



**RECONSIDERING THE GEOGRAPHIC ORIGINS BASED ON THE
SYNTHESIS OF ARCHAEOLOGICAL AND LINGUISTIC EVIDENCE
AND THE NEWEST RESULTS IN GENETICS –
A FINNO-SCYTHIAN HYPOTHESIS**

Csaba Barnabas Horvath
Budapest, Hungary.

Corresponding Email: keluolang@hotmail.com

Abstract

Tracking down prehistoric migrations and locating the geographical birthplace (if any) of language families has been subject to great interest for centuries. New statistical data in genetics built up in the first two decades of the 21st century now contribute to a new set of clues in this puzzle, a set of clues that only leads us to plausible conclusions. However, if used with a great degree of caution, and as part of a multidisciplinary synthesis involving scientific data reached by archeology and linguistics as well. This paper attempts to map the origins of the Uralic language family with such methods, reaching a conclusion that is inherently new.

Keywords: Uralic, Finno-Ugric, Genetics, Archaeology, Linguistics.

**Introduction – Previous Research on the Issue by the Author,
Relevant literature, Methodology**

This paper presents a hypothesis regarding the origins of the Uralic language family, reached by a multidisciplinary synthesis, combining scientific evidence reached by archaeology, genetics and linguistics. The novelty of this approach is that besides results of archaeology and linguistics, it also relies on the statistical data of the geographic distribution of recently identified genomic elements. While the paper mostly uses secondary sources in the form of the three aforementioned disciplines, this kind of synthesization of results of the three disciplines involving the newest results of statistical data in genetics has so far happened only sporadically and fragmentally, as the three disciplines largely work independent of each-other. Thus, what enabled this paper to reach conclusions that are inherently novel, is a throughout synthesis of a broad range of archaeological, linguistic and most recently achieved genetic data covering the entirety of the Uralic language family, and possibly related archaeological cultures. The result, locating the Uralic urheimat (ancient homeland of origin) in the territory of what is today Ukraine, in the late Neolithic, the Calcholitic and early Bronze Age is in a stark contrast with the most widespread view on the issue, that before the genetic data involved in this research became available, tended to put the Uralic urheimat in the subarctic Ural region, such as theories by Péter Hajdu [Hajdú 1962], Juha Jahunen [Jahunen 2009] as well as numerous other hypotheses cited by Markku Niskanen [Niskanen 2002]. It was already proposed that the Corded Ware culture, or at least its' eastern flank representing Finno-Ugric languages in my earlier paper "The Story Of Two Northward Migrations origins of Finno-Permic and Balto-Slavic Languages in Northeast Europe, Based on Human Y-Chromosome Haplogroups" published in 2014 (Horváth 2014, and Corded Ware culture in turn representing a demographic continuum formed by a chain of migrations connecting Yamna, Corded Ware, Afanasievo and Andronovo cultures in another paper of mine, "R1a Subclades and Bronze Age Migrations on the Eurasian Steppes" published in 2015 (Horváth 2015). The Idea of Ukraine

as a candidate though, does not lack a precedent. Uralic urheimat in Ukraine was first suggested by Finnish linguist Kalevi Wiik (Wiik 2008) not the least by recognizing key patterns in the geographical distribution of a Y-chromosome DNA lineage, haplogroup R1a, a lineage also playing a key role in the present paper. At the time, Wiik constructed his hypothesis however, research of the named genetic lineage and genomic elements identifying it was still at an early stage, causing misconceptions in the timing and to a certain degree the geographic pattern of the demographic expansion spreading it to vast areas of Europe and Asia. Thus, Wiik got to a conclusion that already identified Ukraine as the urheimat in accordance with the conclusion of this paper that also puts it there, but the time of these events suggested by him suggested the Paleolithic, and also reaching to conclusions on the geographic pattern of the spread of the family, a pattern that followed the ribs of a fan-like geographic network of northwards expansion, with the center of the fan being Paleolithic Ukraine, and its' ribs reaching out Northwards from Scandinavia all the way to the Urals (Wiik 2008, p. 274.). This theory has not been confirmed since then, and thus got mostly rejected by the academic community, presumably not the least because contradictions caused by its' early timing. New genetic data however, as we will see, suggests a similar, but significantly different scenario, also compared to the chronology of divergence of Uralic languages, discussed in detail below, and instead of the Paleolithic, starting almost ten thousand years later in the Copper Age, and their expansion also following a somewhat different geographic pattern than the fan shaped network of migrations suggested by Wiik, although to a great degree, covering the same areas of Europe. These results retrospectively confirm that the approach of Wiik as much as linking the Uralic family to the geographic range of haplogroup R1a was basically right, as right as at the given stage of genetic research of the lineage it could be. On the other hand, the Uralic urheimat in Copper Age Ukraine gives such a place to the Uralic language family in the history Bronze Age Europe and Central Asia, that is inherently novel even compared to the Paleolithic hunter and gatherer societies in the center of the hypothesis of Wiik.

The hypothesis presented in this paper also implies origins of the Indo-European language family south of the Black Sea, in contradiction with the widely popular Kurgan hypothesis proposed by Maria Gimbutas, the currently perhaps most popular theory on the origins of the Indo-European language family, that puts Indo-European urheimat on the Pontic Steppe (Gimbutas 1970). Such a context, again, is not without precedent, as Colin Renfrew constructed a hypothesis on the origins of the Indo-European family, putting its' urheimat there (Renfrew 1990). Here again, in a similar way as regarding the theory of Wiik, conclusions of this paper support a geographic pattern of migrations in Europe similar to the one proposed by Renfrew, thus retrospectively confirming the validity of his approach, but instead of a single major wave of migration in the Neolithic as proposed by Renfrew, ends up suggesting a series of consecutive migrations following similar geographic patterns from the Copper Age all the way to the Iron Age, presumably each related to the spread of new agricultural and metallurgical technologies from the ancient Middle East. But the closest match to what this paper implies is an Indo-European urheimat in Eastern Anatolia in the 5th-4th millennium BC, proposed by Thomas V. Gamkrelidze and Vjaceslav V. Ivanov (Gamkrelidze & Ivanov 2010). Linguist Kathrin Krell also argued against the steppe origins of Indo-European, based on the reconstructed vocabulary of proto-Indo-European that implied a society with advanced agricultural and navigational skills (Krell 1994), alien to the steppes, but fitting Anatolia.

Regarding genetic data, the approach of this paper is mainly focused on Y-DNA. The main advantage of Y-DNA along with mtDNA, is that both of them are non-recombining genomic elements, thus each of the two represents a phylogenetic "family tree" of markers, where the place of each and every single sample can be clearly determined along that family tree, and not only spatial distribution of the statistical occurrence of Y-DNA and mtDNA lineages, so called haplogroups is known, but also the chronology of when such a lineages separated from its' sibling lineages as well. The reason why the main focus of this paper in genetics is Y-DNA rather than mtDNA, is that genetic statistical evidence confirms that in the preindustrial world,

men usually took part in migrating and conquering populations in larger part than women, while on the other hand, from the conquered populations, women survived and integrated in the society of the conquerors in a significantly larger part, than men did. An apparent example for this is the Mexican Mestizo population where the mtDNA cocktail is in majority of Native origins (Guardado-Estrada et al. 2009) while the Y-DNA cocktail is predominantly European [Martinez-Cortes et al. 2012], and in accordance with this, mtDNA patterns virtually everywhere in the world are more archaic, than Y-DNA. This meant that migrations tended to change the mtDNA cocktail of populations less than they tended to change the Y-DNA cocktail of them, and consecutive waves of such migrations going on almost everywhere in the world for millennia, amplified that anomaly to such a magnitude where the traces of migrations in Y-DNA distribution are much more apparent and thus much more practical to use than mtDNA. On the other hand, however, while in the direct use of genetic statistics this paper focuses on Y-DNA, it also uses secondary sources of analyses of different kind of DNA as well, in cases where the evidence is clear enough to determine if a certain migration in a certain direction at a certain time did take place. Large scale shifts in mtDNA occur only in cases where the conquering population exterminates or expels the conquered population rather than merging with it, or represents a population density so many times higher than the native one, that even that factor on its' own makes the representation of native lineages in the new cocktail extremely low. Such is the case of modern North America, where not only Native Y-DNA, but also Native mtDNA's presence in the gene pool is extremely low. Statistical data, however, suggests that prehistoric migrations of this latter kind were rare, and migrations of the former kind were the norm, with populations blending and merging, rather than replacing each-other.

Because of the very nature of non-recombining Y-DNA of course, the presence of Y-DNA haplogroups in almost every human population is highly diverse and mixed, as even the traces of populations of thousands of years and dozens of major migrations away are present. Another common occurrence that can seem to be confusing at fist is the phenomenon of so called "bottleneck effects", where (usually during major migrations) due to mostly random factors, certain genealogical lineages reproduce in disproportionately higher numbers, than the rest of the very same population, and if their Y-DNA (or for that matter mtDNA) lineage was up until that point otherwise atypical to the given population, this can cause a seemingly major change in the Y-DNA cocktail, despite of the population remaining the same, and merely expanding in numbers. One such example is N-Z1933 in Finland [Hay 2018 a]. Misleading evaluation of such data however can be easily avoided as the timing of separation of the successive lineages of such a shift can be easily determined from statistical data due to the phylogenetic and non-recombining nature of these genomic elements, and besides of defining the timing of their expansion, it can also be defined whether if they come from a lineage that has already been part of the given society at the time when the demographic expansion started, or if it newly arrived from elsewhere. Still, however, due to such factors, except for some extreme cases, certain Y-DNA lineages cannot simply be associated with certain ethnolinguistic identities.

How can Y-DNA distribution still help us to solve the puzzle then? Where, despite of all these problems, it can indeed help, is that it can clarify whether if two certain prehistoric cultures were linked to each-other or not, it can also help to decide which cultural shifts took place by migration, and which by cultural diffusion. Checking such connections of a certain number then clarify which geographical directions a certain culture had demographic links, and which geographical directions it did not. If we have a sufficient number of such links, these links start to show the form of chains and networks of migrations linking certain cultures to each other. Due to bottleneck effects, the Y-DNA haplogroup that confirms culture "B" formed by a demographic expansion from culture "A" can be different than the one that than another one that confirms culture "C" being formed by a demographic expansion from culture "B" but still, if it is confirmed that a demographic expansion from culture "A" formed culture "B", and a few centuries later culture "C" was formed by a demographic expansion from culture "B", with culture "B" at the same time confirmed not being effected by any other external demographic

impact in the meantime, we still get an apparent chain of migrations from A to B to C. Then, at a certain point such chains and networks reach a culture with a documented ethnolinguistic identity, that can help to define the identity of other members of the chain. Where linguistics come to the picture, is that linguistics studying language families can identify branches of each language family, and the timing of the separation of such branches, if the timing of such a separation coincides with the timing of a described chain of migrations splitting to two branches, that can also contribute to the picture. Regarding the Uralic language family, the classical theory of its' phylogenetic pattern suggests proto-Uralic splitting to proto-Finno-Ugric and proto-Samoyedic around 3000 BCE, then proto-Finno-Ugric splitting to proto-Finno-Permic and proto Ugric around 2000 BCE, with proto-Finno-Permic splitting to proto-Finno-Volgaic and proto-Permic around 1500 BCE, while proto-Ugric splitting to Magyaric and Ob-Ugric around 500 BCE (Csepregi 1998). Of course, we should also be cautious whether a certain migration necessarily means conquest or not. Here archaeology comes to our help, since if a sudden large-scale migration is accompanied by a major and sudden change of the character of a certain culture, especially that of the elites, this usually suggests conquest.

How to determine then, if, in case of a certain population conquering another population supplants its' own language on the conquered natives, or adopts their language? Here, in historically documented cases, we can see the former option, the conquered population shifting to the language of the conquerors clearly the more frequent, we can even say the usual scenario: The Romanization of Hispania and Gaul, the Anglicization of the British Isles, the Slavicization of the Balkans, the Arabization of North Africa, the Turkification of Anatolia, the Hispanicization of the Mesoamerican and Andean societies suggests the conquered shifting to the language of the conqueror even in cases where the former population actually outnumbers the latter. We can see examples for the opposite only in cases where the conquered population not only outnumbers the conquerors by several times, but also has a significant edge in agricultural technologies, urbanization and administrative structures. Perhaps the most notable examples for this scenario are the Romanization of conquering Germanic tribes settled in Italy, Gaul and Hispania, the Slavicization of the Varagian Vikings settled in the Rus, the Sinicization of virtually all nomadic groups that managed to conquer China, and last, but not least the almost total lack of Mongolisation in the Mongol Empire, outside the core ethnic Mongolian regions. Therefore, based on the historical record, this paper takes an approach of assuming language shift to the language of the conqueror in cases where the conquering culture either represents a population at least of a similar size as the conquered, or represents more advanced agricultural technologies, administration and urbanization. The conqueror adopting the language of the conquered is only considered in cases where both numbers and economic factors are on the side of the conquered.

1. Genetic Evidence Suggesting A Chain of Bronze Age Migrations in Eastern Europe and Central Asia

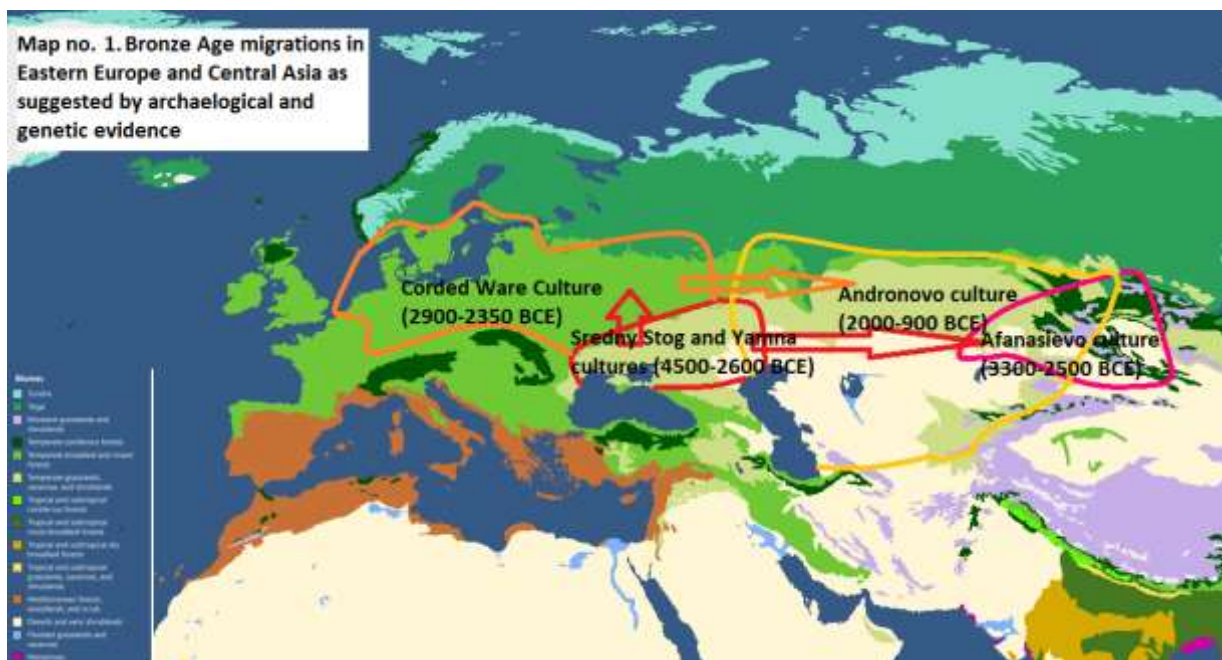
In an earlier paper of mine, "R1a Subclades And Bronze Age Migrations On The Eurasian Steppes" I reached a conclusion that the phylogenetic and geographical distribution of haplogroup R1a-Z645/S224 combined with the pattern of archeological cultures suggests a continuous structure of migrations covering most of Calcholithic and Bronze age Eastern Europe and western parts of Central Asia. As the issue can be found extensively discussed in that paper, here I share only a short summary of the findings: R1a-Z645/S224 is the most recent common ancestor of R1a-Z283 and R1a-Z93. Out of this two, R1a-Z283 is widespread in Eastern and Northern Europe from Northern Germany and Scandinavia all the way till the Urals and the Black Sea, while R1a-Z93 has been proved to have first spread in Central Asia. The first archeological culture where R1a-Z283 became dominant, was the Corded Ware culture spreading in a vast area, reaching from present day Northern Germany to the Ural-Idel region, while the first culture where R1a-Z93 was apparently dominant, was the Afanasievo culture of Central Asia. The migration that carried R1a-Z93 to the range of the Afanasievo culture apparently occurred along the steppe belt, from somewhere around present-day

Ukraine in Eastern Europe. Therefore, the location of divergence of these two branches most likely occurred there as well. In accordance with this, archeologists put the most likely origins of the Corded Ware culture to present day Ukraine as well, to the Yamna culture. The estimated time when both Yamna and Afanasievo cultures started, is around 3300BC, which also more or less fits the estimated time of separation of R1a-Z283 and R1a-Z93. Thus, however, the Yamna culture cannot be the ancestor of Afanasievo culture, but only a sibling to it, with R1a-Z93 already at Afanasievo by then, suggesting a common origin a bit earlier, and perhaps further west, as the common parent haplogroup of these two, R1a-Z645/S224 separated from its' sibling subclade, R1a-CTS4385 in a way that the latter one did spread in Central and Western Europe putting the most likely common origins of these latter two somewhere to the North Balkans, from where the Cucuteni Tripolye culture offers the suitable route towards the steppe belt. Later a culture also likely being an offshoot of Yamna, the Andronovo culture also spread to Central Asia (Horváth 2015).

Since then, new evidence confirmed the main concept of a migration deriving from Chalcolithic Ukraine forming the Corded Ware, the Afanasievo and the Andronovo cultures, but also meant that some details in it had to be revised. What confirmed the main concept, is that ancient DNA evidence proved the genetic link between the populations of Afanasievo and the Yamna cultures [Quiles 2018], as well as Corded Ware culture formed by a population arriving from the Yamna culture [Haak et al. 2015], and ancient DNA evidence confirmed the Andronovo culture in turn being formed by migration arriving from the Corded Ware culture [Allentoft et al. 2015]. Afanasievo culture originating from the Pontic Steppe, presumably from the Yamna culture was already proposed by David W. Anthony in 2010 [Anthony 2010, pp. 307-311.]. As Yamna (3300-the As Yamna (3300-2600 BCE) and Afanasievo (3300-2500 BCE) cultures started roughly the same time, Afanasievo is rather a sibling, than a descendant of Yamna, the genetic link between the populations of Yamna and Afanasievo rather suggests them originating from the predecessor of Yamna, the Sredny-Stog culture (4500-3500BC) with the population founding Yamna remaining on the spot, while the population founding Afanasievo migrating eastwards. Regardless, this new evidence from ancient DNA confirms the populations of Corded Ware, Andronovo and Afanasievo cultures being of the same demographic stock, spreading out as part of a massive demographic expansion from Chalcolithic Ukraine. Details that I had to revise are the following two: Afanasievo culture, although can be really linked to Chalcolithic Ukraine, in ancient DNA evidence, this link is mainly represented not by R1a-Z93, but by R1b-L23 [Quiles 2018], (that has been confirmed to be overwhelmingly its' eastern subclade R1b-2103 in the case of Yamna [Eupedia 2019 b] while samples from Afanasievo weren't tested for enough SNP-s to find out more) and despite Corded Ware (already R1a-M417 dominated as confirmed by ancient DNA evidence [Eupedia 2019 a]) representing a demographic continuity with Yamna, Yamna is also dominated by R1b rather than R1a. The most likely solution for this puzzle of confirmed demographic continuity, but differences in dominant Y-DNA haplogroups is a bottleneck affect, where the proportion of R1a was relatively low in the original Yamna and Cucuteni-Tripolye population, but disproportionately expanded during or right before the expansions that created Corded Ware and Andronovo cultures, and the same happened in the migration creating Afanasievo as well. Such bottlenecks are a common phenomenon, one example is the particular N-Z1933 subclade in Finland now [Hay 2018 a], that became widespread during the formation of the modern Finnish population, but was relatively minuscule among the ancestors of Finns when they were living further south from Finland before their migration arriving there, thus made Y-DNA haplogroup distribution significantly different among Finns than among Estonians despite of their otherwise extremely close genetic links. These bottlenecks are also supported by the fact that while R1a-Z283 and R1a-Z93 were the subclades that directly broke away from R1a-Z645/S224, in both cases it was a single first degree subclade of them, that became responsible for the bulk of the later expansion, namely R1a-Z282 in case of R1a-Z283, and R1a-Z94 in case of R1a-Z93 [Hay 2018 b]. It is also part of this problem whether if Z93 moved eastwards with the migration that created Afanasievo, or with Corded Ware, and the Andronovo. What still

supports the former option is that had Z93 moved eastwards first with Corded Ware, since it already separated from R1a-Z645/S224 by then, then its' earliest-derived subclades (presumably ones derived earlier than Z94) should be found in present or past populations of the Corded Ware range. They are however not present there, despite of the fact that ancient haplogroups survived much more often in the Corded Ware range than in the steppe belt where the Yamna=>Afanasievo migration must have taken place. This supports Z93, forming a bottleneck, moving eastwards as part of the population that formed Afanasievo, and was then still dominated by R1b-L23. One more detail where I had to do a slight revision is the ancient movement of R1a-Z645/S224, the common parent subclade of R1a-Z283 and Z-93, and its' own separation from its' sibling subclade, R1a-CTS4385. One possible scenario for the origins of this is R1a-Z645/S224 and R1a-CTS4385 separating from their parent subclade, R1a-M417 at the Kőrös-Stračevo culture (5800-4500 BCE) of the Northern Balkans, with R1a-CTS4385 moving westwards from there and R1a-Z645/S224 moving eastwards from there, settling at the Cucuteni-Tripolye culture (5500-2750 BCE) then R1a-Z283 and R1a-Z93 splitting from it there, with the carriers of R1a-Z283 moving as part of the population that formed the Yamna culture (3300-2600BC) and carriers of R1a-Z93 moving as part of the population that formed the Afanasievo culture (3300-2500 BCE). The other possible scenario is R1a-Z645/S224 and R1a-CTS4385 separating from R1a-M417 in the Cucuteni-Tripolye culture (5500-2750 BCE), then R1a-CTS4385 moving westwards, and R1a-Z645/S224 moving from where eastwards to the Sredny Stog culture (4400-3500 BCE) in present day Ukraine. Then, R1a-Z283 and R1a-Z93 splitting from R1a-Z645/S224 there, with the population containing R1a-Z93 first, forming the Afanasievo culture (3300-2500 BCE) and the population containing R1a-Z283 moving out from the region somewhat later, forming the Corded Ware culture (2900-2350 BCE).

Although, when I wrote my earlier paper described, I was leaning for the former one, since then new phylogenetic evidence emerged that lead us to support the latter scenario in at least two ways: Ancient DNA evidence Andronovo culture proved it being derived from the Corded Ware culture, and not from the Yamna culture [Allentoft et al. 2015] suggesting the expansion of R1a-Z283 starting with Corded Ware, and not with Yamna. This is also underlined by the fact that R1a-Z284 subclade of R1a-Z283, that is predominant in North Germany and Scandinavia, thus apparently represents a lineage of the western flank of Corded Ware culture was the one breaking away from it the earliest while other lineages deriving it in the east separated only later (Hay 2018 b). Regardless, in both cases we are facing traces of an ancient migration in which Corded Ware, Afanasievo and Andronovo cultures were all founded by a demographic expansion spreading out from Calcholithic Ukraine. These patterns apparently suggest an interconnected system of consecutive migrations. At a first glance, this looks anything but what one would associate with Uralic languages. But as we will see in this paper, taking a closer look shows a series of puzzles, to virtually all of which Uralic ethnolinguistic identity of this migration offers the most plausible solution. See map no. 1. [All maps in this paper are own work by the author of this paper Csaba Barnabas Horvath, using as background: Terpischores 2012. Categorized as "Labeled for reuse with modification" by Google]



Map 1.

2. Genetic Evidence, Corded Ware Culture and Finno-Ugric Languages

In another earlier paper of mine “The Story Of Two Northward Migrations origins Of Finno-Permic And Balto-Slavic Languages In Northeast Europe, Based On Human Y-Chromosome Haplogroups,” (Horvath 2014) I discussed the apparent anomalies that patterns in the spread of haplogroup R1a and its’ subclades present in Eastern Europe, patterns that make the widespread belief of Corded Ware culture being Indo-European, and its’ eastern flank being Balto-Slavic, to appear a highly unlikely scenario. As I extensively discussed the issue in the named research paper, I only give a short briefing of its’ findings here: I realized two main factors, that are confusing regarding this widespread belief: One is that Corded Ware culture in Eastern Europe spread further north, than the confirmed linguistic boundary between Balto-Slavic and Finno-Ugric languages in Eastern Europe as recently as around 1000 CE [Cartographia 1991 p. 107.] [Kuussaari 1935] [Pounds 1990, map 4.3] (not to mention estimates for earlier position of the linguistic boundary, reaching back in estimates till 500BC, and showing positions significantly further south to even that [Pounds 1990, map 2.3]). As migrations and the spread of virtually all agricultural and military technologies in the region in both historical and archeologically explored prehistoric times happened in a direction from south to north, this made it to seem highly unlikely that if Corded Ware culture in the region was really Indo-European, then Finno-Ugric-speaking cultures would have had any chance any time to regain those territories, so that the linguistic boundary could be south enough by 1000 CE at, where it really was then. This problem is underlined by three more facts: First, not only the linguistic boundary was further south, but also, as Finno-Permian populations south of the northern boundary of Corded Ware culture were largely agriculturalists, Finno-Ugric populations north of it were largely nomads, as well as hunters and gatherers, which means that population density in the south was several times higher than that in the north, making any north-to-south move even more unlikely. Second, as recently as 1000 CE, also the majority of individual Finno-Ugric languages were located south of the line, with the exception of Sámi and Permian languages, again suggesting territories south of the boundary rather being the center of the Finno-Ugric world, and territories to the north being the mere periphery of it. Then, I identified patterns in the distribution of the R1a subclade that further confirmed this suspicion: While the subclades of R1a-Z283 in general was indeed widespread among Finno-

Permian populations except for the northernmost ones, in accordance to the notion that they originally spread with the Corded Ware culture, and later continued to spread northwards by the south-to-north migration of Finno-Ugrian peoples. At the same time however, I identified another pattern, namely the pattern of distribution of the R1a-M458 subclade of R1a-Z283, that showed a striking coincidence with the spread of Slavic languages. While non-M458 R1a-Z283 subclades are present in both Slavic and Finno-Ugrian groups, the presence of R1a-M458 is high among Slavic groups, then dramatically drops at the linguistic boundary, and is extremely low in Finno-Ugrian populations. The geographic distribution of another haplogroup, I2a shows a pattern very same as that of R1a-M458, suggesting that they spread with the same migration of the same population. This suggests not one, but two northwards migrations in the region characterized by a high presence of R1a: One that of the original Corded Ware populations, apparently of the Finno-Ugrian origins, and another later one characterized by the high presence of R1a-M458 and I2a, apparently representing the Slavic expansion. As R1a-M458 and I2a is only present in a significant quantity only among Balto-Slavic and not among Finno-Ugrian populations, but at the same time there is a relatively high presence of non-M458 R1a also among Balto-Slavic populations besides Finno-Ugrian populations, this suggests large Finno-Ugrian populations being integrated and assimilated into the advancing Balto-Slavic societies (Horváth 2014).

Besides the scenario of the first wave of migration characterized by non-M458 R1a-Z283, and apparently spreading the Corded Ware culture being Finno-Ugrian, and the later migration wave characterized by the presence of R1a-M458 being Balto-Slavic, the only possible alternate explanation for this pattern would be the first wave of being of the Baltic branch of Balto-Slavic languages, and the latter being Slavic proper. However the relatively high presence of R1a-M458 and I2a among Lithuanians [Povilionis V. 2019] [Hay 2018 b] show a pattern resembling the one among Slavs, and not that of Finno-Ugrians, thus indicating the population and a migration characterized by the presence of R1a-M458 and I2a that was not just Slavic, but Balto-Slavic, and virtually excluding the possibility of a Baltic scenario, as in this means that in that case Lithuanians would be not Baltic, but Slavic speakers by now. (Besides this genetic evidence, this also solves the problem of the location of most Finno-Ugrian populations south of the northern boundary of Corded Ware culture as recently as 1000 CE, while the Baltic scenario wouldn't offer any solution to this problem.)

Similar to the case of the East European plain, phylogenetic patterns in Scandinavia suggests two major demographic shifts as well. There is an apparent difference between the phylogenetic patterns of R1a-Z283 and R1b-U106 show a considerable difference here. While both of them seem to be present in Scandinavia since the Corded Ware culture, their further derivations are different: While the subclade of R1a-Z282 dominant in Scandinavia, R1a-Z284 shows the most significant wave of internal derivation at the time of the Corded Ware culture [Hay 2018 b], on the other hand most subclades of R1b-U106 that are now the dominant representatives of their parent group in Scandinavia, derived in the late Nordic Bronze Age period, such as the R1b-Z18 subclade of R1b-U106 [Hay 2017]. This supports a secondary migration arriving to Scandinavia during the Nordic Bronze age, supporting the Uralic substrate hypothesis for Germanic languages of Kalevi Wiik [Wiik 1997]. See map no. 2. [All maps in this paper are own work by the author of this paper, Csaba Barnabas Horvath using as background: Terpischores 2012. Categorized as "Labeled for reuse with modification" by Google]

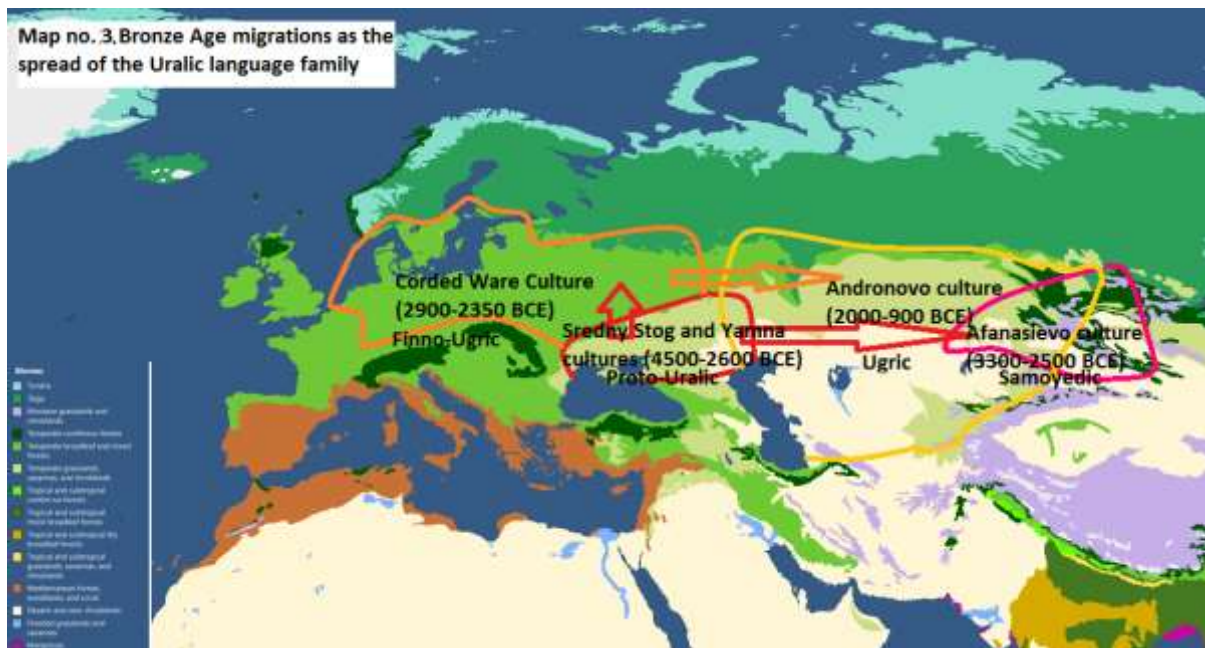


Map 2.

The next question is which cluster of the Uralic language family did the Corded Ware culture represent? The entire Uralic family, and its' branches derived only later? Only the Finno-Ugric languages? Or only the Finno-Permic branch, as Uralic languages spoken in former Corded Ware areas throughout the last millennium were almost exclusively Finno-Permian? If Corded Ware included the entire proto-Uralic realm, then separation of Finno-Ugric and Samoyedic languages should have happened after this period, and geographically deriving from this. However, as we will see in a later chapter of this paper, linguistic, genetic, and archeological patterns of Samoyedic do not fit the scenario of being originated from the Corded Ware realm, but rather from the Altai region, and also linguistic research shows them to have separated from Finno-Ugric much earlier, around 3000BC (Csepregi 1998). Therefore, Corded Ware could have only be either Finno-Ugric, or Finno-Permian, but not covering the entire Uralic family. Ugric languages are and were located eastwards from Finno-Permian, so here we have to look for demographic flow east from Corded Ware. For Corded Ware to be just Finno-Permian, the population that later developed Ugric languages should have left eastwards before Corded Ware culture was formed, presumably deriving from the Yamna culture, the direct parent culture of Corded Ware. For Corded Ware to be Finno-Ugric, they should have left later, presumably directly deriving from Corded Ware itself. Here both linguistic, archeological, and genetic evidence support the latter version: Ugric languages separated from Finno-Permian sometimes around 2000 BCE [(sepregi 1998). Within this timeframe, one major culture emerged directly east of Corded Ware, the Andronovo culture (2000-900 BCE), with it' start in 2000 BCE supporting the latter scenario. As we can see, ancient DNA also confirms Andronovo originating from Corded Ware [Allentoft et al. 2015]. Thus, all evidence suggests Corded Ware to be not merely Finno-Permian, but Finno-Ugric, with Ugric languages breaking away from it presumably along the eastwards demographic flow that formed the Andronovo culture, and Finno-Permian originating from the language of that part of the Corded Ware population that remained between the Baltic Sea and the confluence of the Volga and Kama rivers, and continued to expand its' territory northwards there. This conclusion has also significant implications on the supposed Uralic substrate of Germanic languages discussed earlier. It is important to note, that this means that the time when the western flank of Corded Ware culture, reaching Scandinavia, separated from its' eastern flank in the East European Plain, and the split of its' characteristic R1a-Z283 subclade, R1a-Z284 all occurred earlier than

the separation of the Andronovo culture from Corded Ware and the estimated time of the split between Ugric and Finno-Permian, this still in the Finno-Ugric period. Thus, such a Scandinavian and pre-Germanic western branch of Finno-Ugric languages must have been at least as different from both the Finno-Permian and Ugric branches of the group, as these two are from each-other. In addition to their early separation from each-other, it is also very likely that the population of the Funnelbeaker culture that this western flank of Finno-Ugrians encountered in Scandinavia was most likely of a completely different ethnolinguistic character, than the population of the Comb Ceramic culture of apparently Siberian origins, that the eastern flank, the ancestors of both the Finno-Permian and Ugric groups encountered in the east, therefore linguistic substratum from pre-Uralic languages, if any, must have also been different in the western flank, being absorbed from the languages of the Funnelbeaker culture and not from the Comb Ceramic culture, further adding to the difference. Given the fact how tremendous the linguistic distance between Finno-Permian and Ugric languages already is, likening it to the difference between Slavic and Romance languages by linguists [Karanko et al. 1985] and thinking about an extinct branch of Finno-Ugric with even greater linguistic distance from those two being the linguistic substrate of Germanic also explains why linguistic evidence regarding the Germanic substrate hypothesis has been inconclusive so far.

Thus, the conclusion that we get here is that the population of the Corded Ware culture was most likely of Finno-Ugric ethnolinguistic identity, and Balto-Slavic as well as Germanic languages spread in the region by a later migrations, marked by the spread of R1a-M458 regarding Balto-Slavic languages in Eastern Europe, and R1b-Z18 regarding Germanic languages in Scandinavia. What makes this conclusion even more interesting, is that the formation of Corded Ware culture in 2900 BCE and that of the Afanasievo culture in 3300 BCE by populations coming from the territory of present day Ukraine matches the date of separation of Finno-Ugric and Samoyedic branches of the Uralic, that the classical view puts to 3000BC [Csepregi 1998], and the formation of the Andronovo culture in 2000 BCE exactly matches the date where the classical theory puts the separation of Ugric from the rest of Finno-Ugric languages. Thus, here we can see a striking match of both linguistic, archeological and genetic evidence all consistently suggesting a scenario of the Uralic urheimat being Chalcolithic Ukraine in the Sredny Stog (4500-3500 BCE) period and earlier stage of the Yamna period (the entire Yamna period lasting from 3300 to 2500 BCE, so presumably the first half of this period) with the population that developed the Samoyedic languages moving eastwards through the steppe belt from it, forming the Afanasievo culture (3300-2600 BCE) around the Altai, and the population that developed the Finno-Ugric languages, moving north forming the Corded Ware culture (2900-2350 BCE) reaching from North Germany till the Ural-Idel region. Then, the population that developed the Ugric languages splitting eastwards and forming the Andronovo culture (2000-900 BCE) on the steppes between the Ural-Idel region and the Altai, and the population that developed Finno-Permian, remaining at the eastern territories of the Corded Ware culture. See no. 3. [All maps in this paper are own work by the author of this paper, Csaba Barnabas Horvath using as background: Terpischores 2012. Categorized as „Labeled for reuse with modification” by Google]



Map 3.

Before we take a look on the question if Andronovo and the Ugric languages, in the next chapter first we discuss the question of haplogroup N, usually associated with Uralic languages by the mainstream, and the question of Samoyedic languages, whether if they can really be linked to the Afanasievo culture.

Before taking a look at haplogroup N and the issue of Samoyedic languages however, as an epilogue of this chapter, we have to take note of a migration pattern also illustrated by the apparent two waves of northwards migration in Northeast Europe described: A trend of consecutive south-to north migration in Subarctic Eurasia throughout known history: Balto-Finnic and Permian populations pushing Sámi and Samoyedic populations to the north in Northeast Europe, while they themselves being pushed northwards by Balto-Slavic populations from the south. In eastern Siberia the two main indigenous ethnic groups of the region today, the Sakha and the Evenki both arrived to their present area by such migration, both linked to pastoral or agricultural societies of the steppe belt: The Evenki arriving from the broader Manchuria region, while the Sakha being a Turkic-speaking population arriving from the steppe belt, presumably from the upper Yenisei region, both of this groups replacing Paleo-Siberian (presumably Yukaghiric and Yeniseian) populations in the vast regions of East Siberia. Thus we can see the same trend from the Baltic to the Pacific as long as known history goes back. The explanation could be very simple: Natural environment can determine the direction of migration, and natural environment doesn't change that easily, thus can trigger consecutive migrations along the same routes in the same directions. Here, such natural factors are that areas further south are both more suitable for agriculture, and closer to the great demographic centers of Eurasia, thus tend to have population densities several times higher and more advanced agricultural as well as military technologies than subarctic territories adjacent to them. Therefore, in any military confrontation between pioneers coming from the south and existing populations of the subarctic of any given time, the former tended to have the edge in both numbers and military technologies. Besides this, higher population densities of the more southern areas could have contributed as a push factor, and lower population densities in the north offered free land, even if that of a much poorer agricultural quality. Such potential pioneer groups always emerged ensuring the newer and newer consecutive waves of migration, each pushing descendants of the former ones to the north, almost as a mechanism of a conveyor working in Subarctic Eurasia. Also, as the Subarctic is a more remote

environment, it can assure the survival of languages and cultures that already disappeared from their ancestral homeland further south, such as Finno-Permian languages that disappeared from present day Latvia and central parts of European Russia where they dominated as recently as 1000 CE, but are still in good shape in Finland and the Komi Republic, regions in a large part conquered by modern Finns and Komi since that time. In the east, we can see a similar pattern with the Sakha and the Evenki still abundant in the north, while ethnic Russian population dominating the upper Yenisei region, and Manchu language all but extinct in Manchuria.

3. The Question of Haplogroup N, The Seima-Turbino Migration, and Samoyedic Languages

Now, we have to analyze both the issue of haplogroup N, often associated by Uralic languages by the mainstream, and the issue of the Samoyedic languages, that branch of Uralic languages where haplogroup N is the most dominant in the populations speaking those languages. The two subclades of N1a1 that are most widespread in Europe, N1a1-Z1936 and N1a1-CSTS2929 are both relatively young, as both of them diverged around 2000BC. Their ancestral groups however show a pattern of them being present in Europe much longer before, suggesting that both N1a1-Z1936 and N1a1-CSTS2929 being formed already in Europe, and spread out by bottleneck affects. Both of them are descendants of N1a1- L708, a subclade that appeared about 6800 BCE. Although, within this branch, most European N1a1 belongs to the N1a1-L392/L1062 subclade that appeared only around 3000 BCE, but as N1a1-Y9022, a subclade that broke away from the parent lineage of L392/L1062 (that parent lineage being M2126/P298) around 5400BC [Hay M. 2018. a] [International Society of Genetic Genealogy 2019], is also based in Europe (in the Volga-Ural region), this means that the separation of Y9022 and M2126/P298 already happened in Europe. This means that N1a1-L708 must have arrived to Northeast Europe as early as around 5400 BCE. presence of N1a1 in the Comb Ceramic culture has been confirmed by findings in 2014. (Hay M. 2018. a; International Society of Genetic Genealogy 2019)

Phylogenetic distribution of N1a1-CSTS2929 shows most of its' divergence taking place in the present day Baltic states, Belarus, and adjacent regions of Russia (Phylogeographer 2019; Ilumäe et al. 2016) what combined with its' age suggests it being spread in the region as part of a population already mixed with R1a-Z280. As R1a Z280 apparently arrived from the southwest, and N1a1-CSTS2929 from the east, they could not have arrived to their blending spot together from the same direction, but arrived from opposite directions and met on the spot. This blending could have taken place in three possible ways: The carrier population of R1a-Z280 or its' ancestors settling as a superstrate on the population that carried N1a1-CSTS2929 or its' ancestors, the other way around the population carrying N1a1-CSTS2929 settling as a superstrate over a substrate population carrying R1a-Z280, or as a third possible option, without large scale migration, by a minuscule blending between two neighboring populations, followed by a bottleneck effect that later caused a large scale expansion of the haplogroup that took part in it. Archeological cultures suggesting major migrations in the region and timeframe of the blending of R1a-Z280 and N1a1-CSTS2929 are the Comb Ware culture, arriving from the east, followed by the Corded Ware culture arriving from the west, and then the Seima-Turbino culture again arriving from the east. R1a-Z280 apparently arrived by the Corded Ware culture, but then for the arrival of N1a1-CSTS2929, in case if it took place as part of a large scale migration we have two possible suspects at the first look: The Comb Ceramic culture, and the Seima-Turbino culture. The latter however, can be excluded do to the fact, that it hasn't reached the apparent geographical region where the blending between R1a-Z80 and N1a1-CSTS2929 took place: The high presence of N1a1-CSTS2929 among Baltic and Balto-Finnic populations as far south as Lithuania indicate it being spread to the southwest as far as the upper reaches of the Dnieper river. Throughout the known history of both Balto-Slavic and Finno-Permian peoples in the region in the last two millennia, a constant demographic flow was present from the south and southwest to the north and northeast, (from

more densely populated and technologically more advanced regions to more sparsely populated and technologically less advanced regions) [Horváth 2014] thus it is highly unlikely for any major demographic group to have moved south or southwest from its' past location during that period. Therefore, if N1a1-CSTS2929 originates from some large scale migration from the east, such a migration should have reached those regions spreading till the upper reaches of the river Dnieper earlier than that period. The Seima-Turbino migration does not fulfill these criteria, so if N1a1-CSTS2929 or its' ancestors arrived to the blending spot by a large scale migration, that could only have been the Comb Ceramic culture. Due to the age of N1a1-CSTS2929, this on the other hand, most likely means the arrival of its' ancestors than itself, with N1a1-CSTS2929 spreading later, as part of migrations of populations in which it was present already alongside R1a-Z280 after the blending already took place. This suggests a bottleneck effect involving a relatively small population containing the ancestors of N1a1-CSTS2929 arrived with the Comb Ceramic culture, assimilated by the Corded Ware population bringing R1a-Z280, and then spreading out already mixed with them. In this case, if the blending took place by Corded Ware populations assimilating a Comb Ceramic culture substratum, the language of the succeeding population must have been the language of the eastern Corded Ware culture, as this culture was the one taking over the other, and arriving from a more densely populated and technologically more advanced region. Besides the above option of the N1a1-CSTS2929 or its' ancestors arriving as part of a large-scale migration, the other possible option could have been a minuscule random blending between two neighboring populations and cultures as part of a bottleneck effect. However, even if the admixture happened this way, it doesn't changed the facts that the Corded Ware culture took over the Comb Ceramic culture, and that the Seima-Turbino culture didn't reach southwest enough to reach the region where the blending of R1a-Z280 and N1a1-CSTS2929 took place. Therefore, if the blending took place this way, with N1a1-CSTS2929 being introduced not as part of a major migration but as part of a random bottleneck interaction between two neighboring cultures, in this case too it became part of a culture that most likely spoke the language of the eastern Corded Ware population, and neither that of the Comb Ceramic culture, nor that of the Seima-Turbino culture.

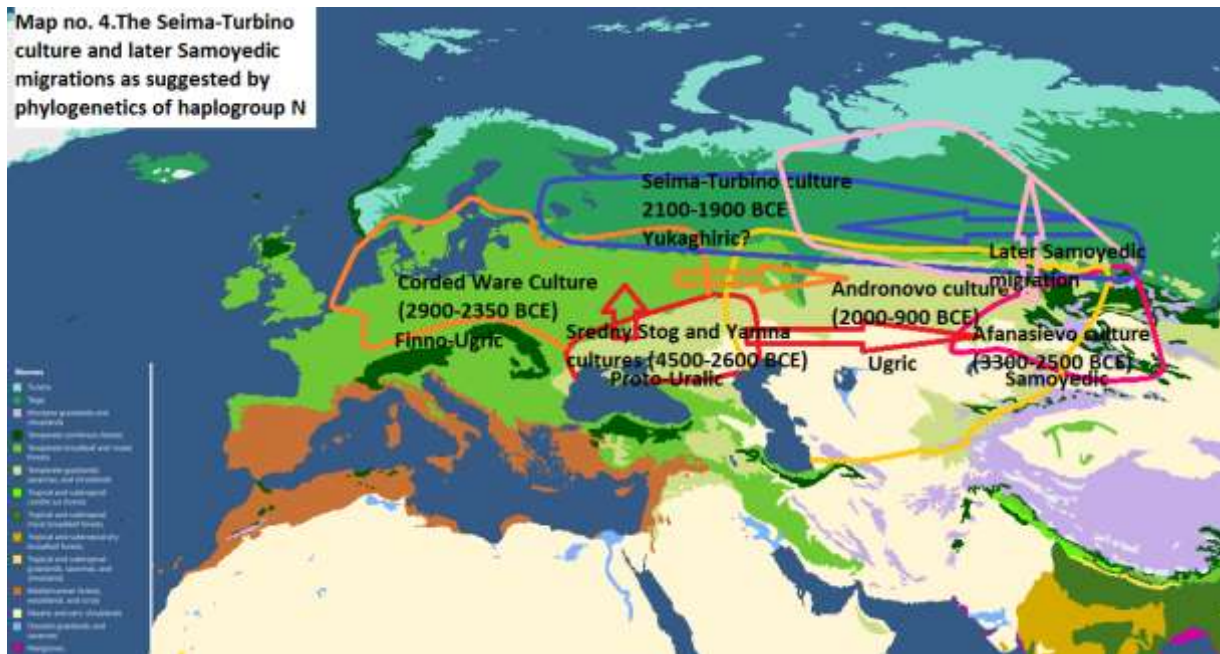
Getting to this conclusion, now the next question to be answered is whether if the Finno-Permian languages of Northeast Europe are indeed descending from the language of the eastern Corded Ware culture, as concluded in the previous chapter, or the Seima-Turbino culture a culture that arrived from Siberia around 2000 BCE, and is often associated by Uralic languages. (Again, as the Corded Ware culture arrived from the west in Europe, and the Seima-Turbino culture from the east in the region of the Altai mountains, their languages must have been greatly different if they represented two different demographics.) This question is answered by the fact that even as recently as 1000 CE, the majority of Finno-Permian populations and languages lived southwest of the boundary of the spread of the Seima-Turbino culture: The ancestors of Estonians, Livonians, the Merya, the Muroma, the Meshcher and Mordvins all living beyond this boundary (Cartographia 1991 p. 107.; Kuussaari 1935; Pounds 1990, map 4.3). Present day Finland and Republic of Karelia, territories within the Seima-Turbino range were still being mostly populated by the Sámi, thus even ancestors of the Finns themselves mostly lived south of the boundary around 1000 CE (Lamnidis et al. 2018). The mostly agriculturalist Finno-Permian speaking populations beyond the southwestern boundary of the earlier Seima-Turbino culture is being multiple times larger than the mainly nomadic and hunter-gatherer populations north of this line. Showing this massive south-to-north migration that later took place from south of the boundary of the Seima-Turbino range to within it, had Finno-Permian languages spread by the Seima-Turbino migration, then by early modern time only the Sámi and the Permians would have been speaking Finno-Permian languages, and each and every population that we know as Finno-Permian besides them would have been Balto-Slavic speaking by then. This virtually excludes the possibility of Finno-Permian languages of Northeast Europe descending from the Seima-Turbino culture, and thus putting their origins from the Corded ware culture.

The next question that we have to examine is the identity of the Seima-Turbino culture. The range and time of the expansion of the culture coincides with the age and spread of haplogroups N1a1a1a1a2 aka N1a1-Z1936, and N1a2b aka N1a2-P43 [International Society of Genetic Genealogy 2019], with its originating in the Altai region, thus showing sufficient correlations to accompany it with a large scale migration and ethnolinguistic identity on its' own, as well as demographically taking over any Post-Comb Ceramic culture populations. Regarding N1a1-Z1936, this suggests a bottleneck effect, with N1a1-Z1936 after being formed in Europe as part of a relative small human population, spreading out with the Seima-Turbino migration that arrived from the east, bringing N1a2-P43 with it. As described above, Finno-Permian identity can be excluded, as the range of the Seima-Turbino culture doesn't reach southwest enough to be responsible for the Finno-Permian language group. This is further confirmed by the phylogenetic patterns shown by N1a1-Z1936 and N1a2-P43, as they also largely remain north of the Neva-Upper Volga line (Ilumäe et al. 2016: Phylogeographer 2019). Besides all above, an additional fact adding to the case against the possibility of Uralic languages in general being introduced by the Seima-Turbino migration is the extremely high and uneven diversity among them. The linguistic distance between Hungarian and Finnish is as great as the one between Spanish and Russian (Karanko et al. 1985) making Finno-Ugric much more diverse than the Balto-Slavic languages adjacent to it to the south, and the linguistic distance between Samoyedic and the rest of Uralic languages is even greater than that. Had Balto-Slavic languages already arrived with the Corded Ware culture, quite the opposite would be the case. Also the Uralic languages are not only extremely diverse, but this diversity is uneven. According to the traditional model of the Uralic language family tree, no less than 1000 years took place between the separation of Samoyedic and Finno-Ugric languages, 500 years between the separation Samoyedic and Finno-Ugric, and the separation of the two main branches of Finno-Ugric, that is Ugric and Finno-Permian languages, another 500 years between this separation and the separation of the two main branches of Finno-Permian, Finno-Volgic and Permian, and the separation of the Finno-Volgic and Permian branches of Finno-Permian and a further 500 years before the next step of divergence within Finno-Volgic took place with the separation of the "Volga-Finns", then Mari and Mordvinic (Erzya and Moksha) from Finno-Lappic (Csepregi 1998). On the other hand, the entire Seima-Turbino migration took no more than a mere two centuries all the way from the Altai till Finland (Marchenko et al. 2017), in which the entire Seima-Turbino migration took place, and what would have allowed merely decades between each of those steps of divergence. Had the geographical expansion of the Uralic languages occurred along the geographical trail where the Seima-Turbino migration took place, the real pattern of divergence of Uralic languages would require a much slower and gradual migration, with a first phase with ancestors of Finno-Ugrians breaking away from the ancestors of the Samoyeds, moving westwards from an Uralic urheimat (what in this hypothetical case would presumably have been somewhere around the Altai) to a Finno-Ugric urheimat (in this case presumably somewhere around the Urals) then 1000 years later the ancestors of Finno-Permians breaking away from ancestors of the Ugrians and moving from there to a Finno-Permian urheimat (in this case presumably somewhere around the Volga-Kama confluence), then 500 years later the ancestors of the Finno-Volgans breaking away from the ancestors of the Permians and moving to a Finno-Volgan urheimat (what in this hypothetical case would presumably have been somewhere around the upper reaches of the river Volga) and then after an additional 500 years the population of the Finno-Lappic languages breaking away from the Volga-Finns, and reaching the region between lake Onega and the Gulf Of Finland. Thus in this case a time span of around 2000 years would have been required from the Altai to the Gulf of Finland along a geographical trail identical to that of the Seima-Turbino migration, as opposed to the mere 200 years long time span in which the Seima-Turbino migration took place in reality along the trail in question. Therefore, the Seima-Turbino migration does not even remotely fit the phylogenetic structure of the Uralic language family the way the chain of Bronze Age East European and Central Asian migrations in the focus of this paper does. The starting point of the Seima-Turniono migration in the Altai region

makes simply an Ugric identity highly unlikely as well. The Afanasievo culture was already present in the Altai region from 3300 BCE. This means that even if the Afanasievo culture was formed by a migration diverging from the same stock of which Corded Ware culture (the apparent carrier of the Finno-Ugric languages) originated, the separation couldn't have taken place later than 3300 BCE, a time far too early for the separation of Finno-Permian and Ugric languages that most linguists put at 2000 BCE (Csepregi 1998). Had Uralic languages in general spread out with the Seima-Turbino migration, their divergence would be much more even. The separation of Finno-Lappic and Volga-Finnic languages would have happened within 200 years to the separation of Finno-Ugric and Samoyedic, therefore the linguistic distance separating the former two would be hardly distinguishable from then one separating the latter two. At first glance, the option of Samoyedic identity would be supported by the fact that Samoyedic languages were also highly diverse in the Altai-Sayan-Upper Yenisei region before their extinction there, while are more homogenous to the north, which in turn also supports the possibility of the Upper Yenisei region being the original homeland of Samoyeds. Samoyedic identity however is made unlikely by the following facts: First, Samoyedic languages in the north are not diverse enough to have spread out at the time of the Seima-Turbino migration. Second, the presence of by a distinct subclade of N1a2-P43, its' second-degree subclade N1a2b1b1 aka N1a2-B170 is that within the Seima-Turbino range, virtually limited to Samoyedic and adjacent Ob-Ugric populations and is largely absent from other parts of the range of Seima-Turbino culture, while non-B170 N1a2b subclades can be found all over the range [International Society of Genetic Genealogy 2019]. This suggest a later migration performing the expansion of the Northern Samoyedic languages, migration by a population characterized by N1a2-B170, taking over and assimilating a substrate presumably of Seima-Turbino origins, and characterized by N1a1 and non-B170 N1a2b. One more possible option is a "Para-Samoyedic" ethnolinguistic identity with the Seima-Turbino languages being of the Samoyedic stock of Uralic languages, but of a different lineage than the proto-Samoyedic ancestor of all present day Samoyedic languages, as the latter diverged from each other only around 100 BCE (Csepregi 1998). However, a strong case against the Para-Samoyedic option is the linguistic connection between Uralic languages and the language isolate of Northeast Siberia, Yukaghiric (Aikio 2014). If Uralic languages had spread by either the Seima-Turbino migration or the Comb Ceramic culture from east to west, then their urheimat could have been located east enough to make an „Uralo-Yukaghir” language family, thus an early common ancestry and consecutive divergence with the Yukaghir possible. However, as we could see, both of those Comb Ceramic and the Seima-Turbino options can be excluded. In this case the linguistic connections between Uralic and Yukaghiric languages can only be originated from long lasting interactions between two neighboring language families. A Para-Samoyedic ethnolinguistic identity of the Seima-Turbino migration however would hinder such a pattern of interaction: First, from then on, Para-Samoyedic languages would have been stucked between the rest of the Uralic languages and Samoyedic, making any interaction after 2000 BCE virtually impossible. (With the only possible exception of the Northern Samoyeds who may have had a chance to reach the Arctic Ocean before Yeniseian and Tungusic languages pushing between them and the Yukaghir.) Second, if Seima-Turbino was not Yukaghiric, then the only migration taking the ancestors of the Yukaghirs to a proximity close enough to the Uralic languages could have been the Comb Ceramic culture. The migration that formed this culture however took place so early, already arriving to Europe around 4200BC, that with such an early date of divergence, the linguistic distance between the language of this population and Yukaghiric would be more distant than the distance between Finno-Ugric and Samoyedic, making the impact that this language had on Uralic hardly associable with Samoyedic. However, if the language of the Seima-Turbino migration was Yukaghiric, that makes all branches of Uralic living in close proximity of Yukaghiric till relatively recently and the languages of Seima-Turbino close enough to Yukaghir proper to be recognizable. Therefore the Para-Samoyedic option also seems to be rather unlikely, though not outright impossible. Thus the Seima-Turbino culture most likely represents a pre-Uralic ethnolinguistic identity. The

most likely candidate for this identity is the Yukaghiric language family given the combination of its' geographic location and its' linguistic links with the uralic language family that suggests a past strong interaction between the two language families at a certain point in history. Besides, N1a1 is also frequent among Yukaghirs (Duggan et al. 2013). One more fact supporting such a scenario is the suggestion by historical sources of an apparently pre-Samoyedic population called the "Sikhirtya" tribe, with mongoloid characteristics in appearance (therefore distinct of Finno-Permians in appearance) living along the Barents Sea as recently as the Early Medieval era, and then even in later centuries the Nenets discussed as consisting of two distinct subgroups, the "white Samoyed" of more Europid appearance and more southern distribution, and the "black Samoyed" with more Mongoloid appearance and more northern distribution (Birgisson 2014, in Veia 2017). All this suggests a pre-Samoyedic population with Mongoloid appearance, probably speaking a pre-Uralic language belonging to the Yukaghir language family, and characterized by a high presence of N1a1 and non-B170 N1a2b, later assimilated by the Samoyeds, who first appeared with the Seima-Turbino phenomenon, were characterized by a high presence of N1a2-B170, and more Europid traces in appearance.

Regarding the spread of the Samoyed languages, what apparently happened by a migration characterized by the presence of N1a2-B170 also explains the linguistic position of Samoyedic languages among the Uralic languages, and can contribute to the Expansion of N1a2-B170 starts from the Altai region, a region on the northern periphery of the Afanasievo culture. Here again we can see some very interesting matchings between archeological, linguistic and genetic evidence, that in turn links this N1a2-B170-characterized Samoyedic migration to the greater migration scheme described in the earlier part of this paper, linking it to the Afanasievo culture that we could see ancient DNA evidence to confirm to have derived from Yamna: The Afanasievo culture appeared around 3300BC, while linguists estimate the separation of Samoyedic languages from the rest of Uralic around 3000 BCE [Csepregi 1998]. The geographic location of Samoyedic languages known in historical times, and the secondary migration indicated by the geographic pattern of N1a2-B170 suggests a scenario, where Afanasievan languages, while being assimilated and replaced by Andronovan and other languages in the steppe regions of the Andronovo culture, survive at its' northern periphery, in a large part adopted by an N1a2 dominated population most likely of Yukaghiric origins, (one more example of the common pattern of taiga-nomads adopting the language of adjacent steppe-nomads, usually by smaller demographic fragments of the steppe-nomad group moving into their territory) and then spreading further northwards shown by the distribution of N1a2-B170. Such a scenario is also supported by the odd linguistic phylogenetics of Samoyedic languages that indicate all of the documented Samoyedic languages of historical times deriving from a single proto-Samoyedic language as recently as 100 BCE, while at the same time this linguistic lineages separating from the rest of Uralic as early as 3000 BCE [Csepregi 1998]. This odd combination of a linguistic lineage separating from the rest of the family that early, but all of its' present existing members separating from a single member language that recently in a geographic region otherwise characterized by a high fluctuation of nomadic migrations suggests something that we could call a "linguistic bottleneck": Most members of the given linguistic lineages becoming extinct, and all today's member languages descending from the sole surviving member. Therefore this correlation in archeological, linguistic and genetic evidence strongly suggests Samoyedic languages being the sole survivors of the Afanasievan languages, is otherwise extinct perhaps a millennia earlier. For a visual representation of this chapter, see map no. 4. [All maps in this paper are own work by the author of this paper, Csaba Barnabas Horvath using as background: Terpischores 2012. Categorized as „Labeled for reuse with modification” by Google]



Map 4.

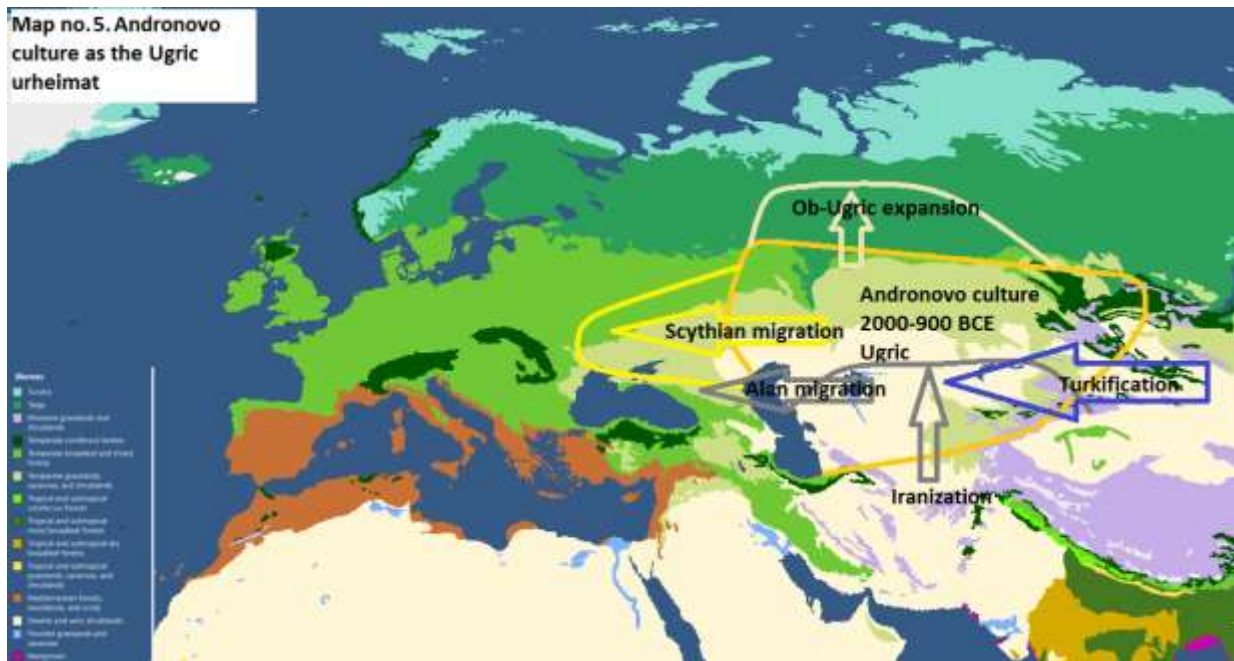
4. Andronovo Culture and the Ugric Languages

Now we arrive to the issue of Ugric, and Andronovo culture. As we could see, the population that formed Andronovo culture, arrived by a mass migration from the Corded Ware culture Allentoft et al. (2015). As Corded Ware culture seems to be Finno-Ugric as we have discussed, and Andronovo culture split away from it at 2000BC, right at the time when linguists estimate the split of Finno-Ugric languages to Finno-Permian and Ugric, this strongly suggests Andronovo to be the culture that developed Ugric languages. Moreover, Ob-Ugric languages, Khanty and Mansi are located on the northern periphery of the geographic range of the Andronovo culture. This can be easily explained by the phenomenon described earlier in this paper as the conveyor of Subarctic Eurasia, with demographic fragments and offshoots of pastoral and agricultural cultures of the temperate and steppe belt pushing into the subarctic belt, assimilating the population found there, as happened with the Finns and Karelians arriving to territories of present day Finland and the Republic of Karelia, and taking over from the Sámi, and the Sakha and the Evenki pushing into the subarctic and arctic belts of East Siberia, and taking it over from the Paleo-Siberian populations living there. Also, this phenomenon is often accompanied by the language of such populations surviving even after their ancestral culture in the south long became extinct. The Khanty and Mansi perfectly fit into this pattern as subarctic offshoots of the Andronovo culture, and this is also confirmed by the genetic evidence of both R1a-Z283, apparently descending from the Corded Ware culture, found among the Khanty and Mansi Quiles (2019), and N1a2-B170 otherwise characteristic to Northern Samoyedic populations, being also widespread among these two Ugric groups, suggesting earlier Samoyedic populations living in their present area, being assimilated by them as they arrived from the south.

Regarding Hungarian, also a member of Ugric languages, it has been confirmed by now by genetic evidence that much of non-M458 R1a-Z283, (that is in this case mostly belongs to the subclade R1a-Z280) being of common stock than the one found among the Khanty and Mansi [Quiles 2019], thus confirming the Andronovo ancestry of ancient Magyars, the ancestors of modern Hungarians. This also explains the pastoral lifestyle of ancient Magyars, same as that of the Andronovo culture. As the Scythian culture is a direct descendant of the Andronovo

culture, this raises the question of the ethnolinguistic identity of the Scythians, and the role of the Scythian culture in the migration of ancient Magyars, or in a broader sense, Hungarian or Magyar language from the Transural steppes to the Carpathian basin. Scythian culture is often associated with Iranian languages, however Scythian was an umbrella term in ancient Greek and Roman historiography, used for a number of groups of whom not necessarily all spoke the same language, and evidence for these cultures all being entirely Iranian-speaking is extremely scarce and limited, with no actual text in the Scythian language or languages spoken in the Pontic Steppe having survived. Where the use of Iranian languages is strongly confirmed within Scythian horizon is Khorasan and Transoxiana, basically the regions adjacent to, or even part of the Greater Iran area, south of a line connecting the Aral and Balkhash lakes, where the Betpak Dala desert and semi-deserts adjacent to it constitute as a barely habitable natural barrier between the greener pastures of today's northern Kazakhstan, Southern Siberia and the Pontic Steppe (this three forming an ecological continuum) north of it, and the fertile river valleys of Transoxiana accompanied by montane steppes to the south. Also, the earliest such evidence was recorded centuries later than the original Andronovo expansion. Linguistic evidence north of the Aral-Balkhash line is scarce, sporadic and inconclusive. Evidence in these regions is mostly limited to names of individuals, however examples from our age emphasize how weak such evidence is: Indo-European as well as Biblical names are and have been throughout the past millennium, widespread among today's Estonian, Finnish and Hungarian population, but that doesn't make these languages being neither Indo-European, nor Afro-Asiatic. It has been the same with names of Latin as well as Biblical origins in Germanic-speaking countries, as well as names of Persian origin in Turkic speaking societies, or names of Chinese origin in Korean and Vietnamese culture. Still, this obviously doesn't make Germanic languages Romance, Turkic languages Iranian, or Korean and Vietnamese Sinitic. Therefore it is extremely common for names to suggest a mere cultural diffusion or linguistic superstrate. The first well documented Iranian language north of the Aral-Balkhash line is Alanian, first appearing on the Pontic Steppe as recently as the first millennium CE. The Alans however, are described by contemporary sources as being recent migrants in the Pontic Steppe, having arrived from Central Asia, originating from, and being identical to the Massagete [Ammianus Marcellinus], an ethnic group living up until then near the lake Aral, south of the Aral-Balkhash line. Thus as matchings in genetic, archeological and linguistic evidence suggest Andronovo culture to be Ugric speaking, and Hungarian, an Ugric language originating from the steppes north of the Aral-Balkhash line being present on the Pontic Steppe up until the 9th century CE, with the first confirmed Iranian language on the Pontic Steppe being Alanian having arrived the only as recently as the 3rd century CE confirmed by contemporary sources having originated from the Massagete south of the Aral-Balkhash line, it can very well be the case that only that part of the Scythian world that lied south of the Aral-Balkhash line went through full Iranization, presumably introduced from the Iranian Plateau, (presumably signaled by the presence of J2 and other haplogroups of southern origin in the region) while in parts of the Scythian world north of the Aral-Balkhash line a widespread presence of Ugric languages survived up until the Tukrification of the West-Eurasian steppe belt (presumably started by the Hunnic conquest intruding there from the Mongolian Plateau), and the arrival of the Alans. The split of Ugric languages to Magyaric and Ob-Ugric put to around 500BC by linguists Csepregi (1998) also closely correlates with the expansion of the Scythians from the original Andronovo area to the Pontic Steppe, sometime in the 600s BCE. Thus the split between Ob-Ugric and Magyaric coincides with the split of the European and Asian Scythians, making it a plausible scenario of Magyaric descending from the Ugric language of the European Scythians, while Ob-Ugric descending from the Ugric language of the Asian Scythians. This raises the possibility of Hungarian Magyar population being formed by the blending of multiple such Ugric-speaking groups arriving to the Carpathian basin as part of different waves of nomadic migration arriving from the Pontic Steppe, such as the scenario described in the "double conquest" theory of Gyula Laszlo Laszlo (1978), or possibly starting even earlier, as Sarmatian groups have been present in the Carpathian basin since at least the arrival of the

Iazyges in 20 CE. Such series of multiple groups of Ugric speaking populations arriving to the Carpathian Basin, and the demographic buildup that such a series of consecutive migrations results, would explain the survival of Hungarian language there, and why it didn't become extinct the same way Turkic-Bulgar in Bulgaria did, or if a steppe language survived in the Carpathian Basin, then why wasn't it a Turkic or Iranian one, something that otherwise would be hard to explain. Bilingualism with Turkic used alongside their own Ugric language mentioned in contemporary sources Dümmerth (1977) being present among early Magyars can also be a sign of bilingualism among such tribes during the process of Turkification of the steppes. This record of bilingualism combined with high presence of R1a among Turkic-speaking populations of the steppes suggests a great portion of Scythian and Sarmatian populations surviving conquest by the Turkic-Speaking elites starting with the Hunnic empire of Attila, and this in turn makes a widespread bilingualism possible for a considerable period with Turkic-speaking elites and Ugric speaking commoners in the Pontic Steppe up until the 9th century among multiple populations that are generally viewed as Turkic due to the language of their elites, as Ugric-speaking Magyars were apparently present there up until Árpád's conquest at the end of the 9th century CE. This date of Árpád's conquest also coincides with the time when, by the arrival of the Pechengs, a Common Turkic speaking people [Spinnielt & Bădulescu 2003], Oghuric Turkic languages in the Pontic Steppe were replaced by Common Turkic, suggesting such builingualism existing in populations ruled by Oghuric elites, but no longer in Common Turkic populations. It is interesting to note, that in the 13th century, a Magyar speaking population was recorded in Volga Bulgaria, the last stronghold of Oghuric languages by Friar Julian, a Hungarian Dominican monk (Györffy 1986). Reasons for further Turkification being hindered among Magyars could be factors such as the peripheric location of the Carpathian Basin within the Eurasian steppe world, and the adoption of Roman Christianity, severing cultural links with the Pontic Steppe, as well as the replacement of Oghuric languages by Common Turkic at the Pontic Steppe. What makes this scenario likely, is that the alternative to it requires extreme coincidences that make alternative scenarios unlikely: A scenario where relatively large Scythian populations of the Pontic Steppe, of Andronovo origins, speaking an Ugric language, first adopting an Iranian language, and then centuries later (instead of continuing to speak their Iranian language or adopting a Turkic or nearby Finno-Permian, or Ob-Ugric language) re-adopting an Ugric language with exactly the same position on the family-tree of the Uralic languages where its' original Ugric language should have been as a result of the split between European Scythians and Asian Sycythians in the 600s BCE, would be nothing less than a miracle. Therefore, such populations simply continuing to speak its' original Ugric language all the way from the Andronovo period (possibly with some later adopted Iranian and Turkic linguistic superstrate) seems to be the most simple and most likely scenario, even by following the principal of Occam's razor. For a visual representation of the conclusions of this chapter see map no. 5.



Map 5.

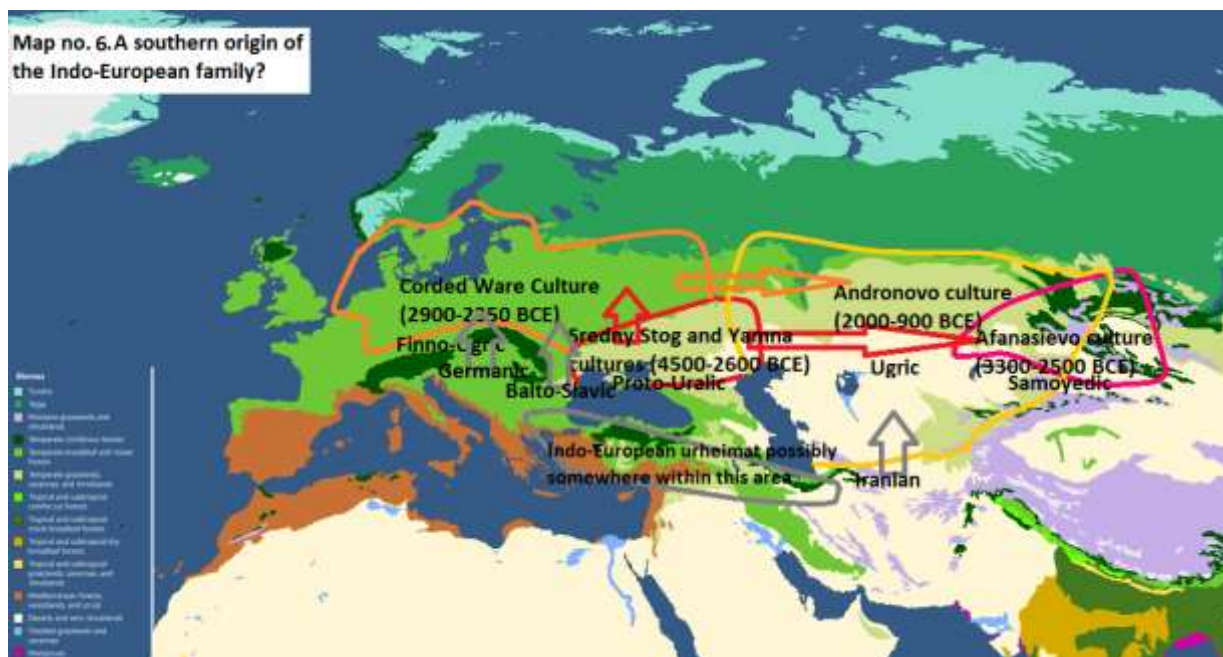
5. The Question of Loanwords

If the presented hypothesis is true, then regarding loanwords exchanged between Uralic languages and neighboring language families, when identifying loan words, one should also avoid prejudices based on the traditional view of Uralic language family exclusively representing subarctic societies, thus a revision of the categorization of loanwords dropping that prejudice may also be justifiable: Words in Uralic languages in common with Yukaghiric or other non-Uralic languages of the Siberian subarctic can easily happen to be actual pre-Uralic, presumably Yukaghiric loanwords instead of being ancient Uralic vocabulary, especially if they are absent from the more southern members of the family, such as Estonian, Livonian, Mordvinic languages and Hungarian, given the massive substrate population absorbed in the north, likely being of pre-Uralic origins as described in this paper. Words in Finno-Ugric languages that are in common with the Germanic, Balto-Slavic and Iranian branches of Indo-European, can easily be Finno-Ugric loan words in Germanic, Balto-Slavic or Iranian, instead of being the other way around, especially if they are absent from the rest of Indo-European languages, except for these branches, sharing massive substrate populations of Finno-Ugric origin, if the hypothesis described in this paper is true. Words in Ugric languages in common with Turkic can very well be Ugric loanwords in Turkic instead of being the other way around, especially if such words are absent from the rest of Altaic languages, excluding Turkic loanwords in Mongolian, given the likely possibility of Ugric languages being widespread in the steppes of the Aral-Balkhash line up until the Turkification of the steppe belt and thus a massive substrate population of Ugric origins in early Turkic societies described above.

6. The Question of Indo-European Origins

The hypothesis described in this paper obviously contradicts to the widespread Kurgan hypothesis of the origins of the Indo-European language family. Thus, though this issue would require a research on its' own, we have cannot leave the question of the Indo-European language family completely unaddressed, thus has to be briefly addressed different scholarly hypotheses on the issue, as well as other genetic data that may have implications on the issue. As we will see, although the hypothesis presented in this paper contradicts to the Kurgan

hypothesis, there actually are other hypotheses for the origins of the Indo-European family, which it does fit, and some more genetic evidence also seems to support such a scenario. At first, let us take a look at the interactions between the Indo-European and Uralic families earlier described in this paper, and their implications for the big picture. Balto-Slavic languages pushing into the Post-Corded-Ware area from the southwest carried by a demographic superstrate as suggested by the distribution of I2a and R1a-M458, Germanic languages arriving to Scandinavia from the south as well, but with Corded Ware being Finno-Ugric, originating from no further east than present day Romania, and Iranian languages spreading into Central Asia from the south, from the direction of the Iranian plateau, points to an expansion coming from a belt connecting the Balkans, Anatolia, and the Iranian Plateau. See map no. 6. [All maps in this paper are own work by the author, using as background: Terpischores 2012. Categorized as „Labeled for reuse with modification” by Google]



Map 6.

The idea of an Indo-European Urheimat within that belt is nothing new, as such scenarios have been proposed by a series of notable scholars, including Colin Renfrew who proposed Indo-European urheimat in Anatolia [Renfrew 1990] as well as Tamaz V. Gamkrelidze and Vyacheslav Ivanov who proposed it being located in Eastern Anatolia sometimes in the 5th-4th millennium BCE [Gamkrelidze & Ivanov 2010], exactly matching both the location and the timeframe where conclusions of this paper would imply Indo-European urheimat to be. Katherin Krell argued that reconstructed Indo-European vocabulary suggests a society with highly advanced agricultural and navigational skills, excluding the possibility of the Kurgan hypothesis [Krell 1994]. This in turn also fits Anatolia, a region with highly sophisticated agricultural societies, right next to the Fertile Crescent, the ancient cradle of agricultural civilization, as well as to the Mediterranean Sea, the cradle of highly advanced early maritime cultures. This would suggest an Anatolian urheimat, with Indo-European expansion spreading out from there sometimes in the early Bronze Age, their expansion apparently aided by their highly sophisticated skills in agriculture and metallurgy, resulting in a higher population density and better weaponry than their neighbors in Europe and Central Asia, most likely signaled by the spread of such haplogroups originating from the Balkans-Anatolia-Iranian Plateau belt, as J2 and E-V13, presumably soon carried on by populations characterized by assimilated bottlenecks of R1a and R1b, such as R1a-M458 in Slavic areas. What makes such a scenario as even more plausible than the Kurgan Hypothesis, is that while the rapid military

conquest that nomadic societies could reach often showed to be temporary only, highly advanced agricultural societies after all tended to get the upper hand over nomadic societies on the long run, among other factors due to their higher population densities. Thus, instead of an Indo-European expansion lead by a conquest by nomadic warriors, this scenario would mean an Indo-European expansion fuelled by more advanced technologies of agriculture and metallurgy, an expansion by pioneers using the latest such technologies developed in the Fertile Crescent, pushing into areas where the given technologies were not yet present at the given time, and also making Greek and Armenian probably the most ancient surviving Indo-European languages. As we could see, Germanic languages probably arrived to Scandinavia in the Nordic Bronze Age. Regarding the Slavic expansion, phylogenetic patterns of R1a-M458 suggests its' main demographic expansion taking place around 500 BCE [Lubicz-Lapinski L. & Milewski M. 2019] what in turn suggests the bulk of Balto-Slavic expansion occurring after 500 BCE. In this respect an interesting factor, linking Balto-Slavic languages to their likely ancestry and kinship in the East Balkans, is the close kinship between them and Thracian and Dacian languages proposed by Jonas Basanavičius [Spaudos 2019], later also supported by Daiva Vyčiniene [Vyčiene 2018] and Harvey E. Mayer [Mayer 1996]. This would perfectly fit in the described southern origins of Indo-European languages, with Balto-Slavic languages being born as an offshoot of Thracian in the east Balkans, spreading out to present day Ukraine, and then from there northwards on. Archeological culture fitting such a pattern are the Chernolets culture (1025-700 BCE) in the forest-steppe belt of present day Ukraine, viewed by many scholars to be the Iron Age Slavic homeland [Wiik 2008] and possibly also the Novocherkassk culture (900-650 BCE) in the steppe belt of today's Ukraine enabling the former to constitute a geographic continuum with Thracian and Dacian lands. During this period, the Thraco-Cimmerian findings showing definite cultural links between Thracia and the Pontic Steppe. The Novocherkassk culture is often associated with the Cimmerians, a pastoral people inhabiting the Pontic Steppe right before the migration of the Scythians in the region from the Andronovo culture. Language of the Cimmerians is unknown, but in this case, they could, together with the Chernolets culture very well be proto-Balto-Slavs, with Chernolets culture being the agriculturalist, and Novocherkassk culture being the pastoral branch of the same ethnic group. Thus any possible remaining of Uralic languages descending from the Yamna culture was likely assimilated and replaced by Balto-Slavic Cimmerians arriving from the Balkans in the 11th- 10th century BCE at the latest, they themselves taken over and replaced by Ugric Scythians in the 7th century, with possible surviving demographic fragments of the Cimmerian culture joining their kins in the Chernolets culture, what became the base of Balto-Slavic expansion to the north. Such a scenario would explain the paradox of linguistic connections between Balto-Slavic and Daco-Thracian languages, but the two being separated for long by nomads of Pontic Steppe representing different linguistic stocks. This linguistic kinship would also explain why any surviving Thracian languages, assimilated into South Slavic so easily during the Slavic migrations into the Balkans, in a stark contrast with the resilient survival of Albanian of probable Illyrian origins in the Western Balkans. In Central Asia, the Bactria-Margiana Archeological Complex or BMAC (2400-1600 BCE) is the ideal suspect of being the base of both the Iranization of southern parts of the Andronovo culture, Iranization of certain R1a-Z93 lineages, and the Indo-Iranian expansion in general. In accordance with an Anatolian origins of the Indo-European family, this culture has been linked to a migration arriving from the more western parts of the Iranian Plateau, by Victor Sarianidi [Salvatori 2016]. (Although, following the Kurgan hypothesis, he believed that to be non-Indo-European, and the migration coming from Central Asia being the Indo-European one, but according to the evidence presented in this paper, it rather seems to be the other way around.) Tokharian, an apparently more ancient branch of the Indo-European languages must have arrived must have arrived in the region earlier, and presumably even earlier than Andronovo. Here, the spread of the cultivation of wheat and barley between throughout the 3rd millennium BCE from the Iranian plateau to the Tarim-Basin via Transoxiana Venkatarmakrishnan (2019) may possibly be trace of such a migration.

Patterns of Y-DNA phylogenetics in Western Europe also support such an origin of Indo-European Languages, reaching Central and Western Europe from the Balkans. Here, the main anomaly is the predomination of haplogroup R1b-P-312 and its' subclades in populations that were reported to speak pre-Indo-European languages as recently as the time of Roman conquest of the given territories, showing a similar pattern as the high occurrence of R1a-Z283 suggests a Finno-Ugric ethnolinguistic identity of the Corded Ware range. Non-Celtic Iberians as well as the Etruscans and Raets were all speaking non-Indoeuropean languages at the time of Roman conquest in 3rd-2nd centuries BCE, not to mention the Basque, the only pre-Indoeuropean language surviving in Western Europe up until today. At the same time, all these groups are ancestors of present human populations with predominance of such ancient R1b-P312 subclades [Hay 2017] that suggests direct ancestry for the major part of each of these populations reaching back to at least those times, but more likely even earlier. (This list may or may not be also expanded by the inclusion of Pictish and Ligurian, languages of which Indo-European or non-Indo-European belonging hasn't been conclusively clarified yet, also within the R1b-P312 range.) Based on the general rule that language families being present longer tend to become more diverse and fragmented than those arrived more recently, linguistic traits within this range also support a relatively recent arrival of Indo-Europeans here: While documented pre-Indo-European languages were extremely diverse in the R1b-P312 range, in fact so diverse that Raet-Etruscan and Basque-Iberian are even confirmed to constitute two completely different language families, all Indo-European languages of this range all belong to the Italo-Celtic branch of that family. The solution for this pattern may possibly be a series of three consecutive migrations in the region: The first one forming Bell Beaker culture (2800-1800 BCE) arriving first from the east (also taking over the southwestern corner of the Corded Ware range), marked by the spread of R1b-312 in general and that of the Vasconic languages, being the ancestor of Basque and Iberian (as well as Pictish and Ligurian in case if they were also non-Indo-European languages) the second forming the Urnfield culture (1300-750 BCE) including Villanova, marked by the spread of R1b-U152 matching it in both age and range, with the Tyrsenian language family that included Etruscan and Raet, and the third one forming the La Tene culture (450 BCE-1 BCE) marked by the spread of R1b-L2 matching it in both geographic range and relatively recent age with such a relatively recent spread of the Celtic languages, accompanied by their Italic kins arriving to South Italy possibly directly from the Balkans across the Strait of Otranto, with a spread of some subclades rather of E-V13 and J2. Had the Bell Beaker migration already been Indo-European, Iberian and Basque languages couldn't have survived, and after close to three thousand years of divergence, the Romans would have found in their place Indo-European languages as distant from Latin, as Sanskrit is. Had the Urnfield migration already been Indo-European, the Raet and Etruscan languages couldn't have survived to see the Roman conquest, and Indo-European languages in their place would have been as different from Latin, as Slavic is. The earlier migration bringing Indo-European languages to the Balkans from Anatolia in the first place, could in turn be the Middle Bronze Age Migration of 1900 BCE suggested by James Mellaart [Mellaart 1958], not only matching the arrival of the Greek to Greece, but also happening late enough to allow Etruscan-Related Lemnian language to survive in the island of Lemnos, in case if Tyrsenian languages covered the Balkans before the arrival of the Indo-European languages from Anatolia, and before their own westwards expansion with the Urnfield culture. It is also interesting to note, that the area of present day Catalonia, was part of Urnfield Culture, but apparently not reached by the expansion of the population carrying R1b-U152, thus if Urnfield Culture was really linked to the spread of R1b-U152, then Catalonia is an area on its' periphery, where it reached only by cultural diffusion, and not migration. The fact that the Iberian language spoken there at the time of Roman conquest was Vasconic-related Iberian, and not Etruscan-related, would support such a scenario. See map no. 7. [All maps in this paper are own work by the author of this paper, Csaba Barnabas Horvath using as background: Terpischores 2012. Categorized as "Labeled for reuse with modification" by Google]



Map 7.

This series of possible consecutive migrations may even seem too turbulent a picture, but we must not forget that we are dealing with a timeframe of more than 2000 years, and is supported by the fact, that several consecutive waves of technological innovations did reach Europe from the Middle East along the very same geographical route, from Anatolia through the Balkans: Neolithic agriculture, the use of copper, the use of bronze at third one, and the use of iron just to mention the most apparent ones. Each of these events offered an opportunity to a culture located at the right strategic spot at the right time to expand its' territory northwest into Europe taking advantage of the technological edge that the given innovation provided to it over its' contemporary counterparts in the rest of Europe, thus it is theoretically possible of several of those waves to have represented a different culture, language, or even language family, though obviously it is not necessarily true to each and every of them. With the right spot mostly being the Balkans, as it constitutes a chokepoint on the way from the highly advanced Middle East to the less developed prehistoric European continent throughout these millennia. This also explains why cultures all dominated by subclades of R1b-L23 adjacent to the Balkans (Yamna, Bell Beaker, Urnfield, La Tene) are suggested by the evidence to belong to different language families: Such a gateway role of the Balkans can easily mean one or two of those language families being native to the area, and the rest of them being certain native population bottlenecks characterized by the presence of R1b-L23 being assimilated by consecutive waves of such migration, with each of them being assimilated by a different one. Also, the phylogenetic distribution of R1b-L23 makes such linguistic diversity caused by consecutive events of linguistic assimilation by distinct invading cultures from the south possible. The two main subclades of R1b-L23 are R1b-L51, present in Central and Western Europe and R1b-Z2103 present in Eastern Europe and Central Asia. The main sibling subclade of R1b-L23 is R1b-7562 that is widespread in the Balkans and Anatolia (Hay 2017). This rather suggests a scenario with R1b-L23 and R1b-7562 separating from each other somewhere in the Balkans or Anatolia, with R1b-7562 remaining there and R1b-L23 moving further north, R1b-L51 turning west along the Danube and R1b-Z2103 turning east to the Pontic Steppe, thus the whole process taking place along the spread of early agricultural and metallurgical technologies from Anatolia to Europe. Such a pattern would easily allow such linguistic diversity with one of those split lineages being part of a population that retained its' original language, and the other ones adopting the languages of conquerors that arrived somewhat later, presumably characterized

by the influx of more southern haplogroups, such as J1, J2, and E-V13. Anyhow, it seems that the most recent one of those that did complete the way, was the Indo-European family, as its' surviving hegemony over Europe means that it was the one that wasn't followed by any other. Proving or disproving such a Bronze Age Anatolian origin of the Indo-European language family can be the topic of another research and another research paper, and throughoutly proving or disproving it is beyond the frames of this paper. However, as the hypothesis discussed in this paper regarding the Uralic language family apparently contradicts the widely popular Kurgan Hypothesis regarding the Indo-European family, the issue how the Indo-European family could fit in this puzzle could not have been left unaddressed, thus had to be addressed briefly as done in this chapter, merely for the sake of making suggestions that can demonstrate that such scenarios for the Indo-European family are at least hypothetically possible, although need to be proved or disproved by future research. So, to summarize the way how this hypothesis relates to the question of the Indo-European urheimat, is that though it contradicts the Kurgan hypothesis of Marija Gimbutas, it perfectly fits the urheimat suggested by Gamkrelidze and Ivanov, the findings of Krell, and the logic of migrations mainly driven by economic innovations spreading out from the Fertile Crescent through Anatolia, proposed by Renfrew.

Ancient migrations typified along the romanticized picture of "barbarians" from rough countries taking over settled civilizations seems to focus on a phenomenon that was rather the exception than the norm. Agricultural civilizations tended to have a several times higher population density, and overall better technologies, than warlike nomadic tribes, with the latter usually only having a chance to conquer the former when that was in some deep internal crisis: Germanic tribes could only scratch the surface of the Roman Empire as long as it was in the height of its' power, even had to wage defensive wars against its' expansion as seen in the case of the Teutoburg Forest, and could only take over it when it fell in a deep internal socioeconomic crisis. The nomads of Mongolia and Manchuria only had chance to conquer China in cases when China was in some deep internal crisis, and otherwise they were the one subjugated and Sinicized, with the slow but steady flow of Han Chinese settlers in Manchuria putting a pressure of assimilation on the Jurchen by having been ongoing for centuries. The slow but steady flow of migrants and pioneers from more densely populated agricultural societies to sparsely populated nomadic or hunter-gatherer areas, (even the Great Bantu Migration of Sub-Saharan Africa, and the Austronesian expansion of Southeast Asia can be categorized as representing this trend) though slower and less spectacular than the swift conquests by "warlike barbarians," seems to have been on the long run much more effective than the latter, especially that the latter often represented groups relatively small in numbers.

Conclusion

The final conclusion of this paper is that the scenario synthesis of archaeological, linguistic and genetic evidence makes the most likely for the origins and expansion of Uralic languages is the Sredny Stog and Yamna cultures (5-4th millennia BCE) in the area of present day Ukraine being the urheimat in the proto-Uralic period, with the population developing Samoyedic languages breaking away from it and migrating eastwards in the 4th millennium BCE forming the Afanasievo Culture, from which a much later northward migration marked by the presence of N-B170 carried the modern Samoyedic languages to the subarctic zone, while Afanasievan languages became extinct in their original territory much earlier than that. Finno-Ugric urheimat on the other hand appears to be the Corded Ware culture (2900-2350 BCE) formed by a population migrating to the temperate forest belt of Northern and Eastern Europe. Then, the formation of the Andronovo culture in 2000 BC by a population splitting away from the Corded Ware culture apparently marked the split between Ugric and Finno-Permian languages, with Ugric being developed by the population establishing Andronovo culture, and Finno-Permian by the population remaining in the original Corded Ware area. Then, in the following millennia, Uralic languages wiped out of most of these areas by the expansion of Germanic, Balto-Slavic and Iranian languages from the south and Turkic languages from the east, with the present Uralic languages being the rare fragments surviving these trends. Due to

the suggested Uralic urheimat in the area of what is today Ukraine, and Andronovo culture also appearing to be Uralic, perhaps the phrase “Finno-Scythian” could mirror the essence of this hypothesis the best. These conclusions are supported by matchings between linguistic, archaeological and genetic evidence as show in the following table:

Table 1: Hypothesized events and fitting matchings in linguistic, archaeological and genetic evidence

Event	Linguistic evidence	Archaeological evidence	Genetic evidence
Samoyedic splitting away from proto-Uralic in the Pontic Steppe, with the population that developed the language that Samoyedic originated from, forming Afanasievo culture	Confirms the linguistic split in question to have occurred around 3000 BCE	Population migrates from Sredny Stog/Yamna to Central Asia forms Afanasievo 3700-3300 BCE	R1b links between Yamna and Afanasievo, separation of R1a-Z283 and R1a-Z93 around 4000BCE, modern Samoyedic population linked to Afanasievo area by the spread of N-B170
Finno-Ugric developed from proto-Uralic, with the population that formed proto-Finno-Ugric, moving north from the Yamna culture, forming the Corded Ware culture	Confirms the linguistic split in question to have occurred around 3000 BCE	Corded Ware deriving from Yamna around 2900 BC	Corded Ware population confirmed to have originated from Yamna, Yamna related R1a present in Finno-Permian populations of Northeast Europe today, distribution of R1a-M458 and I2 confirming Balto-Slavic population having arrived later than Corded Ware
Ugric splitting away from proto-Finno-Ugric, with population that formed Ugric, migrating eastwards from Corded Ware, forming the Andronovo culture	Confirms the linguistic split in question to have occurred around 2000 BCE	Andronovo deriving from Corded Ware around 2000 BCE	Andronovo population confirmed to have originated from eastern Corded Ware, much of R1a in Hungarians confirmed originating from eastern Corded Ware

We could also see patterns of consecutive migrations following the same routes that can be explained by the deterministic role of natural environment. It was often the natural environment determining out of two neighboring societies, which one had the edge that enabled it to conquer territory of the other. Thus, if in different historical periods, two societies emerged as neighbors of each-other in the same pattern of natural environment as a different pair of neighboring civilizations a few centuries earlier, it can easily result in a situation where out of the pair, the one gaining the upper hand will be the one living in the same territory as that of the one that gained the upper hand a few hundred years earlier. In certain geographical regions of characteristic natural environments, this seems to have created certain corridors of consecutive historical migrations, with demographic groups of different ethnic and linguistic identity following the same route in every couple of hundred years.

The theory presented of this paper is, of course, highly hypothetical, but so are all the theories on this issue, prehistoric origins of language families, including the widely popular Kurgan hypothesis regarding the Indo-European family, as well as the traditional model of the Uralic urheimat in the subarctic. Hypothetical is however something all theories on this issue

inherently have to be, as evidence from the given period is scarce and mostly indirect. The hypothesis presented in this paper is based on not less, but rather more hard evidence than earlier theories on the issue, as it not only recognizes the results of archaeology and linguistics, but also combines genetic evidence with them, such genetic evidence, that mostly became available only in the last decade. Thus, it wasn't yet available for use when earlier hypotheses on the issue were constructed, therefore they had to be constructed without relying on them. Future genetic results will be the main judge of the hypothesis presented here, and viewing the speed newer and newer subclades of subclades are identified, more and more ancient DNA evidence is acquired, broader and broader statistical databases are built up, we still have to wait years, but perhaps not that many years for this to happen. Such recent results, being published in the past few years, and already referenced in this paper, are for example ancient DNA evidence confirming the population of the Corded Ware culture originating in a massive migration from the Yamna culture, and the population of the Andronovo culture originating from the Corded Ware culture in a similar manner. It is rather a question of when, than if, for such evidence clarifying the nature of relations between other such cultures to emerge.

References

- i. Aikio, A., 2014. The Uralic-Yukaghir lexical correspondences: genetic inheritance, language contact or chance resemblance? *Finnisch-ugrische Forschungen* vol. 62, no. 2014, pp. 1-3.
- ii. Allentoft, M. E., Sikora, M., Sjögren, K. G., Rasmussen, S., Rasmussen, M., Stenderup, J., ... & Malaspinas, A. S. 2015. Population genomics of bronze age Eurasia. *Nature*, vol. 522, no. 7555, p. 167.
- iii. Ammianus Marcellinus Roman History Book XXXI. II. 12
- iv. Anthony, D.W. 2010. *The Horse, the Wheel, and Language: How Bronze-Age Riders from the Eurasian Steppes Shaped the Modern World* s.l.: s.n.
- v. Birgisson, B. 2014. *Den svarte vikingen*. Spartacus.
- vi. Cartographia 1991. *Történelmi világtasz*. Kartográfiai Vállalat Budapest. p. 107.
- vii. Csepregi, M. Ed. 1998. *Finnugor kalauz*. Panoráma.
- viii. Duggan, A. T., Whitten, M., Wiebe, V., Crawford, M., Butthof, A., Spitsyn, V., ... & Pakendorf, B. 2013. Investigating the prehistory of Tungusic peoples of Siberia and the Amur-Ussuri region with complete mtDNA genome sequences and Y-chromosomal markers. *PloS one*, vol. 8, no. 12, p. e83570.
- ix. Dümmerth, D. 1977. Az Árpádok nyomában.
- x. Eupedia 2019 a. *Corded Ware Culture*. Eupedia [Online] Available at: https://www.eupedia.com/genetics/corded_ware_culture.shtml [Accessed 16th July 2019]
- xi. Eupedia 2019 b. *Yamna Culture*. Eupedia [Online] Available at: https://www.eupedia.com/genetics/yamna_culture.shtml [Accessed 16th July 2019]
- xii. 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.
- xiii. 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.
- xiv. Guardado-Estrada, M., Juarez-Torres, E., Medina-Martinez, I., Wegier, A., Macías, A., Gomez, G., ... & Berumen, J. 2009. A great diversity of Amerindian mitochondrial DNA ancestry is present in the Mexican mestizo population. *Journal of human genetics*, vol. 54, no. 12, p. 695.
- xv. Györffy, G. 1986. Napkelet felfedezése. *Julianus barát és a Napkelet felfedezése. Válogatta, a bevezető tanulmányt és a jegyzeteket írta GYÖRFFY GYÖRGY*. Budapest, 27-40.
- xvi. Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., ... & Fu, Q. 2015. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature*, vol. 522, no. 7555, p. 207.
- xvii. Hajdú, P. 1962. *Finnugor népek és nyelvek*. Gondolat.
- xviii. Hay M. 2017. *Haplogroup R1b*. Eupedia [Online] Available at: https://eupedia.com/europe/Haplogroup_R1b_Y-DNA.shtml [Accessed 16th July 2019]
- xix. Hay M. 2018 a. *Haplogroup N1c*. Eupedia [Online] Available at: https://eupedia.com/europe/Haplogroup_N1c_Y-DNA.shtml [Accessed 16th September 2019]
- xx. Hay M. 2018 b. *Haplogroup R1a*. Eupedia [Online] Available at: https://eupedia.com/europe/Haplogroup_R1a_Y-DNA.shtml [Accessed 16th July 2019]
- xxi. Horváth, C. B. 2014. The story of two northward migrations origins of finno-permic and balto-slavic languages in northeast europe, based on human Y-Chromosome haplogroups337. *European Scientific Journal*.

- xxii. Horvath, C. B. 2015. R1a subclades and bronze age migrations on the eurasian steppes. In *3rd eurasian multidisciplinary forum, EMF 2015 19-21 October, Tbilisi, Georgia* p. 195.
- xxiii. Ilumäe, A. M., Reidla, M., Chukhryaeva, M., Järve, M., Post, H., Karmin, M., ... & Ekomasova, N. 2016. Human Y chromosome haplogroup N: a non-trivial time-resolved phylogeography that cuts across language families. *The American Journal of Human Genetics*, vol. 99, no. 1, pp. 163-173.
- xxiv. International Society of Genetic Genealogy, 2019. Haplogroup N Tree *International Society of Genetic Genealogy* [Online] Available at: <https://docs.google.com/spreadsheets/d/1ju7oNjHjMrgMUB1xXmr0EaaL-RctJJ8FreUPM2DzuBY/edit#gid=692817756> [Accessed 16th July 2019]
- xxv. Janhunen, J. 2009. Proto-Uralic—what, where, and when. *The quasiquicentennial of the Finno-Ugric society*, vol. 258, p. 57-78..
- xxvi. Karanko, O., László, K., & Kniivilä, I. 1985. *Finn nyelvkönyv: 1.: kezdőknek és középhaladóknak*. Tankönyvkiadó. p. 9.
- xxvii. Krell, K. S. 1994. *Modern Indo-European homeland hypotheses: A critical examination of linguistic arguments*. University of Ottawa (Canada).
- xxviii. Kuussaari E. 1935. Map off Finno-Ugric Distribution 900 AD – Map VII. *Historical Maps of Finno-Ugric People*. Complete Uralic World Guide [Online] Available at: <http://uralica.com/900ad.htm> [Accessed 16th July 2019]
- xxix. Lamnidis, T. C., Majander, K., Jeong, C., Salmela, E., Wessman, A., Moiseyev, V., ... & Sajantila, A. 2018. Ancient Fennoscandian genomes reveal origin and spread of Siberian ancestry in Europe. *Nature communications*, vol. 9, no. 1, p. 5018.
- xxx. Laszlo, G. 1978. *A kettos honfoglalás*.
- xxxi. Lubicz-Lapinski L. & Milewski M. 2019. *R1a Project. Family Tree DNA* [Online] Available at: <https://www.familytreedna.com/groups/r-1a/about/background> [Accessed 16th July 2019]
- xxxii. Marchenko, Z. V.; Svyatko, S. V.; Molodin, V. I.; Grishin, A. E.; Rykun, M. P. 2017. "Radiocarbon Chronology of Complexes With Seima-Turbino Type Objects (Bronze Age) in Southwestern Siberia", *Radiocarbon*, vol. 59, no. 5, pp. 1381–1397
- xxxiii. Martinez-Cortes, G., Salazar-Flores, J., Fernandez-Rodriguez, L. G., Rubi-Castellanos, R., Rodriguez-Loya, C., Velarde-Félix, J. S., ... & Rangel-Villalobos, H. 2012. Admixture and population structure in Mexican-Mestizos based on paternal lineages. *Journal of human genetics*, vol. 57, no.9, p. 568.
- xxxiv. Mayer, H.E. 1996. South Baltic. *Lituanus -Lithuanian Quarterly Journal of Arts and Sciences*. Volume 42, No. 2 -Summer 1996 ISSN 0024-5089
- xxxv. Mellaart, James 1958. "The End of the Early Bronze Age in Anatolia and the Aegean". *American Journal of Archaeology*. Vol. 62 no. 1, pp. 9–33.
- xxxvi. Niskanen, M. 2002. The origin of the Baltic-Finns from the physical anthropological point of view. *Mankind Quarterly*, vol. 43, no. 2, pp. 121-154.
- xxxvii. Phylogeographer 2019. Bronze Age Migrations of N Haplogroups to the Baltic via Finland. *Phylogeographer.com* [Online] Available at: <https://phylogeographer.com/bronze-age-migration-of-n-haplogroups-to-the-baltic-via-finland/> [Accessed 16th July 2019]
- xxxviii. Pounds, N. J. 1990. *An Historical Geography of Europe Abridged Version*. CUP Archive. map 4.3.
- xxxix. Povilionis V. 2019. *Lithuanian DNA. Family Tree DNA* [Online] Available at: <https://www.familytreedna.com/public/LithuanianDNA/default.aspx?section=yresults> [Accessed 16th July 2019]

- xl. Quiles C. 2019. *R1a-Z280 and R1a-Z93 shared by ancient Finno-Ugric populations; N1c-Tat expanded with Micro-Altaiic*. *IndoEuropean.eu* [Online] Available at: <https://indo-european.eu/2019/04/r1a-z280-and-r1a-z93-shared-by-ancient-ugric-populations-n1c-tat-expanded-with-micro-altaic/> [Accessed 16th July 2019]
- xli. Quiles, C. 2018. *Yamna/Afanasevo elite males dominated by R1b-L23, Okunevo brings ancient Siberian/Asian population*. *Indo-European.eu* [Online] Available at: <https://indo-european.eu/2018/06/yamna-afanasevo-elite-males-dominated-by-r1b-l23-okunevo-brings-ancient-siberian-asian-population/> [Accessed 16th July 2019]
- xlii. Renfrew, C. 1990. *Archaeology and language: the puzzle of Indo-European origins*. CUP Archive..
- xl.iii. Salvatori, S. 2016. *Bactria-Margiana Archaeological Complex: how terminology hides historical processes*. *Trudy Margianskoi Arkeologicheskoi Ekspeditsii-Pamyatii Viktora Ivanovicha Sarianidi [Transactions of the Margiana Archaeological Expedition-Dedicated to the Memory of Victor Ivanovich Sarianidi]*. *Staryi Sad, Moscow*, pp. 449-460
- xliv. Spaudos 2019. *Jonas Basanavicius*. *Spaudos.lt* [Online] Available at: http://www.spaudos.lt/Istorija/J_Basanavicius_b.en.htm [Accessed 16th July 2019]
- xl. v. Spinei, V., & Bădulescu, D. 2003. *The Great Migrations in the East and South East of Europe from the Ninth to the Thirteenth Century*. Romanian Cultural Institute.
- xlvi. Terpischores, 2012. „*Biomes of the World-The 14 biomes of the world according WWF*”. *Categorized as „Labeled for reuse with modification” by Google* [Online] Available at: https://commons.wikimedia.org/wiki/File:Biomes_of_the_world.svg [Accessed 16th July 2019]
- xl. vii. Veia M. S. 2017. *Hjør. Avaldsnes, Norway’s Oldest Royal Seat. Nordvegen Historisenter* [Online] Available at: <https://avaldsnes.info/en/informasjon/hjor/> [Accessed 16th July 2019]
- xl. viii. Venkatarmakrishnan, R. 2019. *Aryan migration: Everything you need to know about the new study on Indian genetics*. *Scroll.in* [Online] Available at: <https://scroll.in/article/874102/aryan-migration-everything-you-need-to-know-about-the-new-study-on-indian-genetics> [Accessed 16th July 2019]
- xl. ix. Vyčiene, D. 2018: *Relationships Between Lithuanian And Balkan Schwebungs-Diaphoneie: Interdisciplinary Search Key*. *Academia.edu* [Online] Available at: https://www.academia.edu/9177602/RELATIONSHIPS_BETWEEN_LITHUANIAN_AND_BALKAN_SCHWEBUNGS-DIAPHONIE_INTERDISCIPLINARY_SEARCH_KEY [Accessed 16th July 2019]
- l. Wiik, K. 1997. *The Uralic and Finno-Ugric phonetic substratum in Proto-Germanic*. *Linguistica Uralica*, vol. 33, no. 4, pp. 258-280.
- li. Wiik, K., & Kádár, G. 2008. *Az európai népek eredete*. Nap Kiadó.