

# **Redefining Pre-Indo-European Language Families of Bronze Age Western Europe: A Study Based on the Synthesis of Scientific Evidence From Archaeology, Historical Linguistics and Genetics**

*Csaba Barnabas Horvath, PhD*  
Corvinus University of Budapest, Hungary

Doi:10.19044/esj.2019.v15n26p1

[URL:http://dx.doi.org/10.19044/esj.2019.v15n26p1](http://dx.doi.org/10.19044/esj.2019.v15n26p1)

---

## **Abstract**

When and how did the Indo-European language family expand into western Europe? What language families were present there before, and where did they come from? Where does Basque and Etruscan come from? These are questions and puzzles that have been in the focus of the archaeological community since long ago. The present paper offers a coherent hypothesis mapping the expansion of language families in the western half of Europe from the Copper Age till the Roman conquest, based on matchings identifies between the newest genetic evidence, and earlier results of archaeology and historical linguistics. The approach focuses on matching phylogenetic and geographical distribution patterns of Y-DNA lineages with archaeological cultures, and the phylogenetics of language families in order to identify migrations and language families in Western Europe in the 3<sup>rd</sup>-1<sup>st</sup> millennia BCE.

---

**Keywords:** Indo-European, pre-Celtic, Basque, Etruscan, Vasconic, Tyrsenian

## **Introduction**

This paper makes an attempt to identify how Indo-European expansion took place in the western half of Europe, and what major pre-Indo-European language families may have been present in the region, and aims to answer this question by a synthesis of archaeological and linguistic data with the newest genetic results. As my earlier paper, “Reconsidering The Geographic Origins Based On The Synthesis Of Archaeological And Linguistic Evidence And The Newest Results In Genetics – A Finno-Scythian Hypothesis” addressed the same issue regarding the Northern and Eastern parts of Europe, this paper can be seen as a sequel of that one, covering the western half of Europe, with the two papers combined mapping a full picture of Bronze Age

pre-Indo-European language families from Iberia to the Urals. As it will be presented in this paper, certain matchings with archaeological, linguistic and genetic evidence all matching each-other suggests a scenario that coherently answers questions regarding the origins of both of the prehistoric languages families Vasconic and Tyrsenian, as well as defining the start of Celtic expansion, all of these as part of a coherent scenario, and this scenario also forming an even broader, but still coherent one and single scenario combined with the one presented in my aforementioned previous paper regarding Northern and Eastern Europe.

The methodology of this paper is identical with that of my earlier paper „Reconsidering The Geographic Origins Based On The Synthesis Of Archaeological And Linguistic Evidence And The Newest Results In Genetics – A Finno-Scythian Hypothesis” , and it can be read in detail there (Horvath, 2019). To summarize the methodology: The method used in both paper is a synthesis of scientific evidence from archaeology, historical linguistics, and genetics in a way that it is focusing on finding matchings between these three sets of data regarding prehistoric migration or conquest events. Then patterns among individual migration events are analyzed to see whether if patterns can be found that suggest certain individual migration or conquest events form larger chains and networks of interconnected or consecutive migrations. Then if such a chain or network of apparently interconnected migrations appears to be linked to a culture where we already have sufficient historical documentation to determine its’ ethnolinguistic identity, then that can possibly suggest the given network of interconnected migrations telling us something about the prehistoric past of the given language family, where a high degree of caution is also needed within the methodology regarding any traces of possible language shifts. Within genetic data, the approach presented here focuses mostly but not exclusively on Y-DNA lineages, called haplogroups due to their non-recombining nature. Literature marks major haplogroups by an alphanumeric system, where, for example R1 and R2 are primary sub-lineages, so called subclades of their parent lineage R, and R1a and R1b are subclades of R1. As the alphanumeric system adds one more character at each additional degree, would mean impractically long codenames for subclades of several degrees, further sub lineages are often marked by the codename of the marker defining them linked to the alphanumeric code of their parent group, such as R-L51 or R1b-L51. The focus on Y-DNA is also due to the fact that data in distribution of genetic lineages suggests that in the preindustrial world, men usually took part in migrating or conquering groups in a larger part than women did, and from conquered cultures, women managed to integrate in the new society ran by the conquerors in a larger part, than men of conquered cultures did, as demonstrated along examples in detail in the aforementioned earlier paper of mine. Therefore, among the non-

recombining genomic elements, statistical distribution of Y-DNA haplogroups tended to change more as an impact of migrations and conquest, than the distribution of mtDNA haplogroups, and thus function as a more practical tracking indicator for mapping migrations. This approach used in both papers however does not automatically identifies certain Y-DNA lineages with certain language families, as bottleneck effects and language shifts can cause different Y-DNA lineages being dominant within the same linguistic community in different phases of its' history. The focus instead is rather to identify individual chapters of migration with the help of the newest available Y-DNA data, and then as described above, try to see whether if such migrations constitute a broader network or not, regardless if different sections of such a chain of migrations were characterized by different Y-DNA haplogroups due to bottleneck effects. For further details on my research methods, see methodology described in detail in my previous paper on the issue, „Reconsidering The Geographic Origins Based On The Synthesis Of Archaeological And Linguistic Evidence And The Newest Results In Genetics – A Finno-Scythian Hypothesis” (Horvath, 2019).

The issue of the spread of the Indo-European language family has been analyzed by several notable scholars. While Gimbutas locates the original homeland, “urheimat” in the Pontic Steppe (Gimbutas, 1970), Renfrew puts it to Anatolia (Renfrew, 1990), and Gamkrelidze and Ivanov to the Armenian highlands (Gamkrelidze & Ivanov, 2010). As described in my aforementioned, earlier paper, my hypothesis contradicts to the hypothesis of Gimbutas, but overall fits the hypotheses of Renfrew, Gamkrelidze and Ivanov, as it will be further discussed in this paper as well (Horvath, 2019). Regarding the issue of pre-Indo-European language families in Western Europe, we can mention the Vasconic substratum hypothesis of Theo Vennemann (Vennemann, 2003) that proposes Vasconic, the hypothetical language family of Basque to have covered Iberia, Gaul and Britain before the Celtic takeover, and the Tyrsenian hypothesis of Helmut Rix, that proposes a language family named “Tyrsenian”, including Etruscan of Central Italy, Raetic of the Alps, and Lemnic, the pre-Greek language of Lemnos (Rix, 1998).

### **Defining the geographical range analyzed in the present paper**

The range analyzed in this paper is the range covered by the Bell-Beaker Culture (2800-1800 BCE), that is basically overlapping both in time and space the range of the R1b-P312 subclade of haplogroup R1b. Italy is also included in the analysis, as it later became covered by the Villanova culture being an offshoot of the Urnfield culture, that had its' core in the Bell Beaker range, and its' population has been confirmed to have been dominated by the presence of R1b subclades, the R1b-U152 subclade being the most dominant (Grugni et al. 2018). Italic languages, the first documented Indo-European

languages are within the family closely related to Celtic that later settled the original Bell Beaker range, one more reason to include Italy as well. Thus, geographically this paper covers the Bell Beaker range and Italy, what basically means all of Europe west of the Hamburg-Trieste line. As the issue of Germanic and Balto-Slavic languages and their assumed pre-Indo-European substrate has been addressed in my aforementioned paper those issues are not addressed here.

Vast majority of today's population of the region is dominated by haplogroup R1b, therefore suggesting a large-scale population exchange to have happened in the region when this haplogroup was introduced. As the most recent common ancestor of R1b-P312 is dated to 2500 BCE, that of R1b-Z2118, also showing a distribution overlapping the Bell Beaker range to 2700 BCE, and their common parent group, R1b-L52 to 3700 BCE, (YFull, 2019) this evidence suggests that most of today's Y-DNA lineages of the range were introduced by a migration that established the Bell Beaker culture. Regarding Britain, ancient DNA evidence also confirms large scale migration introducing the Bell Beaker culture in the region (Callaway, 2017). R1b-S1161, the most widespread subclade of R1b-Z2118 also has its' most recent common ancestor dated to 2700 BCE (YFull, 2019).

### **R1b-P312: Introduced by Indo-Europeans or pre-Indo-Europeans?**

It would be tempting to simply identify Indo-European with the Bell-Beaker population characterized by R1b-P312, but such a guess faces several problems, for which it doesn't offer any answers:

One such major anomaly is the high and diverse presence of pre-Indo-European languages in the R1b-P312 range up until the Roman conquest. Iberian in the eastern parts of Spain, Aquitanian, the ancestor of Basque in Southwestern France and northern Spain, Etruscan in northern Italy and Raetic in the alps are all examples for that, out of which Basque survived up until today. Besides them, other languages at the time of Roman conquest within the R1b-P312 range were also suspected, though not proven to be Indo-European, such as Pictish in the northern part of Scotland, and Ligurian in Northwest Italy. To make the issue even more complicated, the two most well documented pre-Indo-European languages of the R1b-P312 range, Basque and Etruscan have been confirmed to be unrelated even to each-other.

The other problem is the relative low diversity of Indo-European languages within the R1b-P312 range: In the evolution of language families, we can see a general rule, that language families being present longer usually become more diverse and fragmented by time, than language families that arrived more recently. In the R1b-P312 range however, while documented pre-Indo-European languages show an extremely high diversity, to such a degree that Etruscan and Basque are even confirmed to be unrelated to each-other, all

Indo-European languages of this range all belong to the Italo-Celtic branch of that family, or to be even more specific on the issue, all documented pre-Roman Indo-European language outside Italy within the R1b-P312 range is actually Celtic, a single branch of Indo-European that shows way too much uniformity to have been present for close to 3000 years before the Roman conquest.

To summarize, if the Bell Beaker migration was already Indo-European, then Iberian and Basque languages would have long been extinct by the time of Roman conquest, and after almost three thousand years of divergence, the Romans would have found in their place Indo-European languages as distant not only from Latin, but from each other as well, as Sanskrit from Latin is. If on the other hand, the Urnfield migration was already Indo-European, then Etruscan and Raetic languages couldn't have survived to see the Roman conquest, and Indo-European languages spoken in their area would have been as distant from Latin, as Germanic is. Thus we can assume the Bell Beaker population to have spoken pre-Indo-European languages, and as human populations of the range show a close demographic interconnectedness at the time of the expansion of said culture due to the distribution of R1b-P312, we can assume these to have belonged to one single language family.

### **Major pre-Indo-European languages within the R1b-P312 range**

Basque language, the only pre-Indo-European language that has survived in the western half of Europe up until today is the most appalling candidate for being a member of the original Bell Beaker language family. German linguist Theo Vennemann proposed a Basque linguistic substratum in Atlantic Europe, covering the Iberian Peninsula, Gaul and the British Isles, and calling this language family Vasconic (Vennemann, 2003), which hypothesis is in accordance with genetic evidence shown in the distribution of Rb-P312, and archaeological evidence in the form of the Bell Beaker culture. Iberian language, spoken in the eastern parts of the Iberian Peninsula at the time of the Roman conquest has also been suggested to be related to Basque by analysis of the language itself as well (Orduna-Aznar, 2005), further confirming Vennemann's theory at least in the Iberian Peninsula. As the distribution of R1b-P312 in time and space and that of the Bell Beaker culture is evidence to sufficient demographic cohesion for the Bell Beaker range to sustain a single language family, this scenario seems to be highly plausible, especially that R1b-L21 and R1b-DF27, the two subclades of R1b-P312 that are dominant in Britain and the Iberian Peninsula, also have their most recent common ancestor dated to the middle of the third millennium BC (YFull, 2019).

Vasconic or not, pre-Indo-European presence in pre-Celtic Britain is strongly suggested by linguistic evidence (Hickey 2002).

The other most well documented pre-Indo-European language of the R1b-P312 range is Etruscan, however represents a much more difficult case than Basque. The paradox of Etruscan language has been a puzzle for scholars since roman times: The language is apparently unrelated not only to Indo-European but to Vasconic as well. The population itself however shows a high presence of R1b-U152, suggesting Central European links, and this is not only true for the present population in the region (Hay, 2017) but archaeological findings also confirmed to have been true to the ancient Etruscan population as well (Grugni et al. 2018). Some antique scholars suggests the Etruscans to have arrived along a maritime route from the Aegean (Turfa, & De Grummond, 1986), but the problem with this scenario is that Raetic language, spoken in the Alps up until the time of Roman conquest, has been confirmed to had been related to Etruscan (Rix, 1998), and due to its' location in the most remote valleys between the highest ranges of the Alps, Raetic could have hardly arrived on maritime route from the Aegean, which again suggests Central European links in accordance with the high presence of R1b-U152. At the same time, J2a-M67\*, suggested to represent maritime links to ancient Aegean cultures, represents only 2.7% of all ancient Etruscan DNA samples (Grugni et al. 2018). Therefore, sea migration from the Aegean to Etruscan lands, if any, was more likely that of a minor superstrate population only. If Etruscan language was the language of the native population of that part of Italy, characterized by a high presence of R1b-U152, and actually a majority presence of Central European haplogroups, such as all R1b subclades and G2a-L497, making up no less than 56.9% of ancient Etruscan DNA samples combined (Grugni et al. 2018), and another genetic study also reached similar conclusions (Perkins, 2017) then why is it (as well as Raetic) unrelated to Vasconic? To make things even more complicated, Etruscan and Raetic have been suggested to be part of a broader language family together with the pre-Greek language of Lemnos Island, called Tyrsenian (Rix, 1998).

Besides Basque, Etruscan, Iberian and Raetic, several other languages in the pre-Roman R1b-P312 range are possibly pre-Indo-European, as due to the lack of sufficient and conclusive evidence, their classification whether if Indo-European or not, hasn't been determined yet:

Tartessian, the ancient pre-Roman language of Andalusia is debated to have been Celtic, Vasconic, or a language isolate (Michalowski, 2017) (Koch, 2009) (Trask, 1995). Ligurian, the pre-Roman language of Liguria, and possibly that of Piemonte as well, is also debated to have been Celtic, Vasconic or language isolate (Malkiel, 1952) (Delamarre, 2003). In case of Pictish, the language dominant in Scotland north of the Forth-Clyde line before shifting to Gaelic and Scots, the undetermined question is basically also whether if it was

Celtic, Vasconic, or neither of the two. (Rhys, 1898) (Vennemann, 2003) (Watson, & Taylor, 1926)

As we can see, the list of confirmed and potential pre-Indo-European languages within the R1b-P312 range that survived up until the Roman conquest is extensive: Etruscan and possibly Ligurian in the Northern half of Italy, Raetic in the Alps, Basque in Southwest Gaul and Northern Iberia, Iberian in Eastern Iberia, possibly Tartessian in Southern Iberia as well, and possibly Pictish in Northern Britain, with their combined areas covering a significant portion of the R1b-P312 range, clearly suggesting a massive pre-Indo-European presence among populations otherwise characterized by the overwhelming majority of R1b-P312 in their Y-DNA composition, raising more than reasonable doubts regarding the possibility of this subclade first being spread by Indo-European expansion.

### **Consecutive waves of prehistoric migrations – arrival of the Tyrsenian language family and the Celts**

To solve the problem of this picture of language families, we take an overview of prehistoric migrations suggested by genetic evidence to have occurred after the original Bell Beaker migration and try to draw a conclusion of what implications they have on language families. Subclades with chronological and spatial distribution patterns suggesting such secondary migrations, can be found both among the subclades of R1b-U152, J2b, and E-V13

Distribution of R1b-U152 shows a pattern suggesting it playing a key role in the formation of the Tumulus (1600-1200BC) culture. Out of its' twelve identified primary subclades, eight were identified to have had their most recent common ancestors between 2000 and 2400 BCE, virtually all showing their base being in Central Europe. Subclades of U152 seem to be responsible for introducing the Urnfield culture to Italy and thus establishing the Villanova culture, with its' tertiary subclade R1b-Z145 having its' most recent common ancestor in 1600 BCE being widespread in Italy, and matching in time the Urnfield culture (Horvath, 2019). At the same time, recent findings in ancient DNA confirmed about half of Etruscan samples identified being R1b, and about half of those belonging to R1b-U152 (Grugni et al. 2018).

Certain subclades of R1b-L2, one of the primary subclades of R1b-U152 at the same time show such a distribution on the peripheries of the Celtic world, that suggests them having spread with the La Tene culture (450 BCE- 1 BCE) with four of its' primary subclades having their most recent common ancestors after 1100 BCE, being widespread at such peripheries of the Celtic World as the British Isles (R1b-FGC11577 and R1b-A197 ), Italy (R1b-FGC5354) the British Isles and the Iberian Peninsula (R1b-Y61958) (YFull, 2019).

It is interesting to note that most lineages located exclusively in the British Isles or Iberia, either have their last common ancestor dated before 2000 BC, or after 1000 BC, but seldom between the two (YFull, 2019), suggesting the Atlantic Bronze Age range being populated by the presumably Vasconic-speaking Bell Beaker population, was left largely intact by the presumably Tyrsenian-speaking Urnfield expansion, and remaining Vasconic-speaking till the Celtic takeover in the middle of the first millennium BCE.

This data shows evidence for consecutive migrations linking the Tumulus, Urnfield, Hallstatt and La Tene cultures to populations characterized by the subclades of R1b-P312, including Etruscan, confirmed to have not just cultural, but direct demographic and genetic links to the Urnfield Culture with the presence of R1b-Z154 subclade of R1b-U152, and ancient DNA also showing high presence of R1b-U152. All this however it doesn't explain the three distinct language families (Vasconic, Tyrsenian, and Indo-European – the latter represented by Celtic) all present inside the R1b-P312 range. If the population that established the Bell Beaker culture and was characterized by the spread of R1b-P312 was speaking Vasconic languages, then without external incursions, all these cultures from Tumulus to La Tene should have been Vasconic as well, even if distinct branches of that. Language shifts to entirely different language families could have come only with external incursions, and with all these cultures spreading from east to west, such incursions must have come from the east, and at the right time to play a role in the formation of those cultures. Can we see any signs of such migrations in the genetic evidence?

Genetic evidence suggests two such incursions to have happened. The first one is an event suggested by the evidence to have happened in the middle of the second millennium BC: Among the subclades of J2b, J2b-Z2507 shows distribution patterns in time and space that strongly suggests it to be connected with the Tumulus and Urnfield cultures. First, its' related lineages are mainly located in Southeast Europe, and its' most recent common ancestor being identified to have lived around 2400 BCE, Second, out of its' two primary subclades in 2200 BC (for Z638) and 1900 BC (for Y15085), and three out of four of its' secondary subclades being dispersed in Central Europe (Z1297, Z38240, Z40052), and one in the Balkans (Y21045). Third, most of its' lineages spread over Central and Western Europe originate from tertiary and quaternary subclades with most recent common ancestors identified between 1000 BCE and 1900 BCE (YFull, 2019). This all suggest J2b-Z2507 to represent an incursion to the upper Danube watershed from the Northern Balkans at the time of the formation of the Tumulus culture (1600 BCE) playing a key role in the formation of that culture. Tumulus culture is the direct predecessor of the Urnfield culture, and the Etruscan civilization emerged in the area of the Villanova culture that is linked to the Urnfield culture not only

by cultural links but by subclades of R1b-U152 both confirmed by ancient DNA findings (Grugni et al. 2018) and phylogenetic patterns of R1b-U152 subclades suggesting for example R1b-Z145 to have evolved at the right time. These traces in Y-DNA haplogroup distribution combined with the fact that the range of Raetic language in the Alps were also located within the Urnfield range, this event apparently offer a plausible scenario for the problem of the Etruscan language: Tyrsenian languages being introduced to the upper Danube watershed by this incursion from the east by a population characterized by the presence of J2b-Z2507 subclades, occurring in the middle of the second millennium BC, establishing the Tumulus culture, that later evolved into the Urnfield culture, and making the Vasconic-speaking majority native population of the upper Danube watershed to shift to the language of the conquerors, what appears to have been Tyrsenian. This explains why Etruscans, though characterized by the presence of R1b-U152, spoke a language unrelated to Vasconic, and it also explains, why the Raetians, who in the Alps lived in valleys of the tributaries of the Danube river, spoke a language related to Etruscan. This solves the problem of the Etruscan and Raetic languages, Etruscan descending from the Tyrsenian language spoken by the Urnfield population establishing the Villanova culture, and Raetic, with river valleys in its' area mostly being open to the north, descending from the Tyrsenian languages spoken by Urnfield populations in the upper Danube watershed. Vasconic languages at the same time apparently survived intact in the range of the Atlantic Bronze Age. The area of Iberian language in Eastern Iberia, while partially included in the Urnfield range, lacks a significant R1b-U152 presence, or presence of any other haplogroup suggesting demographic change between the Bell Beaker and the Roman period, thus the introduction of the Urnfield culture there likely happened by cultural diffusion only, explaining why a Vasconic language survived there.

A second incursion apparently occurred almost a millennium later, also arriving to the upper Danube watershed from the east: Among the subclades of E-V13, distribution patterns of subclade E-CTS9320 suggests a migration from the Balkans to Central and Western Europe in early to middle part of the first millennium BCE, exactly fitting the formation of the Hallstatt C, D, and La Tene cultures: Its' most recent common ancestor is dated around 1000 BCE, and three out of its' main primary subclades, E-Y20805, E-Z17264 and E-Z17107 are apparently based on the Balkans. The fourth of those, E-Z16988 however is spread all over Central Europe, with its' own subclades Z38664 and Z27131 each having its' most recent common ancestor in 500 BCE and 900 BCE suggesting the move taking place in the first half of the first millennium BCE (YFull, 2019). With all its' relatives being present on the Balkans, this strongly suggests E-CTS9320 representing the incursion establishing the La Tene culture and possibly already later phases of the

Hallstatt culture, being responsible for the language shift of the population of the Hallstatt culture from Tyrsenian languages of the upper Danube watershed (what were most likely more closely related to Raetic than Etruscan) to Celtic. Then Celtic languages were further expanded to the British Isles, Gaul, and Iberia by assimilated groups of the local population, marked by the secondary expansion of certain R1b subclades from Central Europe, such as those subclades of R1b-L2 mentioned before.

The issue of the Italic languages suggests a scenario that may seem surprising at first glance. Before the Roman conquest of the Etruscan civilization, the range of Italic languages was limited to Central and South Italy, south of Tuscany. At the same time, these regions of Italy are characterized by a striking low presence of R1b-P312 and R1b-U152 subclades, and R1b in general (Hay, 2017). Had Italic languages arrived from the north, the population bringing them should have been characterized by a high presence of R1b-U152, and it is also problematic, how could they have reached their present range without conquering the Etruscans. A plausible scenario solving this paradox is that Italic languages arrived to Italy from the south, directly crossing from the Balkans across the Strait of Otranto. This would easily explain both the lack of R1b-U152 among them, and Etruscan domination in the north. The viability of such a migration route is proven by the presence of Illyrian speaking tribes in Southeast Italy in early Roman times, such as the Iapygians (Talbert, 1985 p.85.). With Celtic languages arriving to the upper Danube watershed from the Northern Balkans, this suggests the pre-Illyrian Western Balkans being the urheimat of Italo-Celtic languages, and these languages splitting as the ancestors of Celts moved to the northwest along the Danube, while ancestors of the Italic speaking population moving to the southwest crossing the Strait of Otranto to Italy. As South Italy is abundant in the high presence of J1 and J2 subclades that have apparently arrived in consecutive waves of migrations from the Aegean and the Balkans occurring throughout prehistory all the way from the Neolithic to the Iron Age, there is no lack of genetic evidence supporting such a scenario.

Thus how we can link language families to cultures in the R1b-P312 range is the following: Bell Beaker culture was apparently Vasconic, and probably had the edge over earlier population of the region due to skills in metallurgy, and the establishment of this culture occurred with a large scale demographic replacement of the earlier native population by the conquering R1b-P312-characterized Bell Beaker people. Tyrsenian languages were apparently introduced from the Balkans by an incursion of a conquering population characterized by the presence of J2b-Z2507 that established the Tumulus culture, thus the Tumulus, Urnfield, Villanova and most likely Hallstatt A and B cultures were speakers of Tyrsenian languages. For hypothesized position of language families around 1800 BCE, right before

Tyrsenian populations established the Tumulus culture, see Map 1 (All maps in this paper are own work by the author of this paper, Csaba Barnabas Horvath, using as background: Terpischores 2012. „*Biomes of the World-The 14 biomes of the world according WWF*”. Categorized as „Labeled for reuse with modification” by Google).



At the same time, unlike the establishment of the Bell-Beaker culture, this language shift occurred with no major population replacement, an overwhelming majority of their population were direct descendants of the native Bell Beaker people characterized by subclades of R1b-U152 (with the Etruscan Villanova culture being an exception in a way that it was outside the original Bell Beaker range, and Tyrsenian languages in Italy were introduced by a large scale migration of Bell Beaker-related, but already Tyrsenian speaking Urnfield people characterized by the presence of the R1b-Z145 subclade of R1b-U152). Vasconic languages at the same time, survived in Atlantic Europe beyond the western borders of the Urnfield culture, still being predominant in the Atlantic Bronze Age cultural range, and also in the Urnfield-influenced parts in eastern Iberia, where the lack of U152 subclades from the era suggests Urnfield influence being present due to a mere cultural diffusion instead of migration. Then, Celtic languages were apparently introduced to the upper Danube watershed by a migration again arriving from the Northern Balkans, characterized by the presence of E-CTS9320, occurring sometime in the early half of the first millennium BC, again characterized by the lack of a major demographic replacement of the native population, but by a mere language shift of the majority native population instead, due to the influence of a conquering minority. It is difficult to identify the exact time of this conquest, but the most apparent cultural and technological shift in the region that would explain an edge sufficient for such a conquest was the

introduction of iron when Hallstatt C replaced Hallstatt B. Thus having the technology of iron use and iron weapons is highly likely candidate for such an edge, suggesting the scenario of proto-Celtic population arriving from the Northern Balkans introducing iron use in the region, and taking over the Tyrsenian speaking Hallstatt B culture by conquest, and their impact being strong enough for the majority native population to shift to their Celtic language. For hypothesized position of language families around 900, before proto-Celtic establishment of the Hallstatt C culture, see Map 2 (All maps in this paper are own work by the author of this paper, Csaba Barnabas Horvath, using as background: Terpischores 2012. „Biomes of the World-The 14 biomes of the world according WWF”. Categorized as „Labeled for reuse with modification” by Google).



Out of the earlier Tyrsenian Urnfield languages of the region, Raetic language as a relic survived in the mountains of the Alps up until roman conquest, and so did Etruscan in Italy, having better access to new technologies through the Mediterranean Sea. Also sometimes at the turn of the second and first millennia BC, the population speaking the Italic languages crossed from the Balkans to South Italy through the Strait of Otranto.

### The Unetice culture and the question of Germanic languages

An issue linking the range analyzed in this paper to the aforementioned paper of mine is the question of the Unetice culture (2300-1600 BCE), that located in the Eastern areas of what is today Germany, and also the Czech Republic and thus apparently occupied the core of what later became the Iron Age Germanic areas. With its' most recent common ancestor dated to 2700 BCE (YFull, 2019), haplogroup R1b-U106 apparently performed the first wave of its' expansion in the 3<sup>rd</sup> millennium BCE, making it a contemporary

of the Corded Ware and Bell Beaker cultures. Here too, a secondary expansion of its' subclade R1b-Z19 indicates a later expansion that could be responsible for the introduction of the Germanic languages, but the migration that can be associated with the original introduction of R1b-U106 is third millennium BCE, making it likely to have been pre-Indo-European. Its' geographic distribution and its' timing however shows the most matches with the Unetice culture, especially given the fact that most of its' subclades present in Sweden and Norway have their respective most recent common ancestors dated to the late Bronze Age. This combined with the fact that similar to the R1b-P312 expansion, the expansion of R1b-U106 and that of the Unetice culture also started in the Carpathian Basin, makes it likely to have been Vasconic as well, apparently settling in these territories (but not in Scandinavia) over the R1a-Z284 dominated Corded Ware culture, that in turn is suggested to have been Finno-Ugric in my aforementioned paper (Horvath 2019),

A new wave of divergence among key haplogroups in Northern Europe suggests migration introducing the Germanic languages into the region: While R1a-Z284 and R1b-U106 seem to have reached the region early with the Corded Ware and Unetice migrations, and thus likely had been Finno-Ugric and Vasconic, R1b-Z19, subclade of R1b-U106 with its' most recent common ancestor dated to 1700 BCE, that of its' most widespread primary subclade, R1b-FGC79182 to 1500 BCE and its' distribution covering most of the Germanic world, suggests to be linked to a secondary migration in the late Bronze Age. Similar is the case with Q-L804, apparently surviving as a hidden bottleneck for millennia, being formed in 13200 BCE but having its' most recent common ancestor dated only to 1200 BCE also showing a distribution covering most of the Germanic world (YFull, 2019). The lineage that may form a link to the Northern Balkans conquered by the presumably Indo-European mid-Bronze Age migration from Anatolia is R1a-L664, with its' most recent common ancestor dated to 2000 BCE, and those of its' two widespread primary subclades, R1a-S2857 and R1a-S2894 dated to 1400 BCE and 1200 BCE (YFull, 2019). This all suggests a secondary migration from the south in the late Bronze Age, introducing Germanic languages replacing Vasconic in what are today regions of Germany, the Czech Republic and Poland, and replacing Finno-Ugric in Scandinavia, probably with the formation of the Lusatian culture (1300 BCE) and the Late Nordic Bronze Age (1100 BCE). This would also explain the ambiguity of the pre-Indo-European substratum of Germanic languages (Wiik 1997), if at first in former Unetice areas it absorbed as substrate such a version of Vasconic that in turn had a Finno-Ugric substrate on its' own, and then when it later expanded to Scandinavia with already having this Vasconic substrate, it absorbed a secondary, Finno-Ugric substrate there. (Especially if these Finno-Ugric substrates in these areas in turn had a Funnelbeaker substrate on their own,

unlike the ancestors of Finno-Permian and Ugric that had a Pit Comb Ware substrate, making linguistic differences even more significant. (Horvath 2019))

### **Secondary findings: the role of Anatolia**

As secondary findings, we can see, genetic evidence confirms these consecutive waves of migrations arriving from the Balkans. This, in accordance to my findings presented in my earlier paper “Reconsidering The Geographic Origins Based On The Synthesis Of Archaeological And Linguistic Evidence And The Newest Results In Genetics – A Finno-Scythian Hypothesis” that identified Yamna and Corded Ware cultures as being Indo-European, and thus points to the Balkans as the entry point of Indo-European expansion to Europe (Horvath, 2019), also confirms the role of the Balkans as a cradle of prehistoric migrations and language families, apparently due to its’ geopolitical position as the bridge where all new technologies in metallurgy and agriculture must have passed through on their way from the highly advanced societies of Anatolia and the Fertile Crescent to Europe. Uralic language family originating from Neolithic Northeast Balkans, Vasconic and Tyrsenian language families having their origins somewhere in either the Neolithic Aegean, Southern Balkans, or Neolithic Western Anatolia, and the Indo-European language family having its’ urheimat probably somewhere in or near Eastern Anatolia in the 5th-4th millennium BC, as suggested by Gamkrelidze and Ivanov (2010), a location also supported by the linguistic results of Krell (1994). In my aforementioned earlier paper, a possible scenario was also included explaining how the population speaking the ancestors of Indo-Iranian and Tocharian languages likely spread eastwards through the Iranian plateau from the Indo-European urheimat in Eastern Anatolia (Horvath, 2019). Analysis of DNA patterns among human populations speaking the Northwest Caucasian and Northeast Caucasian languages on the northern slopes of the Caucasus also suggest these populations originating from the ancient Middle East (Yunusbayev et al., 2011). These trends combined constitute a complete hemisphere of consecutive prehistoric migrations through all possible routes to the West, North and East from the Fertile Crescent, through the Anatolia-Balkans route as well as crossing the Caucasus, and through the Iranian Plateau basically verifying the farming-language dispersal hypothesis of Colin Renfrew (Bellwood & Renfrew, 2002). As hypotheses presented in this paper and my aforementioned paper both fit in the farming-language dispersal hypothesis, this further enhances their likelihood as part of a bigger picture formed by the farming-language dispersal process.

Indo-European was apparently the last of such prehistoric migration waves reaching Europe from the Balkans, arriving from Anatolia. The most

likely candidate for such an arrival of Indo-Europeans from Anatolia is the Middle Bronze Age Migration of 1900 BCE suggested by James Mellaart (Horvath, 2019). This also solves the problem of the Lemnic language, spoken on the Aegean island of Lemnos near the coasts of Thracia as recently as the first millennium BC before the population shifted to Greek, but likely being part of the Tyrsenian family together with Etruscan and Raetic: As Tyrsenian languages appear to have arrived to the Tumulus and Urnfield range from the Balkans, a plausible solution for Lemnic being part of the family, is that before the arrival of the Indo-Europeans from Anatolia in 1900 BCE, and Tyrsenian expansion into the upper Danube watershed in 1600 BCE, the range of Tyrsenian languages was covering much of the Balkans all the way from Thracia to Pannonia. The Indo-European expansion reaching the Balkans from Anatolia in 1900 BC perfectly fits to such a scenario, with Tyrsenian languages being predominant in much of the Balkans including Thracia up until at least 1900 BC. The Indo-European migration mostly affected the southern and eastern parts of the Balkans at first, occupying these areas by the middle of the second millennium BCE, being able to introduce the core populations of future Balto-Slavic and Germanic speaking populations to the southernmost limits of the Corded Ware range at this time, while Tyrsenian languages most likely survived in the Western Balkans at this point, having been replaced by Indo-Europeans only by the end of the second millennium BCE. As E-V13 apparently arrived to Europe before this migration (most recent common ancestor dated to 2800 BCE) the apparent introduction of Celtic languages to the upper Danube watershed by E-CTS9320 doesn't mean, that Indo-Europeans had arrived to Europe with the migration that brought E-V13 in, but rather that that part of the population of the Balkan from which E-CTS9320 arrived to the upper Danube, was assimilated already in the Balkans by the original Indo-European conquerors. Haplogroups showing distribution patterns that confirm this migration event from Anatolia to the Balkans suggested by Mellaart to have indeed happened, and seem to have participated in it, are J1-L818 with its' most recent common ancestor dated to 1900 BC and G-S9591 most recent common ancestor dated to 1800 BC (YFull, 2019). With chronology matching, and all of them being widespread in the Balkans, having their closest sibling groups in Anatolia, the migration suggested by Mellaart seems to have indeed happened, and was likely performed by a human population characterized by the presence of these two groups.

Phylogenetic distribution of haplogroups R1b, R1a, J2b, J2a, J1 and E-V13 in Europe also support such an “Out of the Balkans” (and Anatolia) scenario:

The first R1b subclade having all its' subclades apparently spread over Europe is R1b-L51 (earliest common ancestor 3700 BC). Its' sibling subclade, R1b-Z2103 however already shows signs of originating in Anatolia: Although

now it is mostly widespread in Eastern Europe and Central Asia, in these regions it is represented by a single subclade of it, R1b-Z2106, while its' older subclades such as R1b-Y13369 and R1b-Y4363 are all present in Anatolia and the Caucasus, suggesting it originating from Anatolia. This in turn suggests R1b-L23, the common parent group of R1b-L51 and R1b-Z2103 originating in Anatolia as well, which is further confirmed by the fact that its' own sibling subclade, R1b-PF7562 is also concentrated in Anatolia, the Balkans and the Caucasus region, thus suggesting an Anatolian location for R1b-M269, the common parent group of the latter two as well. As R1b-M478, the sibling group of R1b-M269 is also concentrated in the Caucasus, this suggest Anatolia for the homeland of their common parent group, R1b-P297 as well. R1b-V1636, the sibling group of R1b-P297 in turn is found in the Middle East, suggesting their parent group, R1b-L389 originating from the middle east, especially underlined by the fact that R1b-V88, the sibling group of R1b-L389 shows the highest variance in North Africa and the Middle East. As R1b-L389 and R1b-V88 represent the first split of R1b (YFull, 2019), that suggests that the original homeland of R1b was located in the Levant.

The case is somewhat similar regarding R1a. While R1a does have some older lineages in Europe, it's main lineage now widespread in Europe, R1a-M417 also has two main lineages, R1a-CTS4385 and R1a-Z645, with the former showing a distribution west of the North Balkans, and R1a-Z645 showing a distribution indicating a base of expansion east of the North Balkans, also suggesting a point of separation somewhere in the North Balkans during the Neolithic. Thus, while R1a likely arrived to Europe in or before the Neolithic, its' R1a-M417 and its' subclades show signs to have taken the advantage of spreading all over the continent by joining Copper and Bronze age waves in the North Balkans. If we look a step further back, we can also see that the situation is similar regarding R1a, as out of its' two primary subclades, R1a-YP4141 (TMRCA 10200 BCE) and R1a-M459 (TMRCA 12000 BCE), each of them has one primary subclade widespread in Europe, and another one widespread in the Middle East, with their TMRCA-s between 6600 BCE and 3500 BCE. As all cultural flow, and as shown by the distribution of other haplogroups, all demographic flow in this period took place from Anatolia to Europe, and not the other way around, this also suggests that haplogroup R1a also originates from the Middle East. (YFull, 2019).

Ancient subclades of both R1b and R1a could have reached Europe even earlier of course, however R1b-L51, R1b-Z2106, and R1a-M417 seem to be the descendants of groups that arrived from Neolithic Anatolia.

The phenomenon is even more apparent regarding J1, J2 and E-V13, all three of them having their related subclades in the Middle East and North Africa (YFull, 2019).

The apparent high linguistic diversity of populations spreading out to virtually all parts of Europe from Anatolia and the Balkans during the 4<sup>th</sup> and 3<sup>rd</sup> millennia BCE as demonstrated in this paper and my aforementioned previously published other paper, is also a puzzle for which clues can possibly be found in the phylogenetic structure of haplogroups being present in the region in the given time. Uralic spreading out from what is today Ukraine with the Afanasievo, Corded Ware and Andronovo cultures s demonstrated in my earlier paper (Horvath, 2019), Vasconic spreading out from the Northwest Balkans with the Bell Beaker culture and Tyrsenian spreading out from the Balkans with the Tumulus and Urnfield cultures, and Indo-European from Anatolia through the Balkans after 1900 BCE, as demonstrated above in these three latter cases represent at least four distinct language families. See map 3 (All maps in this paper are own work by the author of this paper, Csaba Barnabas Horvath, using as background: Terpischores 2012. „*Biomes of the World-The 14 biomes of the world according WWF*”. Categorized as „Labeled for reuse with modification” by Google).



Phylogenetic structures of haplogroups widespread in the region show most of them having arrived to the Balkans from Anatolia in the given period, and the timing of the most recent common ancestor of each of them (indicating when they diverged from their sibling groups that stayed in Anatolia) can help us to have some guesses regarding possible waves of migrations from Anatolia to the Balkans in the given period. As we could see, J1-L818 has its' most recent common ancestor dated to 1900 BC, while G-S9591 to 1800 BC, indicating these two crossing from Anatolia to the Balkans at the beginning of the second millennium BCE, in accordance with the migration hypothesized by Mellaart. Then going further back in time, signs of the first previous possible major wave are R1b-PF7563 having its' most recent common

ancestor dated to 2600 BCE, E-V13 with its' own dated to 2800 BCE, and J2b-Z615 with its' own at 2900 BCE (YFull, 2019), pointing to the earlier half of the third millennium BCE. Right before this, we can see a major cluster at the middle of the fourth millennium BCE, with the following haplogroups and their most recent common ancestors: J2b-Z615 (3400 BCE), R1b-Z2106 (3500 BCE), G-37368 (3500 BCE), J1-Z2331 (3600 BCE), G-Z6029 (3600 BCE), and R1b-L51 (3700 BCE) (YFull, 2019). Thus between (and including) the arrival of R1b-L51 apparently being responsible spreading Vasconic languages in Western Europe, and the migration suggested by Mellaart apparently signaling the arrival of Indo-Europeans and marked by J1-L818 and G-S9591, we can see at least three major clusters. One in the middle of the 4<sup>th</sup> millennium, one in the first half of the 3<sup>rd</sup> millennium BCE, and one at the beginning of the second millennium BCE possibly being sufficient to introduce at least three different language families in the region, in case each represented a different one, what is a plausible explanation if Neolithic Anatolia was as diverge in language as the Caucasus is today. The apparent problem of R1b-Z206 that characterized the population that apparently started the Uralic language family as demonstrated in my earlier paper (Horvath, 2019) and R1b-L51 characterizing the population that apparently started the Vasconic language family as demonstrated in this paper, being genetically closely related, but representing two unrelated language families can be explained in two likely ways: One is that while one of the two carrier populations retained its' original language brought from Anatolia, the other one adopted a native Neolithic language in Europe (while this would at least at one point require a bottleneck effect to allow the native population to retain such a numerical majority that would have enabled them to assimilate the new arrivals). The other one is that as R1b-L51 and R1b-Z2103 (parent group of R1b-Z2106) separated in 4100 BCE, more than 500 years before the apparent time of these events, and there is a 200 years gap between the most recent common ancestors of their branches that arrived to the Balkans, it is also possible that the cluster in the middle of the fourth millennium BCE may represent two distinct waves of migrations closely following each-other but representing two distinct language families. This would also assume that by assimilation or some other event, R1b-L51 and R1b-Z2103 became part of two distinct populations belonging to two different language families, what occurred still in Anatolia half a millennium before the migration to the Balkans started. As some linguists suggests Basque (or Vaconic) and Kartvelian (or South Caucasian) languages to be distantly related to each-other, and we can see subclades of haplogroup G, characteristic to Georgia, to be present in this mid-4<sup>th</sup> millennium cluster, closer to its' earlier phase, this could mean R1b-L51 becoming part of such a population in Anatolia, while R1b-Z2103 remaining outside of it.

The earliest significant wave of migration and likely a group of languages that seem to have spread out from the Neolithic Middle East through Anatolia is the actual first wave of Neolithic expansion to Europe on what the original farming-language hypothesis is based. As described earlier, the migrations that spread R1b-L51, R1b-Z2106 and R1a-M417 were already apparently post-Neolithic. We can also identify however the Y-DNA haplogroups that seem to mark the original Neolithic expansion. Here the pattern that we can see suggests a scenario where agriculture was introduced to Europe by invading groups from Anatolia, but then soon groups of the native population were assimilated by it as well, and then the bulk of the demographic expansion that spread agriculture was done by these assimilated native groups. The two haplogroups that show matchings that suggest them signaling the movement of early-assimilated native populations that performed a demographic expansion introducing agriculture are I2-L158 (9700 BCE) and I2-L701 (7900 BCE) (YFull 2019). Among this two, I2-L158 seems to represent the Mediterranean Neolithic, spreading from Anatolia along the northern coastline of the Mediterranean Sea, while I2-L701 seems to represent the Danubian Neolithic, spreading north and west from the Aegean Sea along the river valleys of the Vardar and the Morava, and from the Morava-Danube confluence, upwards along the Danube river. The haplogroups on the other hand, that represent the link with Neolithic Anatolia, the demographic expansion of Neolithic Anatolian Farmers into Europe (defining them by the split of subclades the spread in Europe from their sibling subclades that remained widespread in Anatolia and the Middle East) with the dating of their respective most recent common ancestors are J1-Y136727 (6400 BCE) J1- Z18463 (5100 BCE) J2a- PF4888 (6000 BCE) R1a-M198 (6600 BCE), G-L91 (8400 BCE) G-F872 (7000 BCE) and G-PF3345 (8700 BCE) in the area of the Danubian Neolithic, while G-Z6885 (9300 BCE) G-V18939 (6700 BCE) J2a- Y13128 (6800 BCE) and J2a- Z438 (7100 BCE) in the area of the Mediterranean Neolithic (YFull 2019). Due to the high diversity of these haplogroups apparently taking part in this Neolithic expansion, it is impossible to tell whether if the language families of these first Neolithic pioneers of Europe were closely related to any known language family or not. Languages of probably Neolithic origins that may have survived long enough to be recorded by the Romans may include Sicanian, the pre-Indo-European language of Western Sicily, the Ilienses and Balares of Sardinia, the Corsi of Corsica, the Talaiotic natives of the Balearic Islands, and last but not least the Tartessians of what is today Andalusia. The pre-Vasconic Neolithic origins of Tartessians is strongly suggested by the fact that while Bell Beaker culture in general seems to have spread into Western Europe from the east by a migration characterized by R1b-P312, copper use in the Tartessian areas already started earlier than in areas where the migration characterized by the R1b-P312

expansion already arrived earlier. Therefore, copper use must have reached the Tartessian areas before the arrival of the population carrying R1b-P312, but if it did so, then it likely provided an edge to this native population to retain its' identity and language.

These geographic patterns of phylogenetic distribution also strongly contradict to the Kurgan hypothesis of Gimbutas, as they only suggest west-to-east migrations in the Pontic Steppe throughout the 4<sup>th</sup> and 3<sup>rd</sup> millennia BCE, and they show no traces whatsoever to any east-to-west migration in the region in this time period, and this is also supported by analysis of Yunusbayev et al. (2011) that got to the conclusion that most migrations to the steppes arrived from the Balkans, while migrations from the Middle East across the Caucasus only reached as far as the areas of the Northwest Caucasian and Northeast Caucasian languages. When talking about phylogenetic distribution patterns of Y-DNA, it is also worth to note, that two subclades, J2a-Y33844 and J2a-Z7429 also have their origins in the Fertile Crescent, and their subclades present in India have their most recent common ancestors in the early- to middle 2<sup>nd</sup> millennium BCE (YFull, 2019), fitting the time of the Indo-Aryan conquest, suggesting a scenario of getting there from the Indo-European urheimat in Anatolia through the Iranian Plateau, and assimilating significant populations characterized by the presence of R1a-Z93. Secondary findings described in this chapter could be topic of another research paper but had to be addressed here briefly due to the coherence of evidence.

## **Conclusion**

As we can see, the findings demonstrated in this paper map a single coherent scenario explaining the distribution of, and demographic relations between the Vasconic, Tyrsenian, and Celtic languages in the R1b-P312 range. This scenario combined with the scenario regarding Uralic languages demonstrated in my earlier paper (Horvath, 2019), in turn from a broader, but still again single and coherent scenario explaining the distribution of pre-Indo-European language families in all of Europe from the latter half of the 4<sup>th</sup> millennium BC till the start of recorded history. This conclusion was reached by identifying matchings between archaeological, linguistic and genetic results as shown in the table below:

<b>Table 1.</b> Proposed events, and matching in archaeological, genetic and linguistic evidence supporting them				
Proposed event	Time	Archaeological evidence	Genetic evidence	Linguistic evidence
Expansion of Vasconic languages with the Bell Beaker culture	starting by 2800 BCE	Bell Beaker culture 2800-1800 BCE	Age and geographical range of R1b-P312 matching that of the Bell Beaker culture	Basque language with R1b-P312 dominance among Basques. Corpus of the Iberian language showing resemblance with Basque, dominant in R1b-P312 dominant areas; Vasconic substratum hypothesis of T. Vennemann
Tyrsenian-speaking population arriving to the upper Danube watershed from the North Balkans, establishing the Tumulus culture (causing language shift among the native population there), later expanding with the Urnfield culture	starting by 1600 BCE	Villanova culture being an offshoot of the Urnfield culture, Raetic language also located within the Urnfield range	The distribution patterns of J2b-Z2507 showing a migration from the Balkans at the middle of the second millennium BC R1b plurality among Etruscan samples with Urnfield-related R1b-U152 majority among them; Urnfield-related R1b-U152 majority in former Etruscan areas up until today.	Raetic language of the Alpine area of the upper Danube watershed being related to Etruscan, and pre-Greek language of Lemnos near Thracia also being related to them, as described in the Tyrsenian hypothesis of H. Rix.
Early Celtic arriving to the upper Danube watershed from the North Balkans establishing the Hallstatt C culture, having iron use as their edge (with the majority native population of the area adopting the language of the minority conquering population), expand with the Hallstatt D and La Tene cultures	starting by 800 BCE	Introduction of iron use at the start of the Hallstatt C culture, Celtic expansion confirmed by historical sources along the La Tene expansion	The distribution of E-CTS9320 suggesting a migration to have arrived to the upper Danube watershed from the Balkans at the beginning of the first millennium BC. Certain subclades of R1b-P312 indicating a secondary expansion to Atlantic Europe during the first millennium BC	Celtic languages spread over the La Tene area

To summarize findings of this paper is that the Bell Beaker culture represented the Vasconic language family. The establishment of the Tumulus culture on the other hand was apparently marked by the introduction of Tyrsenian languages from the Balkans, (where they were most likely widespread at the time) and Tyrsenian languages were further spread by the Urnfield culture to those areas where this culture expanded by mass migration, most notably the Etruscan language of the Villanova culture of Central Italy. Vasconic languages however continued to be dominant in the Atlantic Bronze Age cultural range, that remained beyond the limits of Tyrsenian speaking Urnfield expansion. Indo-European languages then seem to have crossed to the Balkans from Anatolia around 1900 BC, and their Celtic branch took over the upper Danube Watershed by the establishment of the Hallstatt C culture, and their Italic branch crossed over to South Italy from the Balkans through the Strait of Otranto sometime at the turn of the second and first millennia BCE. Combined with findings presented in my paper “Reconsidering The Geographic Origins Based On The Synthesis Of Archaeological And Linguistic Evidence And The Newest Results In Genetics – A Finno-Scythian Hypothesis” (Horvath, 2019) where the main conclusion is that Corded Ware and Yamna cultures were Uralic, this constitutes an overall coherent structure of pre-Indo-European language families and migrations for prehistoric Europe, from the Iberian Peninsula all the way to the Ural Mountains, findings reached by combining the most recent genetic evidence with evidence reached by archaeology and historical linguistics. This also strongly suggests the Indo-European language family originating from somewhere in Eastern Anatolia in the 5th-4th millennium BC, as hypothesized by Gamkrelidze and Ivanov, and supported by the linguistic findings of Krell (1994), and reaching Europe via the Balkans at the turn of the third and second millennium BCE by the migration suggested by Mellaart.

This scenario of the Balkans being home to several consecutive migrations is a confirmation of the farming-language dispersal hypothesis of Colin Renfrew (Bellwood & Renfrew, 2002), albeit in a more complex form regarding its' European branch: Renfrew suggested that the Neolithic technological expansion from Anatolia to Europe may have represented the expansion of the Indo-European language family. Findings presented in this paper and in my aforementioned earlier paper suggests that the recognition by Renfrew of the role of the Balkans as the roadway for migrations into Europe having their edge in introducing more advanced technologies of ancient Anatolia and the Fertile Crescent to Europe is not only right, but was apparently the main driving force of European migrations from the Neolithic all the way to the Iron Age. However while Renfrew's original theory suspected a single such interaction event, one that introduced the Indo-European languages it seems that there was a complex series of such

consecutive migrations arriving on the same route throughout the millennia, each owing its' edge over previous Europeans to the introduction of another new technology from ancient Anatolia and the Middle east, as newer and newer technologies emerged there, and each introducing another language family. Otherwise the findings discussed in this paper as well as in my other, aforementioned paper confirms the farming/language dispersal hypothesis with the inclusion of the Basque, Indo-European, Northwest Caucasian, Northeast Caucasian, Tyrsenian and Uralic language families in the process. This issue is however is merely implied by findings of the two papers, thus needs further analysis, any may be topic for a future research paper on the issue itself.

This scenario of a series of consecutive migrations with each of them owing its' success to the introduction of a more advanced technology from Anatolia, starting with Neolithic agriculture followed by stages of copper, bronze and iron use, most likely with also higher population densities, also fits the mechanism described by Jared Diamond in his work "Guns Germs and Steel" (Diamond, 1999). Thus, it seems already the very beginnings of Europe were a consequence of conquerors owing their success to a combination of higher population densities and new technologies.

Regarding the primary topic of the paper, the question of pre-Indo-European languages in the R1b-P312 range, what basically means west of the Hamburg-Trieste line, with the identification of clues pointing to the presence of the Vasconic and Tyrsenian families, history back until about 3000 BCE is mapped. On the one hand, this means a great timespan compared to first written sources describing languages of the region from the middle of the first millennium BCE, but on the other hand, still leaves a surprisingly broad timespan uncharted, underlying how vast a timeframe millennia of prehistory are: As megalithic culture in Atlantic Europe started before the formation of the Bell Beaker culture, language families of the societies that started those, it still remain unidentified, and so is the language family of the people of Ötzi, and those of the first Neolithic populations of Europe. The hypothesis presented in this paper is of course is highly speculative, but due to the lack of evidence, so are all hypotheses regarding the issue of migrations and language families in prehistoric Europe. However, with the use of the most recent DNA evidence, this paper uses not less, but rather more hard evidence than previous hypotheses, that had to be constructed without these data, as they were not available even a mere decade before now.

## References:

1. Bellwood, P. S., & Renfrew, C. (2002). *Examining the farming/language dispersal hypothesis*. McDonald Institute for Archaeological Research.
2. Callaway, E. (2017). Ancient-genome study finds Bronze Age ‘Beaker culture’ invaded Britain. *Nature News*, 545(7654), 276.
3. Delamarre, X. (2003). *Dictionnaire de la langue gauloise* (2nd ed.). Paris: Editions Errance.
4. Diamond, J. M., & Ordunio, D. (1999). *Guns, germs, and steel*. Books on Tape.
5. Gamkrelidze, T. V., & Ivanov, V. V. 2010. *Indo-European and the Indo-Europeans: A Reconstruction and Historical Analysis of a Proto-Language and Proto-Culture. Part I: The Text. Part II: Bibliography, Indexes* (Vol. 80). Walter de Gruyter.
6. Gimbutas, Marija 1970. "Proto-Indo-European Culture: The Kurgan Culture during the Fifth, Fourth, and Third Millennia B.C.", in Cardona, George; Hoenigswald, Henry M.; Senn, Alfred (eds.), *Indo-European and Indo-Europeans: Papers Presented at the Third Indo-European Conference at the University of Pennsylvania, Philadelphia*: University of Pennsylvania Press, pp. 155–197.
7. Grugni, V., Raveane, A., Mattioli, F., Battaglia, V., Sala, C., Toniolo, D., ... & Torroni, A. (2018). Reconstructing the genetic history of Italians: new insights from a male (Y-chromosome) perspective. *Annals of human biology*, 45(1), 44-56.
8. Hay, M. (2017). *Genetic history of the Italian people*. Eupedia [https://www.eupedia.com/genetics/italian\\_dna.shtml](https://www.eupedia.com/genetics/italian_dna.shtml) Accessed 03.08.2019.
9. Hickey, R. (2002). *Language change in early Britain: The convergence account*. na.
10. Horvath C. B. (2019.) *Reconsidering The Geographic Origins Based On The Synthesis Of Archaeological And Linguistic Evidence And The Newest Results In Genetics – A Finno-Scythian Hypothesis*. Asia-Pacific Journal of Advanced Business and Social Studies. Year 2019. Volume: 4. Issue: 2 pp. 41-70.
11. Krell, K. S. (1994.) *Modern Indo-European homeland hypotheses: A critical examination of linguistic arguments*. University of Ottawa (Canada).
12. Koch, J. T. (2009). A case for Tartessian as a Celtic language. *Palaeohispanica*, 9, 339-351.
13. Malkiel, Y. (1952). old and new trends in spanish Linguistics. *Studies in Philology*, 49(3), 437-458.

14. Mellaart, James 1958. "The End of the Early Bronze Age in Anatolia and the Aegean". *American Journal of Archaeology*. **62** (1): 9–33.
15. Michalowski, P. (2017). Ancient Near Eastern and European isolates. In *Language isolates* (pp. 19-58). Routledge.
16. Orduna Aznar, E. (2005). Sobre algunos posibles numerales en textos ibéricos. *Palaeohispanica*, 5, 491-506.
17. Perkins, Phil (2017). "DNA and Etruscan identity". In Naso, Alessandro (ed.). *Etruscology*. Berlin: De Gruyter. pp. 109–118.
18. Renfrew, C. 1990. *Archaeology and language: the puzzle of Indo-European origins*. CUP Archive.
19. Rhys, J. (1898). *A Revised Account of the Inscriptions of the Northern Picts*. Dr.: Neill.
20. Rix, H. (1998), *Rätisch und Etruskisch [Raetian & Etruscan]* Innsbruck
21. Talbert, Richard J. A. *Atlas of Classical History*. Routledge, 1985, p.85
22. All maps in this paper are own work by the author of this paper, Csaba Barnabas Horvath, using as background: Terpischores 2012. „*Biomes of the World-The 14 biomes of the world according WWF*”. Categorized as „Labeled for reuse with modification” by Google [https://commons.wikimedia.org/wiki/File:Biomes\\_of\\_the\\_world.svg](https://commons.wikimedia.org/wiki/File:Biomes_of_the_world.svg) Accessed 03.08.2019.
23. Trask, R. L. (1995). Origins and relatives of the Basque language: Review of the evidence. *AMSTERDAM STUDIES IN THE THEORY AND HISTORY OF LINGUISTIC SCIENCE SERIES 4*, 65-100.
24. Turfa, J. M., & De Grummond, N. T. (1986). *Etruscan life and afterlife: a handbook of Etruscan studies*. Wayne State University Press.
25. Vennemann, T. (2003). *Europa Vasconica. Europa Semitica*. Trends in Linguistics. Series.
26. Watson, W. J., & Taylor, S. (1926). *The history of the Celtic place-names of Scotland* (p. 228). Birlinn.
27. Wiik, K. 1997. The Uralic and Finno-Ugric phonetic substratum in Proto-Germanic. *Linguistica Uralica*, 33(4), 258-280
28. YFull-Y-Chr Sequence Interpretation Service (2019). *YTree v7.06.01* <https://www.yfull.com/tree/> Accessed 03.08.2019.
29. Yunusbayev, B., Metspalu, M., Järve, M., Kutuev, I., Rootsi, S., Metspalu, E., ... & Yepiskoposyan, L. (2011). The Caucasus as an asymmetric semipermeable barrier to ancient human migrations. *Molecular biology and evolution*, 29(1), 359-365.