostwoud ca. 1881-1646 BC (Olalde et al. 2017), suggesting a connection of lineages between Jutland and the Low Countries. Modern population analysis supports this connection, showing that R1b1a1a2a1a1-U106 distribution peaks today precisely around the Netherlands.

Haplogroup I1-M253 split from the common stem at approximately the same time as I2-M438. The first example is found in Neolithic Linear Pottery culture in Hungary (Szecsényi-Nagy et al. 2015), which suggests its distribution in central Europe before the Corded Ware and Yamna expansions. The next sample found in aDNA records is from the Nordic Bronze Age in Angmollan, ca. 1400 BC (Allentoft et al. 2015).

R1b1a1a2-M269 lineages are found in early (Sebber Skole9, ca. 1410 BC) and late Nordic Bronze Age (Angmollan, ca. 670 BC), while haplogroup I1-M253 is found in Angmollan

---

9 Additional information xP312, xA2150 from Vince Tilroe.
(ca. 1400 BC), and haplogroup I-M170 in Angmollan (ca. 1360 BC) and Abekas (ca. 1255 BC), suggesting a south-north cline in culture and population in Scandinavia during these times of ethnolinguistic change.

The modern population with I1-M253 lineages is centred on northern Scandinavia near Skagerrak strait and Kattegat sea area, and shows a TMRCA ca. 2600 BC. On the other hand, R1a1a1b1a3-Z284/S221 lineage (formed ca. 2700 BC, TMRCA ca. 2300 BC) – probably associated with the expansion of the Corded Ware cultures in Scandinavia – shows a distribution in modern Scandinavian populations located further to the north and west of that zone.
It is difficult to ascertain whether I1-M253 lineages mixed with Bell Beaker lineages in Jutland or Norway before expanding into the rest of Scandinavia, or formed separated communities – maybe inhabiting Scandinavia even earlier than R1a1a1b1a3-Z284 lineages.

During the Bronze Age, at least two social spheres can be described: one of agriculture and husbandry related to the inland; and one of maritime nature, forming a decentralized social organization led by warriors focused on seaborne transport, trade, and raids, related to the sea and the coastal rocks (as well as to rock art and burial cairns). Unlike in the previous Neolithic period, certain parts of the elite invested in the maritime forces of production and long-distance trade (Earle et al. 2015; Ling, Cornell, and Kristiansen 2017).

Southern Scandinavia, with its participation in the larger European network of bronze exchange, may have acted as the periphery of European Bronze Age societies in contact with the Late Indo-European-speaking world. A more complex distinction may be more precisely made on a regional scale between southern, central and northern Scandinavia, with centre-periphery relationships based on the dynamics of regional cycles of production and alliances. Centres of dominant chiefdoms predominate in southern Scandinavia in Early and Late Bronze Age, while in northern Scandinavia inner and coastal groups do not conform to the same culture and subsistence, having probably an indirect relationship based on reciprocal exchange and ideological dominance (Kristiansen 1987).

After an obscure period of internal development (regarding lineages and admixture), the situation in Northern Germany and Scandinavia before the Iron Age probably corresponded loosely to the present situation. R1b1a1a2a1a1-U106 lineages probably occupied Jutland and the northern Scandinavian territory around the Skagerrak strait, while communities with I1-M253 lineages were confined to northern Scandinavia, given the quite late invasion of Jutland by Danes.

The irruption of Germanic peoples in central, east, and west Europe including the Roman Empire – the Barbarian Invasions from Classical sources, renamed the Migration Period since the Romantic era – suggests a R1b1a1a2a1a1-U106-dominated West Germanic area, and Viking migrations point to different clans belonging to R1b1a1a2a1a1-U106, I1-M253, and R1a1a1b1a3-Z284 lineages in the North Germanic area (see below Figure 33).
The modern distribution of R1b1a1a2a1a1-U106 (Figure 27) is roughly coincident with the expansion of West Germanic with the medieval Ostsiedlung, showing a west-east cline compatible with the Germanization of Slavs to the east of the Elbe. Although modern population samples are difficult to assess without genealogical information – due to the expulsion of Germans after World War II –, medieval samples from Podlažice (ca. 1180 AD) in Czech lands and Nicolaus Copernicus’ family origin from Koperniki near Nysa in Silesia before the 14th century (Bogdanowicz et al. 2009) seem to support the expansion of R1b1a1a2a1a1-U106 lineages associated with German settlers of the Holy Roman Empire east of the Elbe (see below Figure 42).

The question of the dialectal nature of East Germanic remains a purely linguistic one, but I1-M253 and R1a1a1b1a3-Z284 lineages scattered throughout Europe seem to support the classical description of East Germanic tribes migrating from Scandinavia to the east of the Elbe, and thus its connection with the Nordic branch.

Figure 27. Modern distribution of haplogroup R1b1a1a2a1a1-U106. Adapted from Myres et al. (2011).
III.8.3.4. Tumulus Culture

It was only after 2000 BC that large-scale mining operations and production which required specialized metallurgical and organizational know-how began in a few centres, and they reached distant regions as far as Northern Scandinavia. And only from 1750/1700 BC began the actual Pan-European tradition of metal work until its consolidation ca. 1600 BC, with different regions in Europe producing their own products, most specially the cultures of the Carpathian basin (Kristiansen and Larsson 2005).

The contacts of Únětice with Carpathian territories are constant, e.g. in the Únětice-Nitra and Únětice-Hatva horizons, where settlement micro-regions and relationships are difficult to assess. Únětice elites controlled trade routes from the Baltic Sea shores to Aegean Sea artisans. Úněticean daggers are found all over Europe and in Anatolia, and the nature of weapons and metal work suggest a chronic state of warfare and the emergence of a warrior class until its demise by the Tumulus culture, born in the area previously occupied by Únětice groups in Southern Germany.

Central European groups from southern Germany would then in this context correspond to a community with a common West Indo-European language ancestral to Italic and Celtic (Kortlandt 2007; Eska 2010), whose continuous development and dialectal evolution is to be followed into the Tumulus culture (ca. 1600-1200 BC). The Tumulus culture was eminently a warrior society which expanded with new chiefdoms eastward into the Carpathian Basin (up to the river Tisza), and northward into Polish and central European and Únětice territories, with dispersed settlements centred on fortified structures (see above Figure 26).

In the subsequent period of crisis, it developed into bands of raiders and mercenaries, and took control of peasant societies, as happened in several regions during the Urnfield and La Tène periods, and similar to the society of mercenaries and warring city states in the Celtic period (Kristiansen 2000). The European world ca. 1450-1100 BC has been compared to the Viking Age, with population pressure and lack of land for young sons with no inheritance leading to war-bands that engage in seasonal raiding, trading, and piracy. This was followed up by more massive colonising ventures and migrations, and a political economy based on a chiefdom form of society where free farmers were the dominant class, with commoners and slaves as dependent groups (Kristiansen 2016; Heyd 2007).
Exogamous and endogamous strategies and variable distances of marriage exchanges to maintain alliances complicate this picture further in the Tumulus/Nordic traditions (Kristiansen 2000), which suggests that post-Únětice societies could especially benefit from the investigation of human ancestry.

The Urnfield culture (ca. 1300-750 BC) is associated with the rise of a new warrior elite, and the formation of new farming settlements and their urnfields. In some areas there is continuity from Tumulus to Urnfield culture, with narrowing and concentration of settlements along the river valleys, but there is also wide-ranging migrations (Figure 28). These migrations are similar to those seen later in the La Tène culture (Kristiansen 2000). Urnfield migrations south of the Pyrenees may have brought the pre-Celtic Sorothaptic language believed to be behind certain toponyms and inscriptions around the Pyrenees (Coromines 1976).

Scarce aDNA from late Urnfield populations from its north-eastern territories in Saxony – near the Lusatian culture –, show a mixture of lineages, which suggest genetic continuity with older cultures: R1a1a1b1a-Z282 lineage was found in Halberstadt (ca. 1085 BC), and of the eight males studied from the Lichtenstein cave (ca. 1000 BC), five were of haplogroup I2a2b-L38, two of haplogroup R1a1-M459\(^\text{10}\), and one of haplogroup R1b-M343 (Schilz 2006).

Given the modern distribution of R1b1a1a2a1a2b-U152 lineages (see below Figure 29), its expansion is probably to be connected to the spread of the Urnfield culture and later offshoots Hallstatt and Villanovan cultures.

\(^{10}\) Reported as Ri and Ri?
III.8.3.4.1. Celtic

From the early Urnfield culture expanded the Hallstatt culture (ca. 1200-500 BC), associated with Proto-Celtic (Chadwick 1970). The later expansion of La Tène culture (ca. 500-1st c. BC) from certain core Hallstatt regions – valleys of Marne and Moselle and neighbouring Rhineland in the west, and a Moravian zone in the east – has been linked to the spread of Celtic languages (Figure 30). However, the Mainz research project of bio-archaeometric identification of mobility has not proven to date a mass migration of Celtic peoples in central Europe ca. 4th-3rd centuries BC, i.e. precisely in a period where textual evidence informs of large migratory movements (Scheeres 2014). La Tène material culture points to far-reaching inter-regional contacts and cultural transfers (Burmeister 2016).

R1b1a1a2a1a2b-U152 lineages are found today (Figure 29), scattered to the north, south, and west of the Alps, reaching the southwest corner of the Iberian Peninsula, and the British Isles. It seems to peak around the current borders between Italy, France, and Switzerland. The first sample of haplogroup R1b1a1a2a1a2b-U152 found in the British Isles (contemporaneous with the first sample found of haplogroup R1b1a1a2a1a1-U106) is from Driffield Terrace ca. 250 AD, after the Roman invasion (Martiniano et al. 2016).

The expansion of I2a2a-M223 lineages, assumed to be already mixed with R1b1a1a2a1a2b-U152 lineages since at least the Tumulus culture, is found from the British isles to Anatolia. I2a2a1a1-M284 lineages concentrated in Great Britain (with mutational divergence suggesting its foundation ca. 300 BC) provide “some tentative evidence of ancient flow with eastern areas that could support the idea that the La Tène culture was accompanied by some migration” (McEvoy and Bradley 2010).

The scarcity of R1b1a1a2a1a2b-U152 and I2a2a-M223 lineages in the modern populations of the British Isles and Iberia – where Celtic languages had clearly spread by the time of the Roman invasion – appear to suggest a successful cultural diffusion of the language from warring Celtic minorities who established new chiefdoms throughout Europe. However, it is also possible that a previous admixture of R1b1a1a2a1a2c1-L21 and R1b1a1a2a1a2a-DF27 lineages in the expanding Celtic population further confounds the genetic change associated with the Celtic expansion.

11 Further information on SNPs by Richard Rocca and Alex Williamson.
Lacking ancient DNA samples and more complex genetic analyses, it may be assumed from the available data that the sociocultural phenomenon associated with the expansion of La Tène culture (and Celtic-speaking peoples) is different from the Neolithic expansion of farming – where Anatolian ancestry spread slowly with technology –, and from the Bronze Age expansion of herding – where male-dominated groups spread rapidly into western Europe and partially replaced or displaced the original population.

Figure 29. Modern distribution of haplogroup R1b1a1a2a1a2b-U152. Adapted from Myres et al. (2011).
III.8.3.4.2. Italic

A spread of early eastern Urnfield cultures from Transdanubia is attested in the Po Valley (Váczi 2013).

The Villanovan culture (ca. 1100-700 BC), expanded from early Urnfield, has long been associated with Proto-Italic (Gimbutas 1965). However, the association of Villanovan culture with Italic languages remains controversial, since Villanovan territory is partially coincident with the later Etruscan-speaking zone, and no clear cultural break is seen between both cultures. Nevertheless, a resurgence of a previous language – akin to the example of Proto-Basque and Iberian languages (see above) – might explain the cultural continuity in Etruria.

Genetic analysis of the modern population show a spread of R1b1a1a2a1a2b-U152 lineages south of the Alps, including north and central Italy, which supports the invasion of this group from the north, through the Alps. However, as with Greece – and even more so – the complexity of the current Y-DNA maps of the region mirrors the Italian Peninsula’s multiple genetic flows since the Palaeolithic (Sarno et al. 2017).

Classical sources since the Aeneid – Virgil’s epic poem linking the foundation of Rome to the flight of Aeneas from the Troy of Homer’s Iliad – have possibly influenced the traditional rejection by famous Italian linguists of the unity between Latin and Osco-Umbrian. While both branches share common innovations, and it is therefore difficult today to reject a shared community by relating all differences to recent contacts, some linguists have tried to reconcile the obvious Italic nature of Latin and its morphological differences compared with Osco-Umbrian with a potential late Anatolian substratum, and have thus supported an eastern invasion through Apulia.

To further complicate the linguistic and archaeological discussion around Latin, there are as of yet unexplained genetic flows between Anatolia and Italian samples since the Chalcolithic unrelated to the known earlier Neolithic expansion of Middle Eastern farmers, which adds to the controversy surrounding the origin of Etruscan (see above Anatolian).

The expansion of Rome (Figure 31) seems not to have been accompanied by a massive migration of peoples, and cultural diffusion is likely to have played a bigger role in the expansion of Latin.
Figure 31. Diachronic map of migrations in Europe ca. 250 BC – 250 AD. Continues on the next page. See full high-resolution version at <https://indo-european.eu/en/maps/classical-antiquity/>.
III.8.3.4.3. Italo-Celto-Germanic and Celto-Germanic contacts

The initial phase of the Elp culture (1800-1200) in the Low Countries is characterized by tumuli related to the Tumulus culture and later to the Urnfield culture. Dutch-German lowland areas share cultural roots with the southern Scandinavian area (Butler, Arnoldussen, and Steegstra 2011/2012) which predate technologic and economic exchanges between Urnfield and Northern Bronze Age Scandinavia (Kristiansen and Suchowska-Ducke 2015).

These complicated cultural-economic networks (see above Figure 26 and Figure 28) that preclude precise ethnic (and thus linguistic) differentiation, supports the maintenance of late contacts between the languages ancestral to Germanic and Celtic, assuming a position of Proto-Celtic to the north of the Hallstatt culture – as supported by the known homelands of the La Tène culture.

Lacking aDNA samples to obtain admixture analysis, careful investigation of I2a2a-M223 lineages – found today distributed among Germanic and Italo-Celtic territories – might bring light to population movements and exchanges during the Bronze Age and the Iron Age in Europe.
III.9. Indo-European in Corded Ware societies

III.9.1. Balto-Slavic

As already stated, samples from late Corded Ware cultures to the east of Únětice show a continuity of R1a1a1-M417 (most likely R1a1a1b-Z645) lineages in Bronze Age and in modern population samples (Figure 32).

The expansion of R1a1a1b1a-Z282 lineages appears therefore to be strongly linked to the spread of the Corded Ware culture, but the original homeland of these lineages is unclear. Based only on current genetic mapping (Underhill et al. 2015):

- basal R1a1a1b-Z645 seems to be distributed following a westward and eastward expansion from a Pit-Comb Ware ancestral homeland;
- R1a1a1b1a3-Z284 seems to have expanded early to Scandinavia and expanded later from a secondary nucleus there (given its late TMRCA);
- R1a1a1b1a1-M458 appears in Central Europe, with diffusion to the East;
- R1a1a1b1a2b-CTS1211 (or M558) seems to be centred on Eastern Europe.

Investigation of ancient DNA (see above Forest Zone and Corded Ware culture) suggest an original spread of EHG ancestry westward from the late Pit-Comb ware culture (Mathieson et al. 2017), and a later expansion westward expansion of steppe ancestry.
associated with Corded Ware cultures (Mathieson et al. 2015; Haak et al. 2015; Allentoft et al. 2015).

The current distribution and older TMRCA of R1a1a1b1a1-M458 compared to the other R1a1a1-M417 lineages could support its position as the original Pre-Balto-Slavic-speaking population. Some late Corded Ware groups in central Europe are thought to have smoothly transitioned to Bell Beaker cultures (Besse 2014), and some of these formed proto-Únětice and Mierzanowice/Nitra groups. Some Úněticean groups later evolved into early Lusatian Tumulus culture (ca. 1700-1400), originally located between the Elbe and Oder basins (see above Figure 26), which later expanded east (ca. 1300-500 BC) into territories of previous Trzciniec culture (see above Figure 28).

Diffusion of West Indo-European isoglosses has already been proposed to be identified with the Úněticean expansion into peoples of mixed ancestry and lineages, and continuity of such admixture from the region of Tumulus/early Lusatian into Urnfield/Lusatian cultures is supported by findings of R1a1a1b1a-Z282 lineages in the Urnfield group from Saxony-Anhalt, close to the proto-Lusatian original territory. That points to the Elbe basin as the original site of cultural breach for R1a1a1b1a-Z282 lineages, between the older Corded Ware tradition and the new Úněticean culture and language.

The Pomeranian and related West-Baltic culture of cairns (ca. 650-150 BC) evolved from the Lusatian culture to the east, following the expansion of the Jastorf and Hallstatt/La Tène cultures (see above Figure 30). Under pressure from Germanic migrations to the south and east from Scandinavia and the German lowlands, represented by Oksywie (2nd c. BC – 1st c. AD) and later Wielbark (1st c. AD – 4th c. AD) cultures in eastern Pomerania.

The Przeworsk culture (3rd c. BC – 5th c. AD) shows continuity in its roots with the preceding Pomeranian culture, but its extension north from the Vistula to the Oder, and south toward the middle Danube from the Dniester to the Tisza valley was accompanied by significant influences from La Tène and Jastorf cultures. The subsequent absorption into the Wielbark culture – related to the East Germanic expansion – make its precise association controversial, and it is sometimes viewed as an amalgam of a series of localized cultures (see above Figure 31).

East of the main Przeworsk zone was the Zarubints culture (3rd c. BC – 2nd c. AD), considered a part of the Przeworsk complex (Mallory and Adams 1997), located between the upper and middle Dnieper and Pripyat rivers. Early Slavic hydronyms are found in
the area, and the prototypical examples of Prague-type pottery later originated there (Curta 2001). It is therefore to be identified as Proto-Slavic (Kobyliński 2005).

Zarubinets came to an end with the migration of its population, linked to the increasingly arid climate. By the 3rd century western parts of Zarubinets had been integrated into the Wielbark culture, and some Zarubinets groups had moved southward into river valleys, moving closer to Sarmatian and Thracian-Celtic groups of the Don region and forming the Chernoles culture. Central late Zarubinets sites gradually turned into the Kiev culture (ca. 3rd-5th c.), widely considered the first identifiable Slavic archaeological culture, from which Prague-Penkov-Kolochin complex of cultures – identified with the expansion of Proto-Slavic (Mallory and Adams 1997) – descended about the 5th c. (Figure 33).

Regarding the conflicting nature of Balto-Slavic, for which a common group with Albanian and Indo-Iranian has been proposed (Kortlandt 2016), it has been hypothesized that the North-West Indo-European language behind Pre-Balto-Slavic – called “Temematic” (Holzer 1989) – would have formed the Pre-Balto-Slavic (especially Pre-Slavic) substratum language, over which a Graeco-Aryan (specifically Indo-Iranian-related) dialect would form its superstratum. However, such differences of Balto-Slavic with North-West Indo-European languages have been disputed (Matasović 2014). The likely Proto-Slavic original territory over layers of previous Cimmerian and Seytho-Sarmatian cultures seem to support a quite recent connection of Slavic and Indo-Iranian (and more precisely Iranian) peoples and their languages.

The division of historical Slavic tribes in territories and cultures in the 5th-7th centuries remains a hotly debated topic (Curta 2001). Nevertheless, the expansion of the Prague-Korchak culture from its nucleus in the older Przeworsk-Zarubinets contact zone to the west – including its expansion as the Sukow-Dziedzice group to the Baltic Sea – can be identified with the West Slavic expansion, and – at least part of – the western peak of R1a1a1b1a1-M458 lineages in modern populations.

The Kolochin material culture was a transformation of the old Kiev culture (Kobyliński 2005), but evidence of Baltic river names in the region have made some propose an original Proto-Baltic population (Mallory and Adams 1997) before the East Slavic occupation. Indeed the Baltic populations have been found to be genetically the closest to East Slavs (Kushniarevich et al. 2015), which is compatible with Baltic-speaking R1a1a1b1a2b-CTS1211 lineages undergoing a cultural assimilation with the East Slavic
expansion. A precise analysis of Finno-Ugric and Baltic populations would be necessary to discern which R1a1a1b-Z645 subclades were associated with which population migrations and expansions.

The expansion of the Penkov culture in the Danube seems related to the expansion of South Slavic. Confusing accounts of the Byzantine Empire of the raids and migrations of a federation of tribes (the Antes and the Sklavenes) in their frontiers give a general idea of the complex interaction of different groups in the Balkans (Curta 2001). This might justify a late assimilation of the language by groups of I2a2a1b-L701 lineages, which are prevalent today in South Slavic territory (Kushniarevich et al. 2015). However, apart from the main peak of haplogroup I2a2a1b-L701 in the south-east Balkan territory, a secondary peak around Bessarabia, as well as its general distribution around the same territory as the Prague-Penkov-Kolochin complex probably point to an earlier assimilation of the group, during the transition to a Proto-Slavic community and before its migration.

The eastern and western peaks in R1a1a1b1a1-M458 lineages might support a west-east migration, as well as an east-west migration, and indeed both in different periods, which is expected to be found if Lusatian is linked to the expansion of Pre-Balto-Slavic, and later younger subclades are linked to the West Slavic expansion to the west.
Figure 33. Diachronic map of migrations in Europe ca. 250-750 AD. Continues on the next page. See full high-resolution version at <https://indo-european.eu/en/maps/antiquity/>.
III.9.2. Indo-Iranian

During the western expansion of Yamna herders in the Bronze Age, the Fatyanovo group emerged early at the north-eastern edge of the Middle Dnieper group, still showing mixed Corded Ware / Globular Amphora traits, substituting the Volosovo culture and occupying the Volga-Kama region. Near it the Balanovo group seems to have been its metallurgical heartland (Anthony 2007).

In the forest-steppe zone of the middle Volga and upper Don, at the easternmost aspect of the Russian forest-zone, the last cultures descended from Corded Ware ceramic tradition, the Abashevo group, emerged ca. 2500 BC or later (Anthony 2007), substituting the late Volosovo culture, and reaching the Upper Ural basin. Abashevo showed a mix of Fatyanovo/Balanovo and Catacomb/Poltavka culture (Figure 34).

Genetic make-up of modern populations show a distribution of basal R1a1a1b1a-Z282* lineages centred on the old territory of Middle-Dnieper – Fatyanovo – Abashevo groups (Underhill et al. 2015), and a sample of haplogroup R1a1alb-Z645 is found later in the Potapovka culture, in Utyevka ca. 2200-1900 BC (Mathieson et al. 2015).

Figure 34. Diachronic map of migrations in Asia ca. 2600-2250 BC. See full high-resolution version at <https://indo-european.eu/maps/chalcolithic/>
Early Yamna continued in the Lower Don – North Caucasian steppe as the Catacomb culture, and in the Volga-Ural region as the Poltavka culture, where human ancestry and cultural continuity implies that eastern languages from the Graeco-Aryan continuum – already separated from the western Paleo-Balkan group – were spoken, i.e. the ancestor of Proto-Indo-Iranian. Herders from the Poltavka culture began to move to the Ural-Tobol steppes probably about 2800-2600 BC. Coinciding with more arid climate after ca. 2500 BC, both Poltavka and Abashevo herders settled between the Tobol and Ural River valleys.

The Poltavka outlier of R1a1a1b2-Z93 lineage found in Potapovka, in the Samara region, dated ca. 2710 BC (Mathieson et al. 2015), clusters closely with samples from the Corded Ware and from later Sintashta and Potapovka cultures. Unlike other samples of R1b1a1a2a2-Z2103 lineage found in the same area, this grave was most likely established on top of an older Poltavka cemetery in the Middle Bronze Age, where a Sintashta culture cemetery was later found (Mathieson 2015, Supplementary materials). The early date, only slightly later to its haplogroup formation, points to a period of population expansion, and probably also to intense early regional contacts between peoples of Abashevo and Poltavka cultures.

Cultures that emerged around 2100-1800 BC in the region – Sintashta in the Ural-Tobol steppes, and Potapovka in previous Poltavka territory – seemed to continue in an early phase the previous Abashevo tradition, but retained and gradually expanded many cultural traits of Poltavka pottery, followed the same burial rites, and settled on top of or incorporated older Poltavka settlements (Figure 35). “It is difficult to imagine that this was accidental. A symbolic connection with old Poltavka clans must have guided these choices” (Anthony 2007). Warring groups were strong enough to take and destroy an entire settlement, signalling an age of fully-fledged conflict, with a succession of changes in the defence systems and planning schemes of the settlements.

Both Sintashta and Potapovka were born from a time of escalating conflict and competition between rival tribal groups in the northern steppes, where raiding must have been endemic, and intensified fighting led to the invention of the light chariot (Anthony 2007). The state of intense warfare was caused by a constant flow of wealth, originating from long-distance metal trade, with formation and destruction of alliances and gathering of large groups of warriors, which created a vicious circle of escalation of conflict, and created new customs, new tactics, and new weapons (Pinheiro 2011).
Figure 35. Diachronic map of migrations in Asia ca. 2250-1750 BC. See full high-resolution version at <https://indo-european.eu/en/maps/early-bronze-age/>.

Ancient DNA samples of Sintashta, Potapovka show a common ancestry which clusters closely with Corded Ware samples (see above Figure 12), and a substitution of R1b1a1a2a2-Z2103 lineages by R1a1a1b-Z645 lineages (Allentoft et al. 2015). However, cultural continuity with Poltavka is not only seen archaeologically in material and symbolic culture, but is also evident from the association of the Sintashta expansion with Andronovo, and therefore with the later expansion of Indo-Iranian peoples and their languages. The most likely explanation for the eastern expansion of Indo-Iranian by peoples with R1a1a1b2-Z93 lineage is therefore the assimilation by Sintashta-Petrovka groups of the Proto-Indo-Iranian language spoken by Poltavka herders.

The process by which this cultural assimilation happened, given the presupposed warring nature of their contacts, remains unclear. It is conceivable, in a region of highly fortified settlements, to think about alliances of different groups against each other, akin to the situation found in Bronze Age Europe: a minority of Abashevo chiefs and their families would dominate over certain fortified settlements and wage war against other, neighbouring tribes. After a certain number of generations, the most successful settlements would have replaced the paternal lineages of the region – with only a slight
drift to steppe admixture observed in PCA compared to Corded Ware –, while the majority of the population in these settlements – including females, commoners and slaves – retained the original Poltavka culture. R1b1a1a2a2-Z2103 lineages were mostly replaced in the region by haplogroup R1a1a1b2-Z93, as demonstrated by the later expansion of its subclades with Andronovo and Srubna cultures, and by present-day distribution of R1a1a1b2-Z93 lineages in Eurasia.

The language spoken by peoples of the Srubna (“Timber Grave”) culture of ca. 1800-1200 BC – heirs of the Pokrovka complex (ca. 1900-1750 BC) created by Potapovka and late Abashevo groups – was probably an Indo-European language of the Graeco-Aryan group. Their repeated violation of the canid-eating taboo across generations point to the connection of dogs with war-bands and with the Indo-European myth of dogs guarding the entrance to the afterlife (Anthony and Brown 2017).

A paleoecological crisis had a significant effect on the economy of the tribes in Late Catacomb and Post-Catacomb times, favouring a higher mobility and transition to nomadic cattle breeding (Demkina et al. 2017). This crisis might have allowed for the westward expansion of eastern R1a1a1b2-Z93 lineages associated with the Srubna culture,

Figure 36. Diachronic map of migrations in Asia ca. 1750-1250 BC. See full high-resolution version at <https://indo-european.eu/en/maps/middle-bronze-age/>.
which replaced the Catacomb culture in the Pontic-Caspian steppe (Figure 36). The further integration of steppe population might account for the slightly increased drift toward steppe ancestry found in this culture, compared to Potapovka and Sintashta cultures.

Samples from the Balkans during the Middle Bronze Age (later than ca. 1700 BC) show an increase in steppe ancestry compared to earlier samples from the Balkans (Mathieson et al. 2017). Together with the finding of an individual of haplogroup R1a1a1b2-Z93 in the Balkans, this further supports that the westward expansion of the Srubna culture from the North Caspian steppe was associated with an homogenization of steppe lineages to R1a1a1b2-Z93 subclades.

Later Cimmerian or Thraco-Cimmerian groups might have emerged from societies related to these expanding western groups of Pontic-Caspian herders, who might have further increased steppe ancestry in the Balkans. Their relationship to Scytho-Sarmatian groups later migrated from south Asia is unclear.

A comprehensive description of Sintashta-Petrovka expansion eastward as part of the Andronovo horizon in Asia – coinciding with the western expansion of the Seima-Turbino phenomenon to the Forest Zone – is given by Anthony (2007). Chariots were probably invented in the steppes, improving warfare and likely playing a big role in Indo-Iranian expansion within the Andronovo horizon (after ca. 1900 BC) and south from the Zeravshan valley into the Bactria-Margiana Archaeological Complex (after ca. 1800 BC), creating between 1800-1600 BC a post-BMAC culture dominated by Tazabagyab-Andronovo herders. The drift toward steppe ancestry found in Andronovo samples point to an admixture with a previous population from the steppe, potentially related to east Yamna and Afanasevo migrants in the area.

In the Murghab region, a change is seen in access to water – essential for life in the region – in the transition from the Late Bronze Age to the Iron Age (Yaz I), evidence of a political-economic system that was shifting toward territorial management. This period is characterized by a combination of sand encroachment from the north, shifts in known watercourses and a possible decrease in flows, as well as the invasion of people represented by ‘steppe’ pottery, with declines and abandonment of major population centres (Rouse and Cerasetti 2016).
After about 1600 BC pastoral economies spread across Iran and into Baluchistan, and ca. 1500 BC Indo-Aryan chariot warriors invaded a Hurrian-speaking kingdom of Mitanni in north Syria. At the same time, as post-BMAC herders spread to the northeast Indian subcontinent, the Rig Veda was probably being composed (see Figure 36 to Figure 40).
Figure 39. Diachronic map of migrations in Asia ca. 250 BC - 250 AD. See full high-resolution version at <https://indo-european.eu/en/maps/classical-antiquity/>.

Figure 40. Diachronic map of migrations in Asia ca. 250-750 AD. See full high-resolution version at <https://indo-european.eu/en/maps/antiquity/>.
The modern distribution of R1a1a1b2-Z93 lineages (Figure 41) shows a clear division between western and eastern subclades – with basal R1a1a1b2-Z93 located east of the Andronovo horizon (Underhill et al. 2015). Whereas the western R1a1a1b2a1-L657.1 subclade has an expected peak in the northern part of the Indian subcontinent – broadly coincident with the spread of Proto-Indo-Aryan and Indo-Aryan languages –, the eastern R1a1a1b2a2-Z2124 subclade peaks at the core of the Proto-Iranian Yaz culture and East Iranian expansion (of languages related to old Bactrians, Sogdians, and Scytho-Sarmatian peoples). Its spread west of the Iranian Plateau, however, is complicated by this region’s condition of place of transit of innumerable cultures and peoples in prehistoric and historic times – as is the case with the genetic make-up of southern Italian and Balkan peninsulas.

III.9.2.1. Ancient North Indian and Ancient South Indian ancestry

Middle Eastern farmers show an ancestry from a Basal Eurasian lineage that diverged from the ancestors of north Eurasian hunter-gatherers and East Asians before they separated from each other.
Two main different ancient populations are ancestral to modern Indians: the Ancestral North Indian (ANI) ancestry, a west Eurasian ancestry, and the Ancestral South Indian (ASI) ancestry. None of them exists in unmixed form, and both contribute a variable amount of the ancestry of South Asians (Reich et al. 2009).

ANI can be modelled as a mix of ancestry related to both western Iranian farmers and people from the Bronze Age Eurasian steppe, and is thus close to Middle Easterners, Central Asians, and Europeans (Lazaridis et al. 2016). Because of that, it is possible that the ANI component was prevalent in the Mehrgarh and later Indus Valley Civilization, expanding eastward into South India in an admixture event associated with the spread of farming, as suggested by mtDNA lineages that entered India from Anatolia, the Caucasus and Iran in the earliest Neolithic sites. The ASI component may have expanded earlier, possibly in different waves, from west and south-west Asia, with the end of the last Ice Age (Silva et al. 2017).

The impact of the steppe invasion in South Asia is heterogeneous. A cline of ANI-ASI ancestry is evident, with ANI ancestry significantly higher in Indo-European than Dravidian speakers. This suggests that Dravidian may have accompanied the expansion of ancestral ASI before mixing with the ANI.

While maternal lineages reflect earlier, pre-Holocene processes (with higher levels of autochthonous variation), paternal lineages predominantly show admixture events of the last 10,000 years, with a particularly strong male-driven invasion from central Asia during the Bronze Age (Silva et al. 2017), also reflected in the significant correlation of ANI with the Y chromosome (Reich et al. 2009). This sex-biased pattern is much less marked in the East Asian fraction, mainly focused on speakers of Tibeto-Burman and Austroasiatic language families.

Among groups, the Mala (a south Indian population) show minimal ANI and ca. 18% steppe ancestry, and the Kalash of Pakistan are inferred to have ca. 50%, similar to northern Europeans (Lazaridis et al. 2016). Regionally the North-West Frontier populations are thus more similar to European and Central Asian populations. Among ethnolinguistic and social groups, there seems to be an influence of north Eurasian admixture (potentially from ancient Indo-European-speaking populations) in forward castes, diminishing in backward castes and Dravidian peoples (Bose et al. 2017).
III.9.3. A common Corded Ware substrate

It has been argued that similarities found in Indo-Iranian and Balto-Slavic languages – like the peculiar phonetic *ruki* development, a similar satem trend in both groups (Meier-Brügger 2003) – suggest a sort of west-east *continuum* between both languages, with certain features running through them (Mallory and Adams 2007).

Since both Únětice (ca. 2300-1600) and Sintashta (ca. 2100-1800 BC) potential language expansions into populations of Corded Ware lineage happened at the same time, it could be argued that both communities happened to speak similar dialects that could have influenced both languages – a North-West Indo-European and a Graeco-Aryan dialect already developed quite differently – in a similar manner. Therefore, a common Corded Ware substrate language – whether non-Indo-European, Pre-Indo-European, or even Indo-European – could explain their similarities.

It has been classically proposed that the Mesolithic language of the Narva and Combed Pit Ware cultures is to be identified with a Uralic community, and dates ca. 4000 BC have been proposed for the common reconstructible Proto-Uralic language (Parpola 2012; Kortlandt 2002). Finno-Ugric has also been shown to have developed in close contact with Proto-Indo-Iranian (Kallio 2002).

According to the theory presented in this paper, the R1a1a1-M417 population of the Combed Pit Ware culture expanded to the east, and then from the Contact Zone – mostly as R1a1a1b-Z645 lineages – with the Corded Ware culture to west and east Europe, so it is possible that their language was indeed Proto-Uralic. On the other hand, if a common Indo-Uralic language is proposed, the division between Middle Proto-Indo-European and Proto-Uralic homelands might be tentatively traced back to the division between Neolithic horse riders of early Khvalynsk (strongly influenced by CHG ancestry) and North Pontic Mesolithic populations.

Therefore, while Wiik’s concept of a Mesolithic European ‘Vasconic-Uralic harmony’ (Wiik 2008) is not tenable today, the proposed Uralic substratum for Germanic and Balto-Slavic may be (Wiik 1999).

From a linguistic point of view, the characteristic palatalization of the consonant system in Proto-Uralic is compatible with the similarly transposed velar system adopted for Late Indo-European dialects by Balto-Slavic and Indo-Iranian speakers, thus explaining the strongest phonetic connection between these dialectally diverse Indo-European languages.
Differences in the Baltic and Slavic satemization processes might point to an early split of the North-West Indo-European dialect ancestral to both, before or during its assimilation by different Uralic-speaking communities of late Corded Ware cultures. A potential connection with the Balkans Chalcolithic, the origin of the Corded Ware horizon, could also explain the potential satem influence found in Anatolian and Paleo-Balkan languages.

This model supports thus the reconstruction of two series of velars: the traditional reconstruction of dorsovelars and labiovelars (Lehmann 1952), which is usually ignored in common textbooks in favour of the older reconstruction of a third series of palatovelars (Bomhard 2015); and Martinet’s glottalic consonants (Gamkrelidze and Ivanov 1995).

A western Corded Ware substratum could also be argued to be the origin of certain common isoglosses found between Germanic and Balto-Slavic. In terms of the “Temematic hypothesis”, Germanic and Temematic would share common western Corded Ware isoglosses, and only later would Proto-Balto-Slavic – already separated from Proto-Indo-Iranian – absorb Temematic as a substratum language (Kortlandt 2016).

To further complicate the dialectal nature of Balto-Slavic, ancient samples show R1b1a1a2a2-Z2103 lineages in western Yamna migrants, as attested in Vučedol and in east Bell Beaker populations (see above). Also, modern populations in central Europe, in regions previously occupied by the Únětice and Lusatian cultures, also show R1b1a1a2a2-Z2103 subclades. Assuming that Yamna lineages corresponded to separated clans that kept a Graeco-Aryan dialect during the western migration, their integration into a common Únětice culture could also explain the Graeco-Aryan features of Balto-Slavic that have been associated with Indo-Iranian.

Balto-Slavic and Indo-Iranian share a special position among Indo-European languages regarding their rather conservative nominal case system. It has been argued that languages with more second language speakers lose nominal cases (Bentz et al. 2015), which would explain the higher simplification of Late Indo-European dialects in west and south-east Europe, compared with the conservation of the original system by speakers of Uralic dialects, known for their large set of grammatical cases.

On the other hand, this could also give support to the theory that Late Proto-Indo-European had in fact a simpler nominal system, derived from a still simpler one of Middle Proto-Indo-European (Adrados, Bernabé, and Mendoza 2016). In this case, Indo-Iranian
and Balto-Slavic morphological differences would be later innovations; however, that would need an explanation as to how Uralic speakers adopting Late Proto-Indo-European added complexity to the language, instead of simplifying it.

The study of precise isoglosses connecting these languages, and their potential relation to specific Uralic proto-languages lies beyond the scope of this paper.

**III.9.3.1. Agricultural Substrate Hypothesis**

A different proposal is found in Kroonen and Iversen’s Agricultural Substrate Hypothesis, which is based on vocabulary and especially noun inflection, in some cases common to Germanic, Balto-Slavic, and Greek. Their authors related it first to the advent of Middle Eastern farmers (Kroonen 2012), and now to the Funnelbeaker culture that was substituted in northern Europe by the Corded Ware culture (Kristiansen et al. 2017).

According to our model, this substrate language would correspond precisely with the language spoken by Corded Ware peoples, that was later substituted by Indo-European from Yamna or Bell Beaker cultures. Kroonen and Iversen’s proposal link the common substrate language to a Middle Eastern language, possibly related to Proto-Semitic. Given the many influences found in Trypillian samples (of which the Usatovo culture is partially the successor), it is unclear which language Corded Ware peoples would have spoken.

The Ice Age refugium of western hunter-gatherers in south-eastern Europe was only partially mixed with Middle Eastern Neolithic farmers up to the Danube. In light of certain macro-family proposals, it might be suggested that the northern region – colonized mainly by R1b1a-L754 lineages (including R1b1a1a-P297 and R1b1a2-V88 subclades) – could have spoken Afroasiatic, or a language common to Afroasiatic and Indo-Uralic, that could have later evolved into the main language of the Old European culture and east-central Europe.

*Figure 42. (On the next pages) Diachronic map of migrations ca. 750 – 1300 AD. See full high-resolution version at <https://indo-european.eu/en/maps/medieval/>.*
IV. Discussion

The core problem addressed by this paper has been the inconsistency found between the prevalent theories on migration routes and the recent research on the genetic make-up of peoples from the Pontic-Caspian steppe.

The Indo-European demic diffusion model proposed here advances the theory that the expansion of Indo-European languages from the steppe was mainly linked to the expansion of peoples of haplogroup R1b1a1a2-M269 in Eurasia.

IV.1. Consequences of the Indo-European demic diffusion model

There is a long-held assumption, since the kurgan hypothesis was laid out (Gimbutas 1963), that Corded Ware herders had helped spread Indo-European languages into Europe and Asia. This assumption, continued into modern times (Beekes 2011), is not fully explained by recent archaeological research (Anthony 2007, 2013; Harrison and Heyd 2007; Heyd 2012), and recent findings in ancient human genetics question it on the grounds of a different path for human migration from the steppes.

To reject this old tenet has wide-ranging consequences:

- The natural trend of Indo-Europeanists to date Indo-European proto-languages all separated at the same time, and usually farther back in time than is warranted by the linguistic evidence (Kortlandt 1990) is challenged, offering a more naturally stepped separation. There is no need to place all known Indo-European branches simultaneously separated in a massive expansion into Corded Ware, Middle Dnieper, Bell Beaker, and Fatyanovo/Abashevo/Sintashta cultures (Anthony 2007, 2013).
• Some Balto-Slavic and Indo-Iranian similarities can be thought of as stemming from a common Corded Ware language substrate, with potential implications for the still prevalent three-dorsal series theory – which seems to be a resilient tradition from the early days of the *centum–sātem* division of Indo-European, and has long been contested with sound linguistic arguments (Lehmann 1952). The association of this language substrate with Proto-Uralic offers an elegant explanation for these developments, and is supported by linguistic, archaeological, and now also ancient genetic data.

• A shared linguistic unity of Italic, Celtic, Germanic, and probably Balto-Slavic, is likely to have existed, probably slightly earlier than Proto-Indo-Iranian, and both later than a potential Paleo-Balkan community.

• Pre-Germanic is more likely to have been imported into southern Scandinavia by peoples of mainly R1b1a1a2a1a1-U106 lineages (maybe already mixed with I1-M253 lineages), marking the transition to the Nordic Bronze Age. The precise pre-North-West-Indo-European linguistic landscape of Scandinavia is unknown, but the previous arrival and likely expansion of peoples of R1a1a1b-Z645 lineages might have brought with them languages of east-central Europe, which had probably in turn replaced earlier Neolithic languages (Kroonen 2012).

• An Italo-Celtic community is compatible with this expansion model, as is their close contact with a Pre-Germanic community, in a period of intense economic exchanges during the Bronze Age.
IV.2. Demic diffusion, cultural diffusion, or founder effect

Ethnos and language are intimately associated, and are known to be much more resistant to change than culture and social stratification, and thus changes in material culture are not to be equated to changes in language, even if ethnicity may take on new meanings (Kristiansen 2000).

Demic diffusion refers to a migratory model of population diffusion into an area that had been previously uninhabited by that group, possibly displacing, replacing, or intermixing with a pre-existing population (Cavalli-Sforza and Minch 1997). It is theoretically the simplest (and thus strongest) link with ethnic and linguistic change, since it shows the predominance of a new people that displaces or absorbs the original one. This is usually accompanied by a decline in Y-DNA variation, since certain chiefdoms and clans usually predominate in the expansion of a population.

Founder effect refers to a loss of genetic variation caused by a colonisation and genetic separation of a subset of the diversity present within the original population, and is different from a bottleneck, where the original population loses its prior diversity by a similar mechanism (Jobling et al. 2014). It seems theoretically second to demic diffusion, in explaining the replacement of genetic make-up without replacement of language. To resort to a founder effect to explain population changes when enough ancient DNA samples are lacking to suggest them is dangerous: the scarcity of ancient DNA samples makes the interpretation of their meaning – in relation to actual ancient areal occupation – a matter of subjective evaluation, in conjunction with archaeological finds (Campbell 2015).

Cultural diffusion in a strict sense (opposed to demic diffusion) refers to the spread of cultural traits – including ideas, technology, and language – between individuals, without a need for a migration. Multiple models have been proposed, but all offer a weaker potential explanation for linguistic change than demic diffusion or founder effect, since all imply that language spreads by way of economic or cultural (e.g. religious) domination. Given the strong ethnic connection of language, examples of such an event were probably exceptional before the creation of the first states. On the other hand, population expansion into certain territories and decline of the original population are followed in some cases by a rising (or “resurgence”) of the original paternal lineages and admixture component (Brandt et al. 2015). A static genetic situation observed after that process could be quite
simply interpreted as cultural diffusion, supposing that no (or almost no) population exchange has taken place. Admixture analyses are quite useful to investigate these cases.

Other potential models can only be weaker than these main three. It seems logical that weaker models should not be used lightly, and clear proof of their applicability (and non-applicability of the stronger models) should be given in each case.

**IV.3. Admixture analysis**

Genetic admixture refers to the analysis of the gene flow between populations that had previously been relatively isolated from one another. Since isolated populations develop linguistic differences relatively quickly, linguistic changes might be expected in a newly hybridised population (Jobling et al. 2014). However, pidgin languages are quite rare, and often one language – usually that of the successful migrants – becomes the superstrate, and another the substrate.

On the other hand, language and culture are unlike a genome in several different ways. While it is possible to obtain admixture percentage of any ancestral population, ancestral language reconstruction and its identification with cultures needs the intervention of careful anthropological investigation. For admixture results to be meaningful, studied loci have to be correctly averaged (and samples should be as complete as possible); genetic drift and selection since admixture have to be taken into account (e.g. distant populations might show a higher differentiation from the original territory); and ancestral populations have to be correctly identified, including their number and precise alleles (Jobling et al. 2014). Therefore, ancient DNA is best collected with the goal of testing specific hypotheses.

Some linguists have used the biological foundations of phylogenetics to extrapolate questionable methods to linguistics, and have thus obtained questionable results (Gray and Atkinson 2003). Similarly, scientists are using the available statistical means to study genetic admixture in modern human populations, extrapolating admixture mapping methods to scarce ancient human samples, and deriving simplistic, far-fetched conclusions. This paper demonstrates the need to include wide anthropological investigation of the historical context of the samples studied, including linguistics, archaeology, and cultural anthropology, as well as careful investigation of haplogroups, to obtain plausible explanations for the complex data obtained in human biology.
It has been proposed that migrating Yamna pastoralists into already expanding Corded Ware groups (Wencel 2015) might have created the necessary environment for the spread of Indo-European languages. Previous mainstream models for Indo-European expansion, based on the “kurgan hypothesis” (Gimbutas 1977) associated the spread of Pre-Germanic (adopted on the Dniester) and Pre-Balto-Slavic (adopted on the middle Dnieper) to the expansion of Corded Ware cultures (Anthony 2007). Given the lack of direct cultural connections between Yamna and the Corded Ware culture, this spread was explained in terms of either an incorporation of languages through centuries of interaction into Funnel Beaker cultures (Kristiansen 1989), or through the emulation of the language of Indo-European chiefs by Corded Ware cultures (beginning ca. 2700-2600 BC) for politico-religious reasons (Anthony 2007).

The component associated with certain Yamna samples is found elevated (up to 76%) in samples of the Corded Ware culture, which has been said to support the migration of Yamna populations into Corded Ware groups. The lower percentage of this component found in Bell Beaker and Únětice groups (50-70%) has been explained as a subsequent, less profound displacement process triggered by western and central European groups (Haak et al. 2015; Allentoft et al. 2015; Mathieson et al. 2015). It has also been found that samples from the Globular Amphora culture do not show evidence of such steppe-related ancestry (Mathieson et al. 2017).

These limited results, apparently challenging archaeological interpretations previously considered established, are propagating quickly within the field of Indo-European studies. David W. Anthony has recently supported the appearance of the Corded Ware culture through the contacts of Yamna immigrants with indigenous people of the Globular Amphora culture in southern Poland (Anthony and Brown 2017), based on their previously known contacts and early dating. Similarly, Kristian Kristiansen has supported the dominance of Corded Ware in central Europe south and north of the Carpathians, asserting that their pottery was apparently shared later by the Bell Beaker culture (Kristiansen et al. 2017).

Many concerns have been raised about obtaining simplistic conclusions based on genetic results (Heyd 2017; Kristiansen et al. 2017):
Figure 43. Modified file from recent papers on ancient samples from Eastern European, Southeastern European, Western European, and Bell Beaker cultures: Left: ADMIXTURE clustering analysis with k=8 showing ancient individuals. E/M/M/LN, Early/Middle/Middle Late Neolithic; W/E/S/CHG, Western/Eastern/Scandinavian/Caucasian hunter-gatherers (Olalde et al. 2017). Center: Supervised ADMIXTURE plot, modeling each ancient individual (one per row), as a mixture of populations represented by clusters containing Anatolian Neolithic (grey), Yamnaya from Samara (orange), EHG (red) and WHG (blue). Dates indicate approximate range of individuals in each population (Mathieson et al. 2017). Right: Ancestral components in ancient individuals estimated by ADMIXTURE (k=11) (Mittnik et al. 2017). All images are originally under a CC-BY-NC 4.0 International license.
IV.3.1. Yamna ancestry: CHG before, during, and after Chalcolithic migrations

IV.3.1.1. Eastern samples

Samples from the Pontic-Caspian steppe – from which ‘steppe’ or ‘Yamna’ ancestry has been defined as a precise combination of EHG and CHG ancestry – are scarce, and the most recent ones mostly from one eastern region (Kalmykia). Because of that, east Yamna was considered the best-known proximate source for the incoming gene flow in Corded Ware samples. The exact source could have been another, yet unsampled, group of people closely related to them (Kristiansen et al. 2017), and a western or earlier (pre-Yamna) steppe population has been suggested as the potential missing link in the chain of transmission of steppe ancestry (Haak et al. 2015).

This so-called ‘steppe’ ancestry from eastern Yamna samples has been found in Corded Ware, Afanasevo, Andronovo, and Srubna cultures, and even a late individual of Bronze Age Bulgaria from Merichleri12, ca. 1690 BC. All of them show higher ‘steppe’ ancestry than some samples clearly identified as from the Yamna culture in Ukraine and in the Balkans, more than a thousand years earlier (Mathieson et al. 2017). A similar western Yamna ancestry is also found in a sample from Vučedol, which probably also descended directly from early western Yamna migrants (see Figure 43).

Samples from central Balkans show in fact a relative increase in steppe ancestry later, during the Bronze Age – unlike west Europe and the southern Balkans, where ancient Indo-European languages were most likely spoken by that time. Furthermore, admixture analyses of modern populations show more steppe proportion in modern north-eastern European populations (including peoples probably speaking Finno-Ugric languages since the Neolithic) than in western European peoples that are known to have spoken Indo-European languages for millennia (Haak et al. 2015, fig. 3).

Most Corded Ware samples are late, almost coincident with the Bell Beaker expansion. Scarce or no samples have been published from potentially controversial areas – like the Contact Zone, north-eastern Europe, the Baltic and the Forest Zone, and western Yamna – during the most relevant periods. Old samples (closer to admixture events) tend to show a higher range of variation, and could inform better of the real impact of migrations, while younger ones – depending on non-random mating processes, influenced by geographic

---

12 The individual from Merichleri, of R1a1a1-M417 lineage might hint, in fact, to an ancient connection of the area with the second Corded Ware horizon.
structure or socioeconomic factors – may falsely show a relatively homogenous high or low ancestral contribution (Jobling et al. 2014).

The oldest sample from Erperstedt (labelled I0104), a second-degree relative to the Erperstedt family (Monroy Kuhn, Jakobsson, and Günther 2017), has been found to cluster the closest to steppe samples, closer than any other Corded Ware sample, previous or posterior, or samples from eastern Corded Ware-derived cultures Sintashta, or Potapovka (Haak et al. 2015). This, connected with the exogamy prevalent among Corded Ware peoples, and the nomadic nature of its culture, precludes a proper interpretation of the ancestry found in the family.

In fact, PCA analysis reveals that early Corded Ware samples, like those from Estonia and Latvia, as well as the Poltavka outlier – in contrast e.g. with Poltavka samples – cluster close to EHG and central European populations rather than to steppe samples. Late Corded Ware samples, as well as Sintashta and Potapovka samples, also cluster closely to central European samples, in contrast with Poltavka and Afanasevo individuals, suggesting an east-central European genesis of the Corded Ware culture.

**IV.3.1.2. Western samples**

The northern Pontic area – from where many Yamna migrants are assumed to have expanded west along the Danube – had been a zone of interaction with peoples from the upper Danube and the Eastern Baltic for millennia – and could thus cluster closer genetically to peoples from Carpathian cultures than the eastern Pontic-Caspian steppe and steppe-forest zone. Individuals from the Balkans at Varna I (ca. 4630 BC), Smyadovo (ca. 4500 BC), and outliers from Ukraine Eneolithic (ca. 3500 BC) and Trypillian culture (ca. 3325 BC) clearly show such steppe ancestry before the main Chalcolithic expansion (Mathieson et al. 2017; Haak et al. 2015).

Yamna migrants from the eastern zone (whose samples are used to define steppe ancestry) had migrated westward to the north Pontic area and beyond along the Danube at least twice: first in the formation process of the early Khvalynsk and Sredni Stog cultures, and later during the formation of the Yamna culture (see above). On the other hand, late Sredni Stog regions seem to have adopted a different culture than the developing Yamna to the east, potentially suggesting a different ethnolinguistic nature. This most likely created a more mixed Balkan and east-central European genetic pool among qualitatively different cultures (probably speaking different languages, with expanding clans dominated by
different lineages), which questions the validity of certain conclusions about the origin of the admixture found in Corded Ware cultures.

The ‘Yamna component’ is paradoxically found in lesser proportion in western Yamna samples. These include the so-called Yamna outlier from Ozera (ca. 3005 BC), one of only three samples from Ukraine; and one sample from a Yamna migrant in Mednikarovo (ca. 2960 BC). Both samples cluster closer to Balkan samples, and potentially to a ‘southern’ ancestry (see above Figure 15), between CHG/Iran Neolithic and Levant Neolithic. This same influence may explain the similar drift in the population of the original Corded Ware culture to the north of the western steppe, compared with its previously mainly EHG ancestry.

A simplistic assumption of recent genetic models, based on proportions of Yamna admixture, suggested that Corded Ware contributed to the creation of the Bell Beaker culture (Haak et al. 2015), which has supported recent proposals of a direct evolution from the former to the latter (Kristiansen et al. 2017; Anthony and Brown 2017). The influence of the Corded Ware outlier from Erperstedt in assessing steppe ancestry of the Corded Ware culture as a whole, and the elevated steppe ancestry found recently in Bell Beaker samples from Hungary and western Europe, demonstrate that this assumption is wrong. The closer position of Bell Beaker samples to Yamna samples – closer than any other sample of the Corded Ware in PCA (Olalde et al. 2017) excluding the I0104 outlier –, as well as the different position of western Yamna samples and Vučedol, makes such a direct connection with Corded Ware migrants even more unlikely. Traditional models of Yamna and Bell Beaker expansion (Harrison and Heyd 2007; Heyd 2007, 2012), accompanied by the expansion of North-West Indo-European (Mallory 2013; Prescott 2012), therefore, seem to be sustained by genetic investigation.

**IV.3.1.3. Ancient and modern samples**

Scattered samples from different periods (by millennia) from the Forest Zone and steppe already showed certain common clusters before the Neolithic and Chalcolithic expansions in global ancestry profiles, in the first articles published\(^\text{13}\). More recently, Estonian samples have shown a genetic component associated with CHG coinciding with the spread of R1a-Z645, which rules out Corded Ware and Yamna as the only origin of this

\(^{13}\) See e.g. Extended Data Figure 2 in Haak et al. (2014), Extended Data Table 2 in Mathieson et al. (2015), Figure 2 in Jones et al. (2017).
component (Saag et al. 2017). A female from Zvejnieki, dated ca. 2885 BC and classified as from Latvian Neolithic/Corded Ware culture (Jones et al. 2017), has been found to cluster quite closely with eastern Yamna samples (Mathieson et al. 2017), in spite of its ancestral component’s supposed origin further south, and necessarily including some generations of admixture with the local population. This is coincident with the ‘southern drift’ observed in samples from western Yamna in the PCA, supporting thus a common genetic drift of forest-steppe and steppe populations, of the northern Pontic-Caspian region.

It is known that the genetic isolation of Eurasian hunter-gatherers came to an end with the arrival of farming and pastoralism. This is seen in the evolution of Middle Eastern ancestries during the Neolithic and Chalcolithic (Lazaridis et al. 2016), and it is becoming clearer too with the genetic flow seen in eastern Europe during the Neolithic and Chalcolithic. Even though samples are scarce and distant, Neolithic individuals from Comb Ware (Zvejnieki), Late Khvalynsk (Samara), and Old Europe (Varna I, Smyadovo) cultures show a clear pattern towards lesser inter-group genetic distances, clearly seen in the appearance of CHG in Admixture, and in their convergence in PCA analysis (Mathieson et al. 2017).

Two female samples from Bohemia were misidentified as Bell Beaker (Allentoft et al. 2015), when they were in fact three millennia younger, from Czech Slavs (Mathieson et al. 2017). PCA or Admixture did not (and cannot) show differences with Bell Beaker or Balkan samples, since parental populations need to be available, or else archaeological context is needed to define demographic models and potential ancestral populations, to ascertain their actual link to the so-called steppe ancestry. In fact, there is a clear north-south cline of steppe ancestry in modern populations, peaking in the Forest Zone, which mimics to some extent its geographic distribution after the Corded Ware and Yamna expansions (Haak et al. 2015), and thus also potentially to some extent a previous situation (Klejn et al. 2017).

**IV.3.2. Demographic issues**

The migration of Pontic-Caspian steppe into Neolithic/Bronze Age central Europeans has been argued to be strongly male-biased (Poznik et al. 2016), with a study suggesting up to 14 migrating males for every migrating female (Goldberg, Günther, et al. 2017), but different in the rates regarding Corded Ware, Bell Beaker, and Únětice. The results of the latter study have been disputed (Lazaridis and Reich 2017), and this in turn contested by
the original authors based on the impact of small, low-coverage ancient samples in admixture analyses (Goldberg, Gunther, et al. 2017). This questions the accuracy of predictions made based on certain samples and methods used.

In terms of mtDNA, Bell Beaker and Corded Ware samples both show cases of common hunter-gatherer haplogroups U5, U2, or U4. Unlike Corded Ware, the Bell Beaker culture shows a higher proportion of haplogroup H (Brandt et al. 2013; Olalde et al. 2017), proper of western Europe, coinciding thus with its expansion to the west. Out of seventeen samples from the Corded Ware culture, four (all of Erperstedt) and one out of five of Sintashta include different J1c subclades, only found previously accompanying the expansion of Neolithic Middle Eastern farmers, including the Globular Amphora culture (in two out of nine samples).

The early Latvian Corded Ware sample (of ca. 2885 BC) is of mtDNA haplogroup U5a1b (Mathieson et al. 2017), a haplogroup found previously in four Sredni Stog samples from Deriivka (ca. 5150 BC). Later it is also found in a central European sample at Benzingerode – dated as of ca. 2275 BC, and dubiously classified as of the Bell Beaker culture, based only on the burial position (Haak et al. 2015) –, and in a sample of the Únětice culture from Przecławice (ca. 1790 BC).

Ascertaining global differences in demographic changes is especially important in light of an apparently mostly peaceful Yamna migration along the Danube (Heyd 2012), contrasting with the potentially violent and strong patrilocality shown by peoples of the Corded Ware cultures (Kristiansen et al. 2017). Peoples of the expanding Corded Ware horizon were of nomadic and exogamic nature, keeping direct contacts with the steppe since its expansion.

Quite relevant for the effect of the invading population is also the population density prior to the invasions, and the actual increase in population estimated after such population expansions, being both greater in south-east Europe (Müller 2013). There are therefore great potential differences in population admixture between both the Corded Ware and the Bell Beaker cultures, roughly expanding westward to the north and south of the loess belt that had previously divided the expanding farmers from hunter-gatherers.

All these differences might have greatly influenced the genetic drift observed, and must be taken into account to make inferences about the actual origin and influence of the population involved in Corded Ware and Yamna/Bell Beaker migrations.
IV.3.3. Technical issues

Shortcomings of methods used for the analysis of ancestral populations are usually not evident, and may affect any theory developed based solely on these methods.

Extraction techniques, analysis in different sequencing centres and compilation in different platforms, classification of poorly known individuals into cultures, and variability of radiocarbon dates obtained in different labs, are just few of the many known issues involved in human evolutionary biology.

The scarcity of samples adds difficulty to the classic problem of characterization of discrete population structure in the presence of continuous patterns of genetic differentiation. The “clines versus cluster” problem in modelling population genetic variation should be addressed taking into account geographical barriers (Bradburd, Coop, and Ralph 2017), which necessarily involves a detailed description of ancient geography, ecology, mobility, etc. for any period investigated.

Principal component analysis (PCA) is a variable-reduction technique, similar to exploratory factor analysis. It reduces a larger set of variables into a smaller set of artificial variables, called principal components (PC), which account for most of the variance in the original variable. This method assumes that there is a linear relationship between variables, that there is sampling adequacy: a precise number of cases is difficult to evaluate, but it is to be assumed that scarce, damaged samples of ancient DNA preclude an ideal sample size. Variables also need to have adequate correlations in order for them to be reduced to a smaller number of components, and there should be no significant outliers.

PCA of ancient DNA samples show usually a large number of principal components, of which the most common ones selected for graphic analysis (PC1 and PC2) can usually explain (have a combined eigenvalue of) no more than 5-10% of the total variance, depending on the samples selected (see Figure 44). This is in line with the prediction that most eigenvalues of the theoretical covariance will be ‘small’, nearly equal, but is in contrast with the expectation that a few eigenvalues will be ‘large’, reflecting past demographic events (Patterson, Price, and Reich 2006).

---

14 This is a guesstimate based on the limited experience of the author with free datasets.
Theoretically, though, under a small number of ancestral populations, with small divergence among them, just two significant eigenvalues will exist, and selecting the two main significant axes of variation captures the most relevant information. On the other hand, ancestral populations – like certain African populations, and ancient hunter-gatherers – may show dozens of significant axes, whose meaning is unclear.

Analysis with STRUCTURE (Pritchard, Stephens, and Donnelly 2000) / ADMIXTURE (Alexander, Novembre, and Lange 2009) is also mainly reported in genetic papers, as PCA, following graphical patterns. It is important to take into account that mixed ancestry in an individual can result – apart from genetic admixture of two isolated populations, which is the object of the study – from shared ancestry (coinheritance of more than one ancestry from the same parental source, incomplete differentiation), and also from assimilation.

The number of ancestral populations selected is based on cross-validation error estimation and graphical analysis. It uses, therefore, a combination of numerical and graphical methods in a similar way to the factor extraction in PCA, but less formally explored (see Figure 45 and Figure 46).
Figure 45. Detail of image, modified from Mathieson et al. (2017). “Unsupervised ADMIXTURE plot from k=3 to 11, on a dataset consisting of 799 present-day individuals and 405 ancient individuals. We show newly 759 reported ancient individuals and some previously published individuals for comparison.” Original under a CC-BY-NC 4.0 International license.

Figure 46. Detail of unsupervised ADMIXTURE plot from k=3-6, 8, 10, and 12, on a dataset consisting of 318 ancient individuals, including Minoan and Mycenaean samples, analysed in Lazaridis et al. (2017). The Corded Ware outlier of Erperstedt (10104) and members of his family have been marked and labelled. Component colours have been used.
In a recent study, evidence supported the selection of 21 ancestries to delineate genetic structure of present-day human populations (Baker, Rotimi, and Shriner 2017), although this is debatable.\(^{15}\) It is unclear whether this ‘ideal’ number would be greater or lesser for ancestral, more isolated populations, and the lack of proper sampling precludes a proper selection of K. The usual small number of inferred ancestral components, selected to show ancestries in a simplified manner (K=2-6), may thus be too simplistic, although a K = ~8-11 appears to be a good range of components for the study of modern populations (see Figure 47).

Therefore, the selection and naming of a population as ‘ancestral’ to another is indeed conventional, and can lead to error when its nature as approximate source or proxy among poorly investigated populations is not fully understood. With new results, the naming of certain ancestral populations may become obsolete, as more ancestral proxy populations are discovered.

\(^{15}\) Iosif Lazaridis (Twitter 3/9/2017), criticises the choice of K=21 as a “minimum”, as well as the concept of “mixed ancestry” meaning the possession of >1 of K=21 components in an admixture analysis over ~19k SNPs.
estimation, admixture graph fitting, and rolloff, included – among other tools – in the free software package ADMIXTOOLS (Patterson et al. 2012). They are robust tools based on statistical methods, but each method is dependent on certain assumptions. So, for example, an estimation of mixing proportions in a three-population test, when phylogeny for the populations studied is incorrect, leaves such proportions without useful meaning. Even discussing mixing from an ancestral population, when an intermediate admixing event occurs, hardly makes sense.

IV.3.4. Cognitive bias, conflicts of interest, contextual bias

In the academic community, prestige, access to grants, and even jobs depend on getting articles published in journals of high impact factor. These journals prefer short articles, mainly based on mathematical methods (preferably with reference to improvements in such methods), groundbreaking conclusions, and self-important titles, with a tendency to “culture-historicism”.

Pressure to publish means also pressure to gather, analyse and interpret the data. However, knowledge and expertise in gathering genetic data from archaeological remains does not mean expertise in statistics and computer science. Statistical knowledge does not qualify one to infer conclusions based on results either, unless one has some previous knowledge of the anthropological subjects involved. Otherwise, researchers concerned with fieldwork and statistical methods are exposed, during the interpretation of results, to the risks of circular reasoning and confirmation bias, by searching only for anthropological information that might fit their results. In this sense, a clear trend can be observed in recent publications, whereby wide-ranging conclusions in genetic papers tend to become outdated in very short periods, as new samples become available.

For the general population, SNP investigation offers a simple view of one’s own paternal line, that a thousand years (or ca. 30 generations) ago would represent a 1,000,000,000th of one’s own genealogical tree; four or five thousand years ago, its contribution to a personal ethnolinguistic definition is non-existent. This, together with the perceived complexity (and lack of familiarity with) intricately linked anthropological disciplines, has made human ancestry investigation quite popular among amateur geneticists, who can easily play with published open source software programs and free aDNA datasets, due to their accessibility. However, the correct use of these programs needs much more than just knowing how to apply certain commands to some data. The quest for one’s own personal and national “ethnic proportion”, often as part of pre-existing simplistic
ethnolinguistic beliefs and socio-political agendas, is also being promoted by commercial
 genetic testing companies to sell their products, in what would certainly be a reason for
 Kosinna’s smile today.

IV.4. SNP comparison
This demic diffusion model relies on the comparison of ancient and modern Y-DNA
SNPs, by observing how patrilineal lineages are replaced in certain areas that belonged
to certain archaeological cultures.

While analysis of whole genomes may be biased, whether by chance (drift) or by selection
(Jobling et al. 2014), careful investigation of SNPs with help of anthropological
disciplines can be used to obtain meaningful conclusions.

Examination of SNPs of the Y-chromosome of ancient individuals one by one seems more
suited to the scarcity of aDNA samples available, and the quality of its recovery, since
defects in the STR sequencing are frequent, and thus only certain SNP markers may be
obtained, with less information – and higher subclades – obtained from the samples.

This technique of careful haplogroup analysis has been used successfully for ancient
populations to detect familial links and paternal clans, their expansion and preferential
marriages among elites, in the Yakut population (Zvenigorosky et al. 2017).

The date calculated for TMRCA of SNPs in modern populations has been used to define
when certain migrations or expansions might have occurred. While it seems a good
starting point for that purpose, it relies on the survival of modern populations related to
such ancient population movements, and as such it could miss initially successful lineages
that are now extinct, and that could have given an earlier date if they had been included
in TMRCA calculations.

Ancient and modern mtDNA distribution analyses – although they can help more clearly
determine migration paths (Brandt et al. 2013) and other interesting characteristics of
ancient cultures, such as female exogamy (Sjogren, Price, and Kristiansen 2016) –, has
not been included in this paper for simplicity purposes.

Potential language relationships have been used to illustrate the Indo-European demic
diffusion model. Many long-term linguistic relationships beyond Middle Indo-European
remain hypothetical at best – when not completely discarded with the current data –, and
it is not the intention of this paper to support or dismiss such connections. Such
relationships – like Indo-European dialectalisation – must be proven by linguistic research, whereas archaeology and genetics can only add precision to it.

While the theory here presented seems rational and scientifically sound, there are many alternative explanations that could have been made of the same data: these have been omitted for the sake of simplicity.

More ancient DNA samples are needed to precisely draw most details of the general theory laid out in this paper.

**IV.5. Alternative models of expansion**

This model uses mainly SNP investigation to support the general view of Archaeologists whereby Yamna migrants did not directly form the Corded Ware culture (Anthony 2007; Harrison and Heyd 2007; Heyd 2007, 2012; Kristiansen et al. 2017; Heyd 2017), correcting thus assumptions based solely on recent genetic research.

Communities formed mainly by R1b1a1a2a-L23 and R1a1a1b-Z645 lineages, which clearly expanded during the Chalcolithic – associated to Yamna and Corded Ware cultures, respectively – are assumed to have evolved differently in eastern Europe, albeit in close contact (probably in a neighbouring region) in light of the common ancestry they share.

Other models are compatible with the state of the art of genetic investigation, though. Listed here are possibilities compatible with the current knowledge, in decreasing order of probability:

**IV.5.1. “Kurgan people”**

Yamna clans of R1a1a1b-Z645 lineages existed and migrated to the north, but have not been sampled yet. This seems to be the preferred model in recent genetic papers, which support a direct Yamna > Corded Ware migration (Haak et al. 2015; Allentoft et al. 2015; Lazaridis et al. 2016).

This framework seems compatible with Kristiansen’s traditional model of Corded Ware culture development (Kristiansen 1989; Kristiansen et al. 2017), and also partially with that of Gimbutas (Gimbutas 1965, 1977), involving a gradual cultural (and population) contribution of Indo-Europeans from the steppe to east-central Europe. Centuries of contact between the Indo-Europeans from the steppe and the Baden and Globular
Amphora cultures should have formed the third Corded Ware horizon that expanded after ca. 3000 BC.

This seems in part supported by steppe ancestry found in individuals from the Balkans at Varna I (ca. 4630 BC), Smyadovo (ca. 4500 BC), and outliers from Ukraine Eneolithic (ca. 3500 BC) and Trypillian culture (ca. 3325 BC). All of them clearly show such steppe ancestry before the main Chalcolithic expansion (Mathieson et al. 2017; Haak et al. 2015). However, it is in contrast with the lack of steppe ancestry in sampled distant individuals from Baden and Globular Amphora cultures (Mathieson et al. 2017). Also, from a linguistic point of view, this long-lasting framework does not allow for the adoption of a Late Indo-European language by the Corded Ware culture, and would still leave the most obvious Late Indo-European-speaking expansion to the south-eastern spread of Yamna migrants.

On the other hand, the model presented in this paper, which supports a Corded Ware Substratum Hypothesis, would be compatible with eastern Yamna clans of R1a1a1b-Z645 lineages expanding a language ancestral to Slavic and Indo-Iranian with the Corded Ware culture, and therefore the substrate language would correspond to the original east-central region where the culture originated in the Late Neolithic. In that case, Kortlandt’s late *Satem* or Indo-Slavonic dialect (Kortlandt 2016) could also be supported, but the common unconnected substratum of Balto-Slavic and Germanic (Kortlandt 2016) would probably need further assumptions, like the Agricultural Substrate Hypothesis (Kristiansen et al. 2017; Kroonen 2012; Iversen and Kroonen 2017).

Problems with this model include:

- This model assumes the existence of differentiated clans with a majority of R1a1a1b-Z645 lineages coexisting with clans with a majority of R1b1a1a2a1-L51 subclades, either in the east or in the west – or, exceptionally, in both, since they are different dialectal areas. Both of them would have expanded during the Chalcolithic: R1a1a1b-Z645 clans to the north (possibly via the Prut), to form the third Corded Ware horizon, and R1b1a1a2a1-L51 and R1b1a1a2a2-Z2103 clans to the east into the Afanasevo culture, and west to form Yamna settlements of the Balkans.

- It would need a framework for a sudden and direct contribution of Yamna to the creation of the Corded Ware culture, which currently does not exist.
- The coexistence of such differentiated communities is in contrast with the natural evolution of male-dominated steppe societies of R1b1a1a2-M269 and R1b1a1a2a-L23 lineages, which underwent at least two important expansions, and must have therefore undergone a decrease in Y-DNA variability since the Mesolithic.

- It is also in contrast with the lack of R1a1a1b-Z645 lineages in Yamna, and in western and south-eastern European cultures related to the expansion of Indo-European languages. A potential explanation for this lack of R1a1a1b-Z645 would be a sample selection bias, whereby western Yamna migrants of R1a1a1b-Z645 subclades and central-east European Corded Ware samples of R1b1a1a2a-L23 subclades have not yet been sampled.

A further assumption of this model is that Corded Ware contributed to the creation of the Bell Beaker culture (Haak et al. 2015; Kristiansen et al. 2017), which is in contrast with mainstream archaeological models, and is based mainly on the influence of the Corded Ware outlier of Erperstedt in assessing steppe ancestry of the Corded Ware culture. The closer position of Bell Beaker samples from the Balkans to Yamna samples – closer than any other sample of the Corded Ware in PCA (Olalde et al. 2017) –, as well as their connection to Vučedol and western Yamna ancestry, makes such a direct connection unlikely.

IV.5.2. Founder effect / resurgence of R1a-M417 lineages, or cultural diffusion
West Yamna was mainly composed of clans of R1b1a1a2a-L23 subclades, as supported in this model, but the admixture seen in Corded Ware samples from central and north-eastern Europe comes precisely from original Yamna migrants of R1b1a1a2a-L23 subclades that have not been sampled. These first generations of northern migrants would have travelled west and north-west, maybe involving those known to have settled north up the Prut River, possibly beginning as early as 3100 BC, but probably from ca. 3000 BC.

This framework could be compatible with Anthony’s proposal of elites from the steppe dominating over the Usatovo culture (Anthony 2007, 2010; Anthony and Ringe 2015).

On the other hand, Anthony’s model would still need a majority of the Usatovo population mainly composed of R1a1a1b-Z645 lineages, and with a previous ancestry closest to the east Yamna population, as well as a resurge of R1a1a1b-Z645 lineages within few
generations, so that previous “native” admixture and Yamna lineages could go unnoticed in the Corded Ware population sampled. To sample individuals from only a few generations is difficult, apart from the indemonstrable proposal (advanced by Anthony) that the general Usatovo population would have adopted the prestige language of the elites (in this case Pre-Germanic) by way of cultural diffusion. It also leaves open the same possibility for multiple (as unlikely and as indemonstrable) cultural diffusions among Corded Ware groups due to the technological superiority of the Yamna culture, e.g. of Proto-Balto-Slavic to the Middle Dnieper culture.

IV.5.3. R1b-L51 from the west and cultural diffusion

West Yamna clans were mainly composed of clans of R1b1a1a2a2-Z2103 lineages, which expanded with Yamna migrants to the west. Subclade R1b1a1a2a1-L51 (and especially R1b1a1a2a1-L151 subclades) may have split after the Neolithic expansion associated with Middle Indo-European, and developed a society in central and southeastern Europe.

It would have later adopted the ‘Yamna package’, developing the East Bell Beaker group that later expanded further to the west. That could be supported by the finding of a R1b1a1a2a2-Z2103 subclade in Vučedol, and by the lack of R1b1a1a2a1-L51 subclades in Yamna. This model could be linked with the proposed origin of Proto-Indo-European in western or central Europe (Cunliffe and Koch 2012).

This is in contrast with estimated dates of haplogroup formation, and the inferred history from ancient samples. The evolution of R1b1a1a2-M269 and R1b1a1a2a-L23 lineages within the steppe since the Mesolithic, the quite late TMRCA for R1b1a1a2a-L23, and the finding of an R1b1a1a2a-L23(xZ2013) subclade in eastern Yamna all point to an internal evolution of these subclades. This and the expansion of R1b1a1a2a1-L151 subclades (but not R1b1a1a2a1-L51) with the Bell Beaker culture, probably evolved from Yamna migrants in the westernmost region, makes the eventual appearance of R1b1a1a2a1-L51 subclades in west Yamna quite likely.

This model would also allow for a Yamna region hiding unsampled clans of R1a1a1b-Z645 lineages, as in the “Kurgan people” model.
**IV.5.4. Anatolian hypothesis**

An alternative hypothesis is that the Proto-Indo-European homeland was in the Caucasus or Iran, and expanded through Anatolia, which can be linked to the most recent proposals of the Anatolian hypothesis (Renfrew 2003).

The current genetic models are not, however, compatible with the Armenian homeland hypothesis (Gamkrelidze and Ivanov 1995), which suggests a more recent expansion of Late Indo-European from the Armenian highland.

A westward movement associated with the CHG ancestry may have thus contributed to the dispersal of Anatolian languages, seen in the contribution of CHG to Anatolian peoples during different Neolithic and Chalcolithic waves. The northward mixture of CHG with EHG ancestry in the steppe may signal the formation of Indo-European-speaking steppe population associated with the Chalcolithic expansion of Late Indo-European (Mathieson et al. 2017; Lazaridis et al. 2017).

In line with this model, one could support the North Pontic Hypothesis, whereby the Proto-Indo-European homeland should be placed north of the Black Sea, linked to Old European cultures, and in contact with potentially related Uralic and Semitic languages (Vander Linden 2004).

However, haplogroup analysis reveals a prevalence of J-M304 lineages associated with CHG expansion into Anatolia, probably related to movements within the Fertile Crescent, which is not seen in the Yamna population. This suggests a different type of spread for this component in the steppe.

No archaeological models show such a strong wave of migrants from the south into the steppe, but it shows continuous contacts of steppe cultures with Transcaucasian cultures since the Mesolithic. It is therefore more likely that the CHG contribution came from a long-term inter-regional gene flow that began early, probably coinciding with Neolithic population movements within the steppe.

Controversial linguistic macro-families are also compatible with other mainstream models of expansion.
V. Conclusion

Careful cross-disciplinary investigation of ancient DNA samples recently published supports a demic diffusion model for the expansion of Indo-European-speaking peoples directly into central and western Europe through the Bell Beaker culture, challenging previous archaeological and linguistic theories based on the expansion through the Corded Ware culture. Potential consequences of this new model in archaeological and linguistic investigation have been outlined in this paper, among them the development of a stable framework of time and space for Indo-European dialectal classification, allowing for a more precise dating of Indo-European branches and their splits and expansions, and why and how they might have occurred.
Acknowledgements

I have received comments and corrections by Fernando López-Menchero, and comments by Frederik Kortlandt in relation to his publications.

I have benefitted from the continuous work of Jean Manco, who keeps an ordered and updated list of the SNPs of most ancient DNA samples obtained to date at ancestraljourneys.org; from Razib Khan’s posts at Gene Expression; from David Wesolowski’s posts at Eurogenes; and from comments of Anthrogenica forum users.

I am also indebted to those readers who have emailed me to correct the information and maps of the first version, and to suggest new articles and books.
References

Andrades Valtueña, Aida, Alissa Mittnik, Felix M. Key, Wolfgang Haak, Raili Allmäe, Andrej Belinskij, Mantas Daubaras, Michal Feldman, Rimantas Jankauskas, Ivor Janković, Ken Massy, Mario Novak, Saskia Pfrengle, Sabine Reinhold, Mario Šlaus, Maria A. Spyrou, Anna Szecsényi-Nagy, Mari Törv, Svend Hansen, Kirsten I. Bos, Philipp W.


Repeated Author. 2010. The horse, the wheel, and language: how Bronze-Age riders from the Eurasian steppes shaped the modern world. New Jersey: Princeton University Press.


Horvath, Csaba Barnabas. 2015. R1a subclades and Bronze Age migrations on the Eurasian steppes. ESJ 2.


Klochko, V. I., and A. Koško. 2009. The societies of Corded Ware cultures and those of Black Sea steppes (Yamnaya and Catacomb Grave cultures) in the route network between the Baltic and Black Seas. *Baltic-Pontic-Studies* 14:269-301.


Lobov, Artem Sergeevich. 2009. Структура генофонда субпопуляций башкир, Институт биохимии и генетики, Уфимский научный центр РАН, Уфа.


Indo-European demic diffusion model

Pokutta, Anna. 2013. Population Dynamics, Diet and Migrations of the Únětice culture in Poland, Gothenburg, Department of Archaeology, University of Gothenburg, Gothenburg.


Przybyła, Marcin S. 2009. Intercultural Contacts in the Western Carpathian Area at the Turn of the 2nd and 1st Millennia BC: Narodowe Centrum Kultury.


Early Divergent Strains of *Yersinia pestis* in Eurasia 5,000 Years Ago. *Cell* 163 (3):571-582.


Scheeres, Mirjam. 2014. High mobility rates during the period of the “Celtic migrations”? 87Sr/86Sr and δ18O evidence from Early La Tène Europe, Fachbereich Biologie, Johannes Gutenberg–Universität Mainz, Mainz.


Szmyt, Marzena. 2013. The circulation of People and Ideas in the Baltic and Pontic Areas during 3rd millennium BC.


Woudhuizen, Frederik Christiaan. 2006. The ethnicity of the Sea Peoples, College voor Promoties, Erasmus Universiteit Rotterdam, Rotterdam.


Zhao, Yong-Bin, Ye Zhang, Quan-Chao Zhang, Hong-Jie Li, Ying-Qiu Cui, Zhi Xu, Li Jin, Hui Zhou, and Hong Zhu. 2015. Ancient DNA Reveals That the Genetic Structure of the Northern Han Chinese Was Shaped Prior to 3,000 Years Ago *PLoS One* 10 (5):e0125676.


*Indo-European demic diffusion model* 185