Indo-European demic diffusion model

2nd 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

June 2017
Preface

This paper is a completely revised version of *Indo-European demic diffusion model*, an essay published in February 2017.

For the past three years I had observed with interest the recent developments in human evolutionary genetics, especially those regarding the potential expansion of Proto-Indo-Europeans.

I saw certain flaws in conclusions derived from the results of the most recent and famous papers on the subject, and because some renown researchers in the academic community and many amateur geneticists seemed to be accepting those results without questioning them, I decided to write a short sketch with my concerns in the free time I had, just after depositing (and before defending) my Ph.D. thesis.

Because I didn’t want to just criticize the limitations of recent papers – something anyone with little knowledge of the many shortcomings of human evolutionary genetics can do –, I opted for taking a constructive approach, by offering my own detailed account of important prehistorical events.

The little I had to offer to this complex field of Indo-European studies came from the different interpretation I think should be made of the most recent genetic data, in light of mainstream linguistic and archaeological models. The work I published – and the revised work I publish now – is obviously heir of essential works written by Volker Heyd, Frederik Kortlandt, David W. Anthony, James P. Mallory, or Kristian Kristiansen, among others, and should be interpreted as a tribute to them.

The positive response I have received from linguists and amateur geneticists, and the recent publication of papers with new samples, especially those of Bell Beaker (Olalde et
al. 2017) and south-eastern European cultures (Mathieson et al. 2017), compels me to publish a thorough revision of this article, with more attention to detail. It seems that new samples and admixture analyses are supporting the traditional linguistic and archaeological view and its consequences, as described in my original paper (confirmation bias notwithstanding), which encourages me to keep working on this theory.

An important change to this version is the dropping of my own terms for haplogroups – which I deemed a gentler nomenclature for academics of anthropological fields – in favour of ISOGG standards, hoping that the advantage gained in exactitude will exceed the sacrificed readability. Another due change involved adding important sources for the maps, some of which have been carefully revised.

Even if the current mainstream interpretations regarding steppe ancestry could eventually prove correct, it is important to question the methods that are being used to achieve those conclusions. This paper offers a good starting point to criticize such methods with the most plausible explanation for the observed data.
Abstract

Introduction.

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.

Results.

SNPs and admixture analyses 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 mainly linked to the expansion of peoples belonging to haplogroup R1b in 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 linked to 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

Preface .............................................................................................................................. 3

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

Table of Contents ............................................................................................................. 6

List of Figures ................................................................................................................... 9

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

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

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

   III.1. Palaeolithic-Mesolithic ...................................................................................... 19

      III.1.1. R1b-M343 ................................................................................................... 19

      III.1.2. R1a-M420 ................................................................................................... 20

      III.1.3. Younger Dryas and Boreal periods ............................................................ 20

      III.1.4. Languages ................................................................................................... 22

   III.2. Mesolithic-Neolithic transition .......................................................................... 25

      III.2.1. Languages ................................................................................................... 26

   III.3. Chalcolithic ........................................................................................................ 30

   III.4. Late Indo-European ........................................................................................... 34

      III.4.1. Contacts with the Caucasus ........................................................................ 35

   III.5. Forest Zone ........................................................................................................ 38

      III.5.1. Expansion of the Corded Ware culture ....................................................... 39
III.6. Tocharian expansion ................................................................. 45
III.7. Late Indo-European expansion .................................................. 48
  III.7.1. Yamna migration ............................................................... 48
  III.7.2. The expansion of Balkan Indo-European ......................... 51
  III.7.3. The expansion of North-West Indo-European in Europe ... 55
III.8. Indo-European in Corded Ware societies ................................. 80
  III.8.1. Balto-Slavic ................................................................. 80
  III.8.2. Indo-Iranian .................................................................. 86
  III.8.3. A common Corded Ware substrate .................................. 96

III. Discussion .................................................................................... 101
  III.1. Consequences of the Indo-European demic diffusion model ........ 101
  III.2. Demic diffusion, cultural diffusion, or founder effect ............... 103
  III.3. Admixture analysis ............................................................. 104
  III.4. SNP comparison ................................................................... 109

IV. Conclusion ..................................................................................... 111

Acknowledgements .............................................................................. 113

References .......................................................................................... 115
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 (continued on the next page). See full version at <https://indo-european.eu/en/maps/palaeolithic/>. ................. 23

Figure 3. Diachronic map of Mesolithic migrations ca. 6500-5000 (Anthony 2007; Piezonka 2015), Uni-Köln (continued on the next page). See full high-resolution version at <https://indo-european.eu/en/maps/mesolithic/>. ........................................ 28


Figure 5. Modern distribution of R1b1a1a2-M269 (xL23) lineages, adapted from Richard Rocca (2012). ................................................................. 33

Figure 6. Diachronic map of Eneolithic migrations ca. 4000-3100 BC (Anthony 2007; Szmyt 2013; Piezonka 2015), Uni-Köln. See full high-resolution version at <https://indo-european.eu/en/maps/eneolithic/>. ................................................... 36


Figure 8. Diachronic map of migrations in Asia ca. 3100-2800 BC. See full high-resolution version at <https://indo-european.eu/en/maps/copper-age/>................ 46

Figure 9. Diachronic map of migrations in Asia ca. 2600-2250 BC. See full high-resolution version at <https://indo-european.eu/en/maps/chalcolithic/>............. 47

Figure 10. Modern distribution of haplogroup R1b1a1a2a1-L51*. Adapted from Richard Rocca (2012). ................................................................. 50
Figure 11. Modern distribution of haplogroup R1b1a1a2a2-Z2103. Adapted from Myres et al. (2011).


Figure 13. Modern distribution of haplogroup R1b1a1a2a1a2c1-R1b-L21. Adapted from Myres et al. (2011).

Figure 14. Modern distribution of haplogroup R1b1a1a2a1a2a-DF27. Adapted from Myres et al. (2011).


Figure 17. Modern distribution of haplogroup R1b1a1a2a1a1-U106. Adapted from Myres et al. (2011).

Figure 18. Diachronic map of migrations in Europe ca. 1250-750 BC (Butler, Arnoldussen, and Steegstra 2011/2012; Wels-Weyrauch 2011; Kristiansen 2000; Przybyla 2009), LDA-LSA. See full high-resolution version at <https://indo-european.eu/en/maps/late-bronze-age/>.

Figure 19. Modern distribution of haplogroup R1b1a1a2a1a2b-U152. Adapted from Myres et al. (2011).


Figure 22. Diachronic map of migrations in Europe ca. 250-750 AD. See full high-resolution version at <https://indo-european.eu/en/maps/antiquity/>.
Figure 23. Modern distribution of haplogroup R1a1a1b1a-Z282. Adapted from Underhill et al. (2015). 85

Figure 24. 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/>. 88

Figure 25. 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/>. 90

Figure 26. Diachronic map of migrations in Asia ca. 1250-750 BC. See full high-resolution version at <https://indo-european.eu/en/maps/late-bronze-age/>. 91

Figure 27. Diachronic map of migrations in Asia ca. 750-250 BC. See full high-resolution version at <https://indo-european.eu/en/maps/iron-age/>. 92

Figure 28. 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/>. 93

Figure 29. Diachronic map of migrations in Asia ca. 250-750 AD. See full high-resolution version at <https://indo-european.eu/en/maps/antiquity/>. 94

Figure 30. Modern distribution of haplogroup R1a1a1b2-Z93. Adapted from Underhill et al. (2015). 95

Figure 31. Diachronic map of migrations ca. 750 – 1300 AD. See full high-resolution version at <https://indo-european.eu/en/maps/medieval/> 99
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, 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, anthropological 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 (1953), impressively 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, taking 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 only, 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 groups 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 used¹ 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-Mesolithic

A sample from the 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 ancestry has been revealed to have contributed substantially to the genetic ancestry of Siberians, Native Americans, and Bronze Age Yamna individuals (Lazaridis et al. 2016), 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).

III.1.1. R1b-M343

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 Iboussieres. East hunter-gatherers from the Iron Gates prove the regional continuity of haplogroup R1b1a-L754 (xR1b1a1a-P297, xR1b1a1a2-M269) – i.e. probably from branches that have not survived in modern populations –, in 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 (Figure 2).

III.1.2. R1a-M420
Haplogroup R1a-M420 originated ca. 25000 BC, and has been proposed to diverge initially in the vicinity of present-day Iran based on a study of modern populations (Underhill et al. 2015). Hunter-gatherers of R1a-M420 lineage were proposed to have migrated from the Iranian area to the forests of Eastern Europe, since early samples were found in the Narva culture (Horvath 2015).

III.1.3. Younger Dryas and Boreal periods
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), and 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 this area.

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 around this time 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.

Samples of R1b1a1a-P297 (xR1b1a1a2-M269) 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).

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).
The oldest aDNA sample of R1a-M420 lineage found in east Europe was at Vasylivka, dated ca. 8690 BC, at the same site where a sample dated ca. 7250 BC was found of haplogroup R1b1a2-V88 (Mathieson et al. 2017).

III.1.4. Languages

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 R1b1a1a-P297 lineages to a territory of previous Caucasian hunter-gatherers. Such Caucasian influence has been supported recently by the finding of a genetic contribution of a pocket of Caucasus hunter-gatherers (ca. 11000–8000 BC), 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 R1b1a1a-P297 lineages. If these people and their language expanded from central and south-east European communities 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), based on 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). Also, haplogroup R2a-M124 seems to be prevalent among (ancient and modern) Dravidians, while haplogroup Q-M242 has links to Asian and Native American populations. There are thus potential links of linguistic macro-groups Eurasiatc, Nostratic (Bomhard 2008) and Borean (Gell-Mann, Peiros, and Starostin 2009) with certain expansions of P1-M45 lineages that could help determine dialectal evolutions, but
macro-languages are speculative and their relationships highly controversial, with such ancient archaeological evolutions – and their relationship to population movements – quite difficult to ascertain.

Figure 2. Diachronic map of Palaeolithic migrations (continued on the next page). See full version at <https://indo-european.eu/en/maps/palaeolithic/>. 
III.2. 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.

However, before the arrival of farmers from the Middle East (Olalde et al. 2015; Szecsenyi-Nagy et al. 2017), pottery was produced in the first half of the 7th millennium BC by hunter-gatherer groups in the Pontic-Caspian steppe, first in the Volga steppes (with the earliest pottery found in the Elshanian culture), probably derived from the Eastern Asian tradition of the Late Pleistocene through Siberia and the Transurals (Piezonka 2015).

To the south- and westward pottery spread into the Northern Caspian culture ca. 6500 BC, and then 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 Criș settlers from the Starčevo–Kőrösi–Criș 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 – 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 (Figure 3).

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), and 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 Vasylivka (see above) – there are samples dated ca. 5500-4800 BC: nine of R1b-M343 lineage, probably from an extinct branch of R1b1a-L754 (xR1b1a1a-P297, xR1b1a1a2-M269; one of R1a-M420 lineage; four of haplogroup I-M170, probably I2a2a1b-L701, in Deriivka; and another four samples of I-M170 and one IJ, probably also all I2a2a1b-L701, in neighbouring Volniensky (Mathieson et al. 2017). The diversity in lineages (all of them extinct) points to a mix in the different groups that emerged in the early Mesolithic period, before the mass expansions that occurred later.

Haplogroup R1a1a1-M417 (formed ca. 6500, TMRCA ca. 3500 BC) is first found in a Karelian hunter-gatherer dated ca. 6850-6000 BC (Lazaridis et al. 2016), and two samples of R1a1a1-M417 dated ca. 6125-4885 BC have also been found in Baikalic cultures at Irkutsk (Moussa et al. 2016) – near the zone where the ancient Mal’ta-Buret’ culture was located. Given the Eurasian origin of the eastern European pottery, and its western expansion into Europe, it seems logical to find a common origin of both populations 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 this area potentially fitting the available data.

III.2.1. Languages

The expansion of R1a1a1-M417 lineages may have disrupted the Early Proto-Indo-European R1b1a1a-P297 community thriving in east Europe. In this context, R1a-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 spoken in the Forest Zone (an equivalent to Early Proto-Indo-European in this macro-family proposal) by R1b1a1a-P297 communities similar to the Narva samples, developing a Proto-Uralic-speaking community.

Given the early sample of R1a-M420 in the Mesolithic north Pontic area – and maybe the rare subclade R1a5-Z645 found in Estonia (Saag et al. 2017) –, it would also be possible, although unlikely given the available archaeological data, to propose that only pottery was adopted from the east, and 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.

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).

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

Even though an aDNA sample of haplogroup N1a-F1206 is found in the Forest Zone dated ca. 2500 BC at Serteya (Chekunova et al. 2014), 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.

Figure 3. Diachronic map of Mesolithic migrations ca. 6500-5000 (Anthony 2007; Piezonka 2015), Uni-Köln (continued on the next page). See full high-resolution version at <https://indo-european.eu/en/maps/mesolithic/>. 
III.3. Chalcolithic

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. The revolution of herding, travel, and raiding – and thus the change in the steppe – came with horseback riding, appearing ca. 4800 in early Khvalynsk, and spreading south- and eastward. The early Sredni Stog culture began about 4400 BC, and it seems that people from the east Pontic Caspian steppe (related to early Khvalynsk) brought a new culture (Anthony 2007), and probably also their Middle Indo-European language.

Within this new culture, a new elite group associated with the Suvorovo-Novodanilovka complex (Anthony 2007) 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 4), and sites of the posterior Cernavodă I culture seem to represent the assimilation of migrants from the steppes, therefore linked to Anthony’s first expansion from the Pontic-Caspian steppes into southern Europe ca. 4200-4000 BC (Anthony 2007; Anthony 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 of the Mesolithic Dnieper-Donets culture (predating Sredni Stog in the same region), dated to a similar time and belonging to haplogroup R1a-M420 (Jones et al. 2017), 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.

Two individuals of haplogroups R-M207 and R1b1a-L754 (dated ca. 4500 BC) in Smyadovo and one of haplogroup R1-M173 (dated ca. 4460 BC) in Varna I cemetery (Mathieson et al. 2017) are not proven to correspond to R1b1a1a2-M269 lineages, and dates are slightly early for the mass migration proposed by Anthony (2007), so they could be assumed to correspond to old Balkan R1b1a1a-P297 branches. However, 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.

Admixture analysis shows steppe-related ancestry (defined as a mixture of European and Caucasus hunter-gatherer similar to individuals from Yamna) in these samples, contrasting with the sex-biased resurgence of hunter-gatherer ancestry in central Europe and Iberia during the Middle Neolithic period — also supported by haplogroup replacement (Mathieson et al. 2017).

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 occurred around this time period, probably from some eastern clans of Pontic-Caspian herders that developed 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 Proto-Anatolian (Kortlandt 1990; Ringe 2006) from a common Middle Proto-Indo-European language (Tischler and Oettinger 1989; Lehrman 1996; Melchert 1998).

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 5). Its modern peak around Kosovo can be explained by posterior founder effects that might have happened during any expansion of peoples in the region in the past four thousand years, and which can tentatively be assigned to a recent Albanian expansion. Its modern distribution in the Alps and in ancient Tyrrenhia might point to an eastern route of the Suvorovo-Novodanilovka settlers of eastern Hungary, hence giving support to the theories describing Etruscan as an Anatolian branch (Adrados 1989, 1994). On the other hand, it could well be a sign of independent back and forth migrations between the Adriatic Sea and the Italian Peninsula.

Figure 5. Modern distribution of R1b1a1a2-M269 (xL23) lineages, adapted from Richard Rocca (2012).
III.4. 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 V² 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, xR1b1a1a2a1-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.

² 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; Oettinger 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-Aryan Indo-European dialects (Adrados 1998; Mallory and Adams 2007; West 2007). In light of the most likely distribution of both dialects during the common Yamna period, the names ‘Western’ and ‘Eastern’ Late Proto-Indo-European would probably be more appropriate.

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 6).

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, and 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).

**III.4.1. 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, and 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).

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.

Figure 6. Diachronic map of Eneolithic migrations ca. 4000-3100 BC (Anthony 2007; Szmyt 2013; Piezonka 2015), Uni-Köln. See full high-resolution version at <https://indo-european.eu/en/maps/eneolithic/>.
III.5. 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% western hunter-gatherer, 30% eastern hunter-gatherer) 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% eastern hunter-gatherer), which proves that a westward migration of peoples accompanied cultural changes (Mathieson et al. 2017). The last sample obtained, 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 found not 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), which 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.5.1. Expansion of the Corded Ware culture

The first horizon of Corded Ware culture appears in the Early Eneolithic Bubanj-Salcuța-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 of the late 5th and middle 4th millennium, after the Suvorovo-Novodanilovka expansion, show steppe ancestry (see above). Also, a sample identified as outlier from Deriivka ca. 3500 BC shows both Caucasian hunter-gatherer and farmer ancestry (Mathieson et al. 2017), which illustrates the complexity of human interaction in this western Pontic region between the main (Middle and Late) Indo-European expansions.

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. The spread of this classical period of the Corded Ware culture has been connected to the evolution of late Funnelbeaker culture infiltrated by late Trypillian groups, after which they could have entered into contacts with Yamna herders on the upper Dniester region ca. 2700-2600 BC (Anthony 2007; Gimbutas 1977). However, no previous direct cultural connection has been found in this area with Yamna (Bulatović 2014).

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), but also encompassing the areas between the Vistula and the Dnieper (including the Małopolska area), with different connecting routes to the north used by Old European – and especially Trypillian culture – societies influencing Baltic cultures of the steppe, forest-steppe and forest zones for millennia (Klochko and Kośko 2009; Szmyt 2013; Czebreszuk and Szmyt 2004; Kadrow and Zakościelna 1999).

The connection between pre-Caucasian (Maykop) and Late Trypillian cultures that had moved to the left bank of the Dnieper (see above) 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 Gordinești group (Ivanova Svetlana and Toschev Gennadiy 2015). These migrations could account for the later steppe-related ancestry found in Corded Ware cultures (Allentoft et al. 2015; Haak et al. 2015; Mathieson et al. 2015), since it is defined by a certain admixture of Eastern European and Caucasian hunter-gatherer ancestry, not found in samples from western Yamna migrants. On the other hand, the expansion

The most recent connection of the north Pontic steppe to Central European areas came from Usatovo (which continued the previous Gordinești group), whose 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).
On the other hand, while Corded Ware culture research is discussed as a purely Central-European phenomenon, recently obtained dates suggest that the appearance of Corded Ware in central Russia (either of early Fatyanovo or maybe proto-stages) may have begun from 2700-2600 onwards, with eastern influence 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).

The most recent sample from Zvejnieki, dated ca. 2885 BC, just before or during the expansion of the third Corded Ware horizon, clusters quite closely with Yamna samples, revealing a break in the regional population with the samples from a thousand years earlier, which were closer to Eastern European hunter-gatherer ancestry (Mathieson et al. 2017).

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 7). 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 below).

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), with one sample in Bergreinfield (ca. 2647), two samples in Eulau (ca. 2600 BC), one sample3 from Kyndelöse (ca. 2670 BC), seven samples 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). Two samples from Tiefbrunn of haplogroups R-M207 (ca. 2755 BC) and

---

3 Published as corresponding to Nordic Middle Neolithic culture, additional information by Vladimir Tagankin revealed a branch typical of modern Nordic (R1a1a1b1a-Z284) subclades, and a new date of ca. 2475 BC, including reduction for high marine signal.
R1b1-L278 (ca. 2725 BC) are of dubious nature – 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.

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 Amphorae (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).

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 (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).

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 east Europe in a relatively short period. This might have provided part of the ‘push’ for the migration of Corded Ware cultures (Anthony and Brown 2017), and might also account for part of the documented differences in population expansion between Corded Ware and Bell Beaker groups, and their demographic consequences.

III.6. Tocharian expansion

The Early Bronze Age Afanasevo culture (ca. 3500-2600) in the Altai-Sayan region has been found to be 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) 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) were found to 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 8, Figure 9). A division between a Mongol Altai and a Siberian area of Indo-European influence was clearly seen in ancient DNA samples from a later period, where only one sample from the Okunev culture (ca. 2300-1800 BC) was of R1b1a1a2-M269 ancestry, with three samples of haplogroup N1a1-M46 (see above) 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

---

4 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.

*Figure 8. Diachronic map of migrations in Asia ca. 3100-2800 BC. See full high-resolution version at <https://indo-european.eu/en/maps/copper-age/>.***
Figure 9. Diachronic map of migrations in Asia ca. 2600-2250 BC. See full high-resolution version at <https://indo-european.eu/en/maps/chalcolithic/>.
III.7. Late Indo-European expansion

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

The most obvious material division within the early Yamna horizon was between east and west (see above Figure 7). 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 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 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, pushing farther up the Danube to the middle Danube valley in eastern Hungary through an Old Europe in crisis – contemporary with late Baden / Cernavodă III (Anthony 2013; Anthony 2007).

III.7.1. 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 7). 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 southwestern 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 –, and it recovered after 3000 BC with a more humid climate that favoured grassland productivity (Harrison and Heyd 2007), accelerating after 2500 BC, which is compatible with the expansion of the horse, the wheel, and pastoralist societies 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 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 (like figurines) in some local traditions of Western Europe seem to predate that of the Yamna horizon, and might have been used quite differently, or for different purposes, in certain local cultures (Robb 2009; Diaz-Guardamino 2014). In fact, it seems to be the destruction of (early) Bell Beaker stelae what marks the arrival of the
East Group of Bell Beakers in Sion, associated with a later expansion of Yamna migrants (Heyd 2012).

Support for the western migrations 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), a subclade 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 R1b1a1a2a1-L51* (Figure 10) in Central Europe all suggest a western migration of peoples from the Yamna culture – and mainly western R1b1a1a2a1-L51 lineages – along the Danube.

Figure 10. Modern distribution of haplogroup R1b1a1a2a1-L51*. Adapted from Richard Rocca (2012).
III.7.2. 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.

R1b1a1a2a2-Z2103 lineages appear in modern populations (Figure 11) 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; and 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.7.2.1. Greek

The southern route of the southern stream of Anthony’s third migration wave has been described as the later expansion of the Yamna culture 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).

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 (which in turn might be related to the central group), but no subclades were given in the study (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 point to an early expansion of R1b1a1a2a2c1a-Z2110 lineages, but the early Anatolian influence over the island precludes a precise identification of their origin.

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 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.

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) – 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 Balkan (IE) individuals, and R1a1a1b2-Z93 probably from eastern (Iranian) migrants.

### III.7.2.2. 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 resurge 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 Armenian, this has 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)\(^5\), suggesting the presence of previous R1b1-L278 lineages in the region – possibly from southward Mesolithic migrations from the Pontic-Caspian steppes –, unrelated to the later Armenian migration.

Also, 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), which is also hindered by the loss of data in modern populations due to the effects of the Armenian Genocide.

\(^5\) 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."
Figure 11. Modern distribution of haplogroup R1b1a1a2a2-Z2103. Adapted from Myres et al. (2011).
III.7.3. The expansion of North-West Indo-European in Europe

III.7.3.1. Bell Beaker

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 and plates, and 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 7). This “proto-package” is found for example in the Maritime Beaker, and expanded ca. 2700-2500 and got 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, with specific correspondences 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 Csepe around Budapest, west of the East Hungarian settlement region of Yamna migrants, and dated ca. 2800-2600, could have been the direct bridge between Yamna on the east and Moravia and Bavaria to the west, through which 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 12).

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 based 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 2014).

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 R1b1a1a2a1a2a-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.

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. There are data suggesting rejection and aversion, but also some form of social discourse between the groups. 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 2014). 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.

Therefore, while 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 –, 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).

Bell Beaker expansion

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

Samples of north-east Iberia, dated ca. 2385, show haplogroups R1b1a-L754 (xR1b1a1a2a-L23), G2-P287, and I2a2-M436, which point to the persistence of old European hunter-gatherer lineages in the region.

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 R1b1a1a2a-P312 lineages are found in Osterhofen ca. 2540 BC, and in Sierentz – Les villas d’Aurèle ca. 2430 BC. The oldest sample of haplogroup R1b1a2a1a2b-U152 is found in Budapest ca. 2335 BC, of haplogroup R1b1a2a1a2a-DF27\(^6\) in Quedlinburg ca. 2290 BC (Lazaridis et al. 2016), and of haplogroup R1b1a2a1a2c-L21 in Amesbury ca. 2290 BC (Olalde et al. 2017).

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 Rhine individuals, and not to the Atlantic façade of western Europe (Olalde et al. 2017). Demic diffusion of R1b1a1a2a2c1-L21 lineages (Figure 13) 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 R1b1a2a1a2c-L21 (or older haplogroups). 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 R1b1a2a1a2c-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 R1b1a2a1a2c1-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  

\(^6\) 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.
exchange of female marriage partners (Brodie 2001) or inter-cultural contact consolidation (Vander Linden 2007), as previously proposed.

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), pointing 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 14).

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.

Two late samples from northern Iberia were found to have a small proportion of steppe ancestry, which supports the early spread of Indo-European speaking peoples into Iberia as early as the beginning of the Bronze Age (Olalde et al. 2017). 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 on the pre-Celtic nature of the languages of Cantabri, Astures, Pellen dones, Carpetani, and Vettones. Also, while the position of Tartessian as Indo-European (Koch 2009) is highly doubted7 – , there is some support for a borrowing of names from a “lost Indo-European language” over the course

7 It was criticized extensively in a special section of Vol. 42 of The Journal of Indo-European Studies (No. 3 & 4, Fall/Winter 2014)
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 (Szecsényi-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.

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, Bell Beaker samples of R1b1a1a2-M269 are found from 2400 BC to 2300 BC, and later an eastern sample is found near the Oder ca. 2170 BC (see below), which – together with long-term and long-distance economic exchange (especially regarding amber imports) during the Bronze Age (Makarowicz 2009) – support the presence of Old European river names in east Europe.
Figure 13. Modern distribution of haplogroup R1b1a1a2a1a2c1-R1b-L21. Adapted from Myres et al. (2011)

Figure 14. Modern distribution of haplogroup R1b1a1a2a1a2a-DF27. Adapted from Myres et al. (2011)
III.7.3.2. 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 15). 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
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.7.3.3. Nordic Bronze Age

In Scandinavia, farming communities had already abandoned their subsistence strategy for the development of transhumance (Jensen 2003). A migration of Bell Beaker groups to Jutland (ca. 2300-1700 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 large chiefly houses similar to Únětice appeared in south Scandinavia, with a growth in interregional contacts. All these changes are interpreted as “a radical reorganization of economy and social organization which wiped out or 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).

The best candidate for an expansion of the Pre-Germanic dialect of North-West Indo-European into Scandinavia is the Barbed Wire Beaker culture of the Low Countries and Northern Lowland (Kristiansen 2009), which would later show a period of change (Figure 16) starting ca. 1850 BC until its 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 I2-M438 was formed ca. 25500 BC, and the modern European population has a TMRCA ca. 19900 BC. Many ancient DNA samples 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. Samples from Sweden since the Mesolithic show exclusively I2-M438 lineages before the Bronze Age.

It has been asserted, based on all samples studied from the Palaeolithic, that western hunter-gatherers (as defined in admixture analysis) represent a population that expanded
from a south-eastern European refugium following the last Ice Age around 15000 years ago – displacing or admixing with the existing population of western Europe (Mathieson et al. 2017)

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 (Szecsenyi-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 Skole8, ca. 1410 BC) and late Nordic Bronze Age (Angmollan, ca. 670 BC), while haplogroup I1-M253 is found in Angmollan (ca. 1400 BC), and haplogroup I-M170 in Angmollan (ca. 1360) 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 in 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) – 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. If invasions from southern to northern Scandinavia are supposed to have happened in a south-to-north route, through the Øresund strait into the Skåne region, peoples of I1-M253 lineages should be assumed to have migrated from northern Europe during the proposed expansion of Beaker peoples into Scandinavia, and thrived after that, both populations pushing back the previous Corded-Ware-associated R1a1a1b1a3-Z284 lineages, which in turn had probably displaced or replaced the earlier Neolithic I2-M438 lineages.

It is difficult to ascertain whether both lineages, R1b1a1a2a1a1-U106 and I1-M253, were already mixed in northern Germany before their northward migration into Jutland, or remained separated until forming a Pre-Germanic community later. If an early mixed R1b1a1a2a1a1-U106–I1-M253 society with a common language is to be supported, it seems to need further explanations as to the clear late differentiation into territorially-
divided lineages, with late founder effects having simplified to a greater extent the situation east and west of the Øresund strait.

After an obscure period of internal development, the situation in Northern Germany and Scandinavia before the Iron Age would have probably corresponded loosely to the present situation, with the R1b1a1a2a1a1-U106 / I1-M253 divide possibly located to the east of the current cline, at the Øresund strait, 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 22).

The modern distribution of R1b1a1a2a1a1-U106 (Figure 17) 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 31).

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 17. Modern distribution of haplogroup R1b1a1a2a1a1-U106. Adapted from Myres et al. (2011).
III.7.3.4. 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 in 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, with Únětice daggers 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 in fortified structures (see above Figure 16).

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; followed up by more massive colonizing 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 2014).
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 more aDNA samples and the application of admixture analysis.

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 18). 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-M459⁹, and one of haplogroup R1b-M343 (Schilz 2006).

Given the modern distribution of R1b1a1a2a1a2b-U152 lineages (see below Figure 19), its expansion is probably to be connected to the spread of the Urnfield culture and later offshoots Hallstatt and Villanovan cultures.

---

Fig. 18. Diachronic map of migrations in Europe ca. 1250-750 BC (Butler, Arnoldussen, and Steegstra 2011/2012; Wels-Weyrauch 2011; Kristiansen 2000; Przybyla 2009, LDA-LSA. See full high-resolution version at https://indo-european.eu/en/maps/late-bronze-age/).

⁹ Reported as Ri and Ri?
III.7.3.4.1. 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 Rhine in the west, and a Moravian zone in the east – has been linked to the spread of Celtic languages (Figure 20). 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 19), 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)10.

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, with a I2a2a1a1-M284 lineage concentrated in Great Britain (with mutational divergence suggesting its foundation ca. 300 BC) providing “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. 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.

---

10 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 also 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 19. Modern distribution of haplogroup R1b1a1a2a1a2b-U152. Adapted from Myres et al. (2011).

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 with Italic 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 historical development 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, a recent Anatolian connection has been found by examining mtDNA in modern populations of present day Tuscany (Brisighelli et al. 2009).

The expansion of Rome (Figure 21) 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 21. Diachronic map of migrations in Europe ca. 250 BC – 250 AD. See full high-resolution version at <https://indo-european.eu/en/maps/classical-antiquity/>
III.7.3.4.3. 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, with Dutch-German lowland areas found to share cultural roots with the southern Scandinavian area (Butler, Arnoldussen, and Steegstra 2011/2012) predating technologic and economic exchanges between Urnfield and Northern Bronze Age Scandinavia (Kristiansen and Suchowska-Ducke 2015).

These complicated cultural-economic networks (see above Figure 16 and Figure 18) 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 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.8. Indo-European in Corded Ware societies

III.8.1. 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 23).

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; whereas R1a1a1b1a2b-CTS1211 (or M558) seems to be centred on Eastern Europe.

Samples of R1a1a1-M417 and R1a1a1b-Z645 lineages from ancient populations and admixture analyses suggest an original spread of European hunter-gatherer ancestry eastward from the late Pit-Comb ware culture (Mathieson et al. 2017), a later expansion of steppe ancestry associated with Corded Ware cultures (Mathieson et al. 2015; Haak et al. 2015; Allentoft et al. 2015), probably from the Contact Zone between Globular Amphorae, eastern Baltic, Old European (especially Usatovo), and later contacts with Yamna migrants up the Prut River (see above Figure 6).

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 16), which later expanded east (ca. 1300-500 BC) into territories of previous Trzciniec culture (see above Figure 18).

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.

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.

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 20). 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, and 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 21).

East of the Przeworsk zone was the Zarubinets culture (3rd c. BC – 2nd c. AD), considered a part of the Przeworsk complex (Mallory and Douglas 1997), located between the upper and middle Dnieper and Pripyat rivers. Early Slavic hydronymy 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 Douglas 1997) – descended about the 5th c. (Figure 22).
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 Scytho-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 Douglas 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), which 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.

Figure 22. Diachronic map of migrations in Europe ca. 250-750 AD. See full high-resolution version at <https://indo-european.eu/en/maps/antiquity/>.
Figure 23. Modern distribution of haplogroup R1a1a1b1a-Z282. Adapted from Underhill et al. (2015).
III.8.2. 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 Amphorae traits, substituting the Volosovo culture and occupying the Volga-Kama region. Near it the Balanovo group seemed to be 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 24).

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 R1a1a1b-Z645 is found later in the Potapovka culture, in Utyevka ca. 2200-1900 BC (Mathieson et al. 2015).

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 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), unlike other samples of R1b1a1a2a2-Z2103 lineage found in the same area, shows that 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. “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).

Ancient DNA samples of Sintashta, Potapovka, and later of the Srubna culture (Figure 25) show a substitution of R1b1a1a2a2-Z2103 lineages by R1a1a1b-Z645 lineages, and recent studies show continuity in steppe admixture in all cultures from the region, Corded Ware, Yamna, Potapovka, and Sintashta (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 evident 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, 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.
Difficult to determine is 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 –, although they probably spoke Indo-European languages of the Graeco-Aryan group.

Cimmerian or Thraco-Cimmerian groups might have emerged from societies related to these western groups of Pontic-Caspian herders. 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.

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 25 to Figure 29).
Figure 25. 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/>. 
Figure 26. Diachronic map of migrations in Asia ca. 1250–750 BC. See full high-resolution version at <https://indo-european.eu/en/maps/late-bronze-age/>. 
Figure 27. Diachronic map of migrations in Asia ca. 750-250 BC. See full high-resolution version at <https://indo-european.eu/en/maps/iron-age/>.
Figure 28. 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/>. 
The modern distribution of R1a1a1b2-Z93 lineages (Figure 30) 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 its 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.

Figure 30. Modern distribution of haplogroup R1a1a1b2-Z93. Adapted from Underhill et al. (2015).
III.8.3. 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 with 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, and thus their similarities could be explained as a common language substrate, whether non-Indo-European, Pre-Indo-European, or even Indo-European.

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 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.

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 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.

On the special position of Balto-Slavic and Indo-Iranian, regarding their rather conservative nominal case system compared to other Indo-European languages, 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), whereas the Indo-Iranian and Balto-Slavic ones 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.
Figure 31. Diachronic map of migrations ca. 750 – 1300 AD. See full high-resolution version at <https://indo-european.eu/en/maps/medieval/>. 
III. 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.

III.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; Anthony 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; Anthony 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–satem 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 (Uralic) languages of east 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.
III.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 colonization 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 of a migration. Multiple models have been proposed, but all offer a weaker potential explanation for linguistic change than demic diffusion or founder effect, since it implies 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 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.

### III.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); ancestral populations have to be correctly identified, including their number and precise alleles (Jobling et al. 2014). 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, 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 admixture of Yamna aDNA samples found elevated (up to 76%) in Corded Ware samples has been said to support the migration of Yamna populations into Corded Ware groups, while the lower percentage 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 Globular Amphorae culture do not show evidence of steppe 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 Amphorae 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):

- Samples from the Pontic-Caspian steppe – from which ‘steppe ancestry’ is defined as a precise combination of West hunger-gatherer and Caucasian hunter gatherer ancestry – are scarce, and recent ones only from one eastern region (Kalmykia).
Steppe ancestry has been found in Corded Ware, Afanasevo, Andronovo, and Srubna cultures, all of which – and even a late individual of Bronze Age Bulgaria from Merichleri\textsuperscript{11}, ca. 1690 BC – show higher steppe ancestry than samples clearly identified as from Yamna migrants in the Balkans. 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).

The north-west Pontic area – from where many Yamna migrants seem to have expanded west along the Danube – had been a zone of interaction with peoples from the Danube and the Eastern Baltic for millennia – and could thus cluster closer genetically to peoples from Carpathian cultures than the eastern Pontic-Caspian steppe. Peoples from this eastern zone, whose samples are used to define steppe ancestry, had migrated to the north-west Pontic area at least twice, first in the formation process of the early Khvalynsk-Sredni Stog cultures, and later in the formation of the Yamna culture (see above), probably creating a more mixed western genetic pool, more similar to the one found in western Yamna migrants.

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\textsuperscript{12}. More recently, Estonian samples have shown a genetic component associated with Caucasus hunter-gatherers coinciding with the spread of R1a-Z645, which rules out Corded Ware and Yamna as the only origin of this component (Saag et al. 2017). Supporting these conclusions, a sample of a female from Zvejnieki, dated ca. 2885 BC and classified as from Latvian Neolithic cultures (Jones et al. 2017), has been found to cluster closely with Yamna samples (Mathieson et al. 2017), closer than samples from Corded Ware cultures, and before any described migration from Yamna could have happened to this region. These recent samples question the validity of assuming a direct gene flow from Yamna migrants when explaining the so-called steppe ancestry found in samples from Corded Ware cultures.

\textsuperscript{11} 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.
\textsuperscript{12} 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).
• 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 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.

• Corded Ware samples are late, almost coinciding with the Bell Beaker expansion. No samples have been published from potentially controversial areas – like the Contact Zone, eastern Baltic 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 structure or socioeconomic factors – may falsely show a relatively homogenous high or low ancestral contribution (Jobling et al. 2014).

• 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)\(^{13}\), 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, Günther, et al. 2017).

• Ascertaining 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). Also relevant is then the actual increase in population due to such expansions – greater in south-east Europe (Müller 2013) –, which further influences the genetic drift observed.

• 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

\(^{13}\) These results could not be replicated in a later study (Lazaridis et al. 2016), but the authors have further explained their methods (Goldberg, Günther, et al. 2017).
to improvements in such methods), groundbreaking conclusions, and self-important titles, with a tendency to “culture-historicism”. This trend is very well represented and remembered in anthropological disciplines by Nature’s paper on the Anatolian origin of Proto-Indo-European, based on glottochronology, by Gray and Atkinson (2013).

- 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 ethnic definition is almost non-existent. This, together with the perceived complexity (and lack of familiarity with) intricately linked anthropological disciplines, has made admixture analysis quite popular among amateur geneticists, who can easily play with published open source software programs, 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 ethnic 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.
III.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.

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, and archaeology and genetics can only add precision to such studies.

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. 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 Szecsenyi-Nagy, Mari Törv, Svend Hansen, Kirsten I. Bos, Philipp W. 115


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


Ivanova Svetlana, V., and N. Toschev Gennadiy. 2015. The Middle-Dniester Cultural Contact Area of Early Metal Age Societies. The Frontier of Pontic and Baltic Drainage Basins in the 4Th/3Rd-2Nd Millennium BC. In *Baltic-Pontic Studies*.


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.


127
Metspalu, M. Meyer, E. E. Eichler, J. Burger, M. Slatkin, S. Paabo, J. Kelso, D. Reich,
and J. Krause. 2014. Ancient human genomes suggest three ancestral populations for

Lazaridis, Iosif, and David Reich. 2017. Failure to Replicate a Genetic Signal for Sex Bias in the
Steppe Migration into Central Europe. *bioRxiv*.

Lee, E. J., C. Makarewicz, R. Renneberg, M. Harder, B. Krause-Kyora, S. Muller, S. Ostritz, L.
Emerging genetic patterns of the European Neolithic: perspectives from a late Neolithic


Li, C., H. Li, Y. Cui, C. Xie, D. Cai, W. Li, V. H. Mair, Z. Xu, Q. Zhang, I. Abuduresule, L. Jin,
H. Zhu, and H. Zhou. 2010. Evidence that a West-East admixed population lived in the
Tarim Basin as early as the early Bronze Age. *BMC Biol* 8:15.

Linden, Marc Vander. 2015. What linked the Bell Beakers in third millennium BC Europe?
*Antiquity* 81 (312):343-352.

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

Lucotte, Gérard. 2015. The Major Y-Chromosome Haplogroup R1b-M269 in West-Europe,
Subdivided by the Three SNPs S21/U106, S145/L21 and S28/U152, Shows a Clear

Maeir, Aren M., Brent Davis, and Louise A. Hitchcock. 2016. Philistine Names and Terms Once
Again: A Recent Perspective. *Journal of Eastern Mediterranean Archaeology &

Major, Candace O., Steven L. Goldstein, William B. F. Ryan, Gilles Lericolais, Alexander M.
Piotrowski, and Irka Hajdas. 2006. The co-evolution of Black Sea level and composition
through the last deglaciation and its paleoclimatic significance. *Quaternary Science
Reviews* 25 (17–18):2031-2047.


Olalde, Iñigo, Selina Brace, Morten E. Allentoft, Ian Armit, Kristian Kristiansen, Nadin Rohland, Swapan Mallick, Thomas Booth, Anna Szécsényi-Nagy, Alissa Mittnik, Eveline Altena, Mark Lipson, Iosif Lazaridis, Nick J. Patterson, Nasreen Broomandkhoshbacht, Yoan


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


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


Scheeres, Mirjam. 2014. High mobility rates during the period of the “Celtic migrations”? <sup>87</sup>Sr/<sup>86</sup>Sr and δ<sup>18</sup>O 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.