

THE STORY OF TWO NORTHWARD MIGRATIONS-ORIGINS OF FINNO-PERMIC AND BALTO-SLAVIC LANGUAGES IN NORTHEAST EUROPE, BASED ON HUMAN Y-CHROMOSOME HAPLOGROUPS³³⁷

Csaba Barnabás Horváth, PhD
Corvinus University of Budapest, Hungary

Abstract

This paper attempts to find a plausible explanation for the origins of Balto-Slavic and Finno-Permian languages in Northeast Europe by a research based on the analysis of statistical databases of human Y-chromosome haplogroups. The mainstream view that associates Balto-Slavic languages with haplogroup R1a and the Corded Ware Culture, and Finno-Permian languages with haplogroup N, fails to solve several contradictions: How come, that the presence of subclades of haplogroup R are as high in most Finno-Permian populations, as the presence of haplogroup N? How come, that Corded Ware culture spread so far north, that it covered the early-medieval range of most Finno-Permic languages? This paper is trying to set up a hypothesis that solves these contradictions.

Keywords: Human Y-chromosome haplogroups; Corded Ware, language shift, migration, assimilation

Introduction

This paper attempts to identify the origins of Balto-Slavic and Finno-Permian languages in Northeast Europe, by the analysis of statistical samples of human Y-chromosome haplogroups. Due to limitations in recommended paper length, this paper can not give a detailed description of the biological nature of human Y-chromosome haplogroups, and also, the topic of this paper is not the biological research of human Y-chromosome haplogroups, but the analysis of what their statistical presence in different human populations could indicate regarding the historical research of prehistoric migrations and the origins of language families. It does worth to note however, that human Y-chromosome haplogroups are distinct by the non-recombining parts of Y-chromosome DNA. Different markers draw a patrilinear phylogenetic tree. Thus the statistical presence of different human Y-chromosome haplogroups in different populations can give a help to historians and archeologists in finding out more about historical and prehistoric migrations.

Main haplogroups of Northeast Europe and mainstream views

In Northeast Europe, we can distinct two main Y chromosome haplogroups, haplogroup R1a and haplogroup N. Haplogroup R1a is dominant among Balto-Slavic peoples but as we will see, also significant among Finno-Ugric populations, while haplogroup N can mainly found among Uralic populations, and it is especially dominant among the Samoyeds. The dominant subclade of haplogroup R1a in Northeast Europe is R-Z283, that was born

³³⁷With special thanks to my friend, Tibor Fehér from FTDNA, who helped me with the latest Y-DNA statistics

about 5500 years ago,³³⁸ and its' spread is often associated with the ancient expansion of the Bronze Age Corded Ware Culture. An other subclade of R1a is also present in the region: R-M458, that was born more than a thousand years later than R-Z283.³³⁹

Haplogroup N is also represented by two different subclades in the region. An older one, N1c, that seems to have been autochthon to the region before the arrival of haplogroup R1a, and a newer one, haplogroup N1b that seems to be a more recent arrival from Western Siberia.³⁴⁰

Two other haplogroups of presumably Central European origins, R1b and I2a are also present in the region.³⁴¹

Since the use of human Y-chromosome haplogroups in the analysis of prehistory and the origins of language families is pretty new, dating back only a few years, research in this aspect is still in early stages. Remarkable works have already been born, but due to limitations in recommended paper length, this paper can not give a literature review on the issue. At this point, the mainstream view regarding the research topic of this paper sees haplogroup R1a as the carrier of Indo-European languages (especially Balto-Slavic in the examined region) and haplogroup N as the carrier of Uralic languages. This explanation however contains several contradictions that the present paper is trying to solve. This paper attempts to identify these contradictions, and offer an alternate explanation for the prehistoric origins of Finno-Permian and Balto-Slavic languages in Northeast Europe based on human Y-chromosome haplogroups.

Questions raised by the dynamics of language shift

A key factor regarding the issue of Uralic and Finno-Permian languages, is that it is quite often the case, that present geographic range of language families does not coincide with their original or past range. It is highly misleading to approach present or recent geographic ranges as something static, since geographic ranges of language families are highly dynamic. A good example of this is the well-known case of the Celtic languages. It is common knowledge today, that the original core geographic range of the Celtic languages approximately covered the area of the La Tené and Hallstatt cultures, an area that roughly coincides with the present German states of Bavaria, Hessen, Baden-Württemberg and Rhineland-Pfalz, and the French regions of Alsace, Lorraine, and Burgundy, as well as parts of Austria. At the same time, present Celtic-speaking areas are located along the western coast of Ireland, along the western coast of Scotland, northern and western parts of Wales, and western Bretagne. What makes the situation bizarre is that not an inch of today's Celtic-speaking areas coincides with the original Celtic core area, and no native Celtic language is spoken any longer there. At the same time, those geographical areas where Celtic languages are still spoken, lie outside the original core area, and are the remnants of remote peripheral outposts of Celtic expansion in the Iron Age. Therefore, if unaware this fact, in order to identify the origins of the Celts, one would collect genetic samples from the present Celtic-speaking areas, ignoring the original core area, the results could easily be misleading.

The case is somewhat similar regarding the Finno-Permic language-group. Today Finno-Ugric languages are mainly associated by arctic and subarctic climates, with taiga and tundra regions, that were often characterized by hunting, gathering and semi-nomad reindeer

³³⁸Rozhanskii, Igor L., and Anatole A. Klyosov. "Haplogroup R1a, its subclades and branches in Europe during the last 9,000 years." *Advances in Anthropology* 2 (2012): 139.

³³⁹Rozhanskii, Igor L., and Anatole A. Klyosov. "Haplogroup R1a, its subclades and branches in Europe during the last 9,000 years." *Advances in Anthropology* 2 (2012): 139.

³⁴⁰Derenko, Miroslava, et al. "Y-chromosome haplogroup N dispersals from south Siberia to Europe." *Journal of Human Genetics* 52.9 (2007): 763-770.

³⁴¹http://www.eupedia.com/europe/european_y-dna_haplogroups.shtml

husbandry up until the modern age. The case however was much different as recently, as a millennium ago. For centuries, the northwards advance of the Slavic-speaking population was steady in the region, slowly but steadily assimilating Finno-Permian speaking groups. This means, that the further we look back into history, the further south we find the Slavo-Finnic language boundary. About 1000-1300 years ago, the area of the present day Central Federal District of Russia was mainly inhabited by Finno-Permian groups, such as the Merya, the Muroma, and the Meshcher.³⁴² This region lies southwards of the subarctic taiga region, and regarding climate, it can be defined more as continental than subarctic, mixed forest and not taiga,³⁴³ while its culture dating back at least to the bronze age is characterized more by some kind of farming agriculture, than by hunting or reindeer husbandry. Therefore, we are talking about a region with a population density much higher than that of the taiga and tundra regions further north. Besides the extinct Merya, Muroma and Meshcher, surviving remnants of the original Finno-Permian languages of this region are the Mordvin and Mari languages. Further west, in a similar latitude, -rather temperate than subarctic, with a traditional lifestyle rather agricultural than hunting- Estonian is such a surviving Finno-Permian language. About 1000-1300 years ago Estonian was accompanied by at least two other Finno-Permian languages right next to it: Livonian in the south, in present day Latvia, and Votic further east, in present day Russia. At the same time, about 1000-1700 years ago, most of what is Finland today was still inhabited by Sámi hunter-gatherers, and the case was the same in case of the area of the present day Republic of Karelia. Also, much of today's Komi Republic was populated by Samoyedic, and perhaps Ob-Ugric groups, and significant areas of present day Archangelsk Oblast were also populated by Sámi and Samoyed nomads.³⁴⁴

The Samoyeds are not Finno-Ugrians, but an other language family among the Uralic languages, that have developed on it's own for at least the last 6000 years. The Sámis do speak Finno-Ugric languages now, but it's widely accepted, that it's the consequence of a language-shift, and originally they had spoken some unknown pre-Finno-Ugric languages.³⁴⁵³⁴⁶

So Finno-Permian presence was not only much more significant in latitudes further south than it is today, but it was also either absent, or represented by a much lower population density in the subarctic taiga and tundra regions that we usually associate with Finno-Ugric languages today. So due to the historically agricultural lifestyle and thus higher population density, we can assume that as of 1000-1300 years ago, the Mari, the Merya, the Muroma, the Meshcher, the Mordvin, the Estonian, the Livonian, and Votic made up a solid demographic majority of Finno-Permian speakers, especially if we also take the above described contemporary situation of present day Finland, Republic of Karelia, and Komi Republic into account. This means that as recently as 1000-1300 years ago, the demographic core of the

Finno-Permian language group was located in today's Central Federal Region of Russia and the Baltic States, a belt south of the

If we go further back in time, what we find only confirms this assumption. About 1500-2000 years ago, the area of present day Finland and Republic of Karelia was almost entirely inhabited by Sámis, the area of the present Komi Republic was almost entirely inhabited by Samoyeds, while the area of present day Archangel Oblast was divided between

³⁴²<http://peacecountry0.tripod.com/900ad.htm#oz>

³⁴³Pounds, Norman JG. *An Historical Geography of Europe Abridged Version*. CUP Archive, 1990. map 1.2.

³⁴⁴<http://peacecountry0.tripod.com/900ad.htm#oz>

³⁴⁵Aikio, Ante. "An essay on substrate studies and the origin of Saami." *Etymologie, Entlehnungen und Entwicklungen: Festschrift für Jorma Koivulehto zum 70* (2004): 5-34.

³⁴⁶Aikio, Ante. "New and old Samoyed etymologies 2." *Finnisch-ugrische Forschungen Band 59-Heft 1-3* (2006).

Sámi and Samoyed populations.³⁴⁷ According to most estimates, the Finno-Ugric/Balton-Slavic linguistic boundary was somewhere around the Riga-Novgorod-Moscow-Tambov line about 1000 years ago, and somewhere around the Vilnius-Kiyev line about 2500 years ago.³⁴⁸

So this all means that if we we want to trace the origins of the Finno-Permian, the gene pool of the Estonians, Livonians, Votes, the Merya, the Muroma, the Mescher, the Mordvin, the Mari, and extinct Finno-Permian groups further south may tell more than Finns of Finland, or the northern Komi.

Besides shifts in the geographic range of certain groups, an other important issue is the dynamics of language shifts within a single population. It is quite common in case of migrations that conquering groups, outnumbered by the conquered indigenous population, but possessing advantages in weaponry, agricultural and industrial technologies, manage to achieve a language shift among the more numerous subjugated population. This happened in the case of the Slavic conquest of the Balkans when the bulk of the more numerous Romance and Greek-speaking indigenous population shifted to the Slavic language of the conquerors, the Anglo-Saxon conquest of England, where the Celtic and perhaps also Romance speaking natives adopted the language of the numerically weaker conquerors, in case of the Arab conquest, when the Aramaic-speaking population of Iraq and Syria, the Coptic-speaking population of Egypt, as well as the Berber-speaking population of the Maghreb largely changed their language to Arabic, despite being more numerous than the conquerors. The same thing happened in the case of Mexico, Peru and Bolivia, where the Native and Métis population (still a majority together) more and more switched to Spanish from their original Nahuatl, Quechuan, and Aymara languages. The medieval Slavic migration to Finno-Permian regions of Northeast Europe was marked by a similar process. Much of the original Finno-Permian population were assimilated by the Slavic newcomers.³⁴⁹

The subjugated native populations seemed to be able to assimilate the conquerors only in cases where besides solid numerical superiority, they also possessed more advanced economic structures, and the conquerors lacked advance in anything but military organization. This was the case in the assimilation of Germanic groups in the successor states of the West Roman Empire, and the assimilation of several nomadic groups by the Han Chinese.

This indicates, that we should find significant presence of genetic traces of the assimilated Finno-Permian groups among the Russian-speaking population of the region, but at the same time, high presence of Slavic genes among the Finno-Permian population is unlikely, since language shift to Slavic was the dominant scenario not only among mixed populations, but even simply among Finno-Permian populations affected by Slavic culture.³⁵⁰ Again, we can see analogic phenomenons from different parts of the world. Native origins are common among the Spanish-speaking populations of Latin-America, but one can hardly find Métis or people of mixed or European origins who use Quechua, Aymara or Nahuatl as their native language.

So if we accept the view of the mainstream, that considers the haplogroup N as the representatives of ancient Finno-Ugric population, and the R1a haplogroup as the emblem of Slavic newcomers, then we woul expect an extremely high frequency of haplogroup N among the Russians, and an extremely low frequency of haplogroup R 1a among the remaining Finno-Ugric speakers

³⁴⁷ <http://peacecountry0.tripod.com/400ad.htm>

³⁴⁸ Pounds, Norman JG. *An Historical Geography of Europe Abridged Version*. CUP Archive, 1990. map 2.3

³⁴⁹ Homén, Theodor: *East Carelia and Kola Lapmark*. Longmans, Green and Company, 1921. pp. 106-119.

³⁵⁰ Homén, Theodor: *East Carelia and Kola Lapmark*. Longmans, Green and Company, 1921. pp. 106-119.

The genetic composition of European Russia, however, is quite the opposite. The frequency of haplogroup N among Russians is only 23%, lower than the proportion of haplogroup R1a in most Finno-Permian populations. At the same time, in most Finno-Permian ethnic groups the frequency of R1a is about as high as the frequency of haplogroup N. Among the Estonians, the combined presence of haplogroup R1a and R1b is 40%, while haplogroup N is just 34%. Among the Mari, R1a and R1b together constitutes 39.5%, while haplogroup N 49.5%, among the Komi R1a and R1b makes up 40% while N 51%, among the Mordvin R1a and R1b constitutes 40% while N just 19.5%, and among the Udmurt R1a and R1b gives 24.5% while N gives 67%. A notable exception is Finland, where R1a and R1b constitute only 8.5%, while N is at 61.5%.³⁵¹ As we can see, in most Finno-Permian groups, the combined presence is higher or about the same as the combined weight of N1c and N1b. There are only two exceptions, where the frequency of N is higher than the frequency of R: The Finns of Finland and the Udmurts. The Finns however, -as described above- have only arrived to present day Finland between 0 and 500 AD, from Estonia and Karelia, wiping out the Sámi aborigines, and this migration caused a significant bottleneck effect that could be misleading regarding the genetic origins of Finlanders.³⁵² Also the particular subclade of haplogroup N1c, that is numerous among the Finns, the N1c1a1a1b (N-L1022) is a pretty recent one³⁵³, its' origins coinciding with the time when Finns arrived to present day Finland from Estonia. So the Finns are one of the newest, one of the most recently formed ethnic group among the Finno-Ugrians. So, their genetic composition, dominated by N1c1a1a1b, can by no means give any basis of speculations about the genetic composition of ancient Finno-Ugrians. So in aware of the extremely recent ethnogenesis of Finlanders, the only real exception with N-majority among European Finno-Ugrians are the Udmurts.

These genetic statistics are especially interesting in the light of the dynamic nature of geographic range of language families described above. As described, the bulk of the Finno-Permian population was located in the mixed-forest belt south of the 60th parallel as recently as 1600 years ago, and thus the gene pool of the Estonian, the Mordvin, the Mari, and the Udmurt population can tell us far more about the character of the prehistoric Finno-Permian population, than samples from ethnic groups further north, that were formed afterwards as offshoots of Northward migrations and expansion from these areas.

In the light of the dynamics of worldwide cases of language shift described above, it is highly unlikely that Finno-Permian groups would have managed to absorb without a language shift, masses of Balto-Slavic conquerors being not only technologically more advanced, but also as numerous as the natives themselves were. Therefore, the assumption that the first wave of R1a migration carried Balto-Slavic languages in the region seems to be highly unlikely.

Archeological findings show a similar pattern. The northern limit of the Corded Ware Culture –usually associated with haplogroup R1a and Indo-European languages – reached far more north, than the linguistic boundary between Slavic and Finno-Permian languages as recently as 1000-1300 years ago, and covered most of the above described key region of Livonian, Estonian, Votic, Merya, Muroma, Meshcher, and Mordvin traditional linguistic areas.

To summarize: Neither the range of R1a, nor the range of Corded Ware Culture coincides with the gap between the Balto-Slavic and Finno-Permian. Both geographical ranges seem to rather overlap with the gap between the ancient southern Finno-Permian groups (Estonian, Vot, Livonian, Merya, Muroma, Meshcher, Mari, Mordvin), and those

³⁵¹http://www.eupedia.com/europe/european_y-dna_haplogroups.shtml

³⁵²Kittles, Rick A., et al. "Autosomal, mitochondrial, and Y chromosome DNA variation in Finland: Evidence for a male-specific bottleneck." *American journal of physical anthropology* 108.4 (1999): 381-399.

³⁵³http://www.eupedia.com/europe/Haplogroup_N1c_Y-DNA.shtml#subclades

lands that were populated by Sámi and Samoyedic societies as recently as 1500-2000 years ago.

Subclade R-M458 of R1a, and haplogroup I2a as an answer to the question

There does exist a difference in gene pools however, that coincides with the gap between the Slavic-speaking population, and the surviving southern Finno-Permian nations. This is the apparently overlapping range of a subclade of Haplogroup R1a, R-M458, and a subclade of haplogroup I, I2a. Haplogroup R M458 is a younger subclade of haplogroup R1a, its' origins are estimated at 4200 before present, with a margin of 450 years.³⁵⁴ As much as there is no gap along the Slavo-Finnic linguistic boundary regarding the range of R1a, there does exist a clear gap regarding the overlapping ranges of R-M458 and I2a, as visible on the following table:³⁵⁵

	Czech	Slovak	Polish	Belarus	Russian average	Ukrainian average	Mari	Estonian	Udmurt
R-M458	30,20%	21,40%	26,60%	12,70%	12,47%	13,45%	0,00%	5,10%	0,00%
I2a	9,00%	16,00%	5,50%	17,50%	10,50%	13,00%	1,00%	3,00%	0,50%
R-M458 and I2a together	39,20%	37,40%	32,10%	30,20%	22,97%	26,45%	1,00%	8,10%	0,50%
Non M458 R1a	7,50%	24,90%	19,30%	42,00%	40,88%	33,73%	31,40%	28,00%	16,70%

These genetic traces indicate not one, but two northwards migrations: A first one, dominated by older subgroups of R1a, with impacts reaching most Finno-Permian populations of Northeast Europe, and a second one, dominated by R-M458 and I2a1, with its' impact largely limited to Slavic-speaking populations, but not reaching Finno-Permian speaking groups.

So, a plausible explanation could be that the Corded Ware Culture (or at least its' eastern flank) carried a first wave of R1a migration with Finno-Permian languages, while a second migration marked by R-M458 and I2a, and introducing Balto-Slavic languages occurred sometime between 3000-4000 BP.

This conclusion is quite close to that of Kalevi Wiik in his book on the origins of the peoples of Europe, except for the fact, that while Wiik placed the Northwards Finno-Ugric migration to the Neolithic, this paper places it to the Bronze Age, with the Slavic migration following perhaps at the dawn of the Iron Age.³⁵⁶

The Seima-Turbino phenomenon, and why it could not be the origin of Finno-Permian languages

A common concept is to view the alleged Seima-Turbino migration, the expansion of a subarctic Bronze Age culture that occurred about 3500 years BP, and spread from South

³⁵⁴Rozhanskii, Igor L., and Anatole A. Klyosov. "Haplogroup R1a, its subclades and branches in Europe during the last 9,000 years." *Advances in Anthropology* 2 (2012): 139.

³⁵⁵ Almost no statistics with the distinction of all these different subclades are available, so the table show a mathematical combination of statistical data from two sources (unfortunately no data was available for both of R-M458 and I2a for the Latvian, the Lithuanian, and the Mordvin population):

- Underhill, Peter A., et al. "The phylogenetic and geographic structure of Y-chromosome haplogroup R1a." *European Journal of Human Genetics* (2014).
- http://www.eupedia.com/europe/european_y-dna_haplogroups.shtml

³⁵⁶Wiik, Kalevi, and György Kádár. *Az európai népek eredete*. Nap Kiadó, 2008.

Siberia through the Urals reaching the taiga belt of northeast Europe,³⁵⁷ as a possible carrier of Finno-Ugric languages.

There are several factors however, that could exclude the possibility of Seima-Turbino migration being the carrier of Finno-Ugric languages:

- First, the range of this migration in the west did not extend into the west and south far enough to reach the medieval linguistic boundary. The Seima-Turbino only extended to the sparsely populated subarctic regions of Northeast Europe, but not into the more densely populated areas of the mixed forest belt of continental climate,³⁵⁸ so it did not reach the key area of southern Finno-Permic groups described above. This indicates that in the densely populated continental areas, the population unaffected by the Seima-Turbino migration vastly outnumbers the subarctic population affected by it.

- If the Finno-Ugric population had arrived with the Seima-Turbino migration, Finno-Ugric languages would be younger than Balto-Slavic languages. Therefore, diversity among Finno-Ugric languages should be lower, than diversity among Balto-Slavic languages. We can see the very opposite however. The linguistic distance between Finno-Permian and Ugric languages is extremely far in the first place, and even Permian languages are so distinct, that some scholars view them as an entirely separate branch of Finno-Ugric languages.³⁵⁹

The time and geographic range of the Seima-Turbino migrations coincide with the dating and geographic range of haplogroup N1b (N-P43Ö, however since haplogroup N1b is estimated to be about 4000 years old, and spread northwards and westwards from Southern Siberia³⁶⁰ exactly the same time and same way as the Seima-Turbino migrations did. This dating and geographical range may coincide with a narrower language group of Uralic however: the Samoyedic languages. They are indeed thought to originate from Southern Siberia, and their diversity support a dating of about 4000 years ago. As we can see above, Samoyedic languages did cover most of today's Komi Republic and Arkhangelsk Oblast as recently as 1600 years ago, so Samoyedic language and haplogroup N1b are plausible linguistic and genetic candidates to be associated with the Seima-Turbino migration.

Other factors that seem to exclude the mainstream view:

If we believe that mainstream view that haplogroups R1a and R1b, and the Corded Ware culture represented Indo-European Slavic speaking cultures, we should assume, that a Southwards Finno-Ugric migration somehow "reconquered" the lands of the Estonians, The Votes, the Latvians, the Merya, the Muroma, the Meshcher, and the Mordvins. The main problem with this scenario, is that, as described above, about 1500-1600 years ago, only Sámi and Samoyedic hunter-gatherer population existed beyond the 60th parallel north. These populations were marked by a far lower population density, and a less advanced technology and level of political association than their southern neighbors. Sámi and Samoyedic populations had neither the demographic, nor the technological resources to conquer far more numerous agriculturalist societies on their southern frontier.

³⁵⁷Koryakova, Ludmila, and Andrej Epimakhov. *The Urals and western Siberia in the Bronze and Iron ages*. Cambridge University Press, 2007.

³⁵⁸Koryakova, Ludmila, and Andrej Epimakhov. *The Urals and western Siberia in the Bronze and Iron ages*. Cambridge University Press, 2007.

³⁵⁹Kittilä, Seppo, Katja Västi, and Jussi Ylikoski, eds. *Case, Animacy and Semantic Roles*. Vol. 99. John Benjamins Publishing, 2011. p. 354.

³⁶⁰Derenko, Miroslava, et al. "Y-chromosome haplogroup N dispersals from south Siberia to Europe." *Journal of Human Genetics* 52.9 (2007): 763-770.

Conclusion

As described above, the conclusion of this paper is that haplogroups R-Z283, R-M458, I2a, N1c, and N1b were all carriers of different language groups. N1c has likely been the dominant haplogroup among a pre-Uralic native population of the region, while Finno-Permian languages were likely introduced in the region by R-Z283 with the eastern flank of Corded Ware migrations in the Bronze Age. N1b likely introduced Samoyedic languages with the Seima-Turbino migrations 3500 years BP, while Balto-Slavic languages were likely introduced by a population characterized by a blend of R-M458 and I2a 3000-400 years ago.

This explanation solves several questions ignored by the mainstream view: How Finno-Slavic linguistic boundary could be that much in the south as recently as early medieval times, despite the far more northern spread of both Corded Ware culture, and haplogroup R1a? How come that in most Finno-Permian populations the presence of haplogroup R is as high as that of haplogroup N, etc.

We have to admit, that this paper not only solves, but also raises questions. How come, that as different language groups, as Finno-Permian and Slavic could originate from two different subclades, R-Z283 and R-M458 of the same haplogroup, haplogroup R1a, while Samoyedic, a language group somewhat related to Finno-Permian, could originate from haplogroup N1b? These questions can be subject of further research, but language shift can be a plausible explanation. The parallel spread of haplogroup I2a (originating from further south) together with R-M458 could indicate such an early language shift for example. The number of questions solved is still higher however, than the questions raised.

References:

- Aikio, Ante. "An essay on substrate studies and the origin of Saami." *Etymologie, Entlehnungen und Entwicklungen: Festschrift für Jorma Koivulehto zum 70* (2004): 5-34.
- Aikio, Ante. "New and old Samoyed etymologies 2." *Finnisch-ugrische Forschungen Band 59-Heft 1-3* (2006).
- Derenko, Miroslava, et al. "Y-chromosome haplogroup N dispersals from south Siberia to Europe." *Journal of Human Genetics* 52.9 (2007)
- Homén, Theodor: *East Carelia and Kola Lapmark*. Longmans, Green and Company, 1921.
- Kittilä, Seppo, Katja Västi, and Jussi Ylikoski, eds. *Case, Animacy and Semantic Roles*. Vol. 99. John Benjamins Publishing, 2011
- Kittles, Rick A., et al. "Autosomal, mitochondrial, and Y chromosome DNA variation in Finland: Evidence for a male-specific bottleneck." *American journal of physical anthropology* 108.4 (1999): 381-399.
- Koryakova, Ludmila, and Andrej Epimakhov. *The Urals and western Siberia in the Bronze and Iron ages*. Cambridge University Press, 2007.
- Pounds, Norman JG. *An Historical Geography of Europe Abridged Version*. CUP Archive, 1990.
- Rozhanskii, Igor L., and Anatole A. Klyosov. "Haplogroup R1a, its subclades and branches in Europe during the last 9,000 years." *Advances in Anthropology* 2 (2012): 139.
- Underhill, Peter A., et al. "The phylogenetic and geographic structure of Y-chromosome haplogroup R1a." *European Journal of Human Genetics* (2014).
- Wiik, Kalevi, and György Kádár. *Az európai népek eredete*. Nap Kiadó, 2008.
- European Y-DNA haplogroups frequencies by country- Eupedia http://www.eupedia.com/europe/european_y-dna_haplogroups.shtml
- Haplogroup N1c (Y-chromosomal DNA) – Eupedia http://www.eupedia.com/europe/Haplogroup_N1c_Y-DNA.shtml#subclades
- Map of Finno-Ugric Distribution 900 AD <http://peacecountry0.tripod.com/900ad.htm#oz>
- Map of Finno-Ugric Distribution 400 AD <http://peacecountry0.tripod.com/400ad.htm>