

# R1A SUBCLADES AND BRONZE AGE MIGRATIONS ON THE EURASIAN STEPPES

*Csaba Barnabas Horvath, PhD*

Pátmány Péter Catholic University, Budapest, Hungary

---

## Abstract

Until recently, our knowledge of the prehistoric world has been blur in many aspects. It was unclear, whether if migrations, conquest, or merely cultural diffusion led from one culture to the other. By the beginning of the 20<sup>th</sup> century, the introduction of genetics into the study of ancient history gave opportunity for a deeper insight to the prehistoric world. As more and more genetic traces are explored, it becomes more and more possible to identify the routes and time of prehistoric migrations. If we can decode the information hidden in genetic data, a turbulent picture of migrations and conquest between certain prehistoric cultures emerges in front of our eyes. This paper attempts to identify prehistoric migrations in Eastern Europe and Central Asia from the Neolithic to the Bronze Age, based on statistical data of genetic samples.

---

**Keywords:** Prehistory, migrations, genetics, archeology, Eurasia

## Introduction

This paper aims to identify prehistoric migrations in Eastern Europe and Central Asia from the Neolithic till the Bronze Age, based on genetic material. For this purpose, the paper mainly uses statistical samples of human Y-chromosome haplogroups. The reason why human Y-chromosome haplogroups can be useful for such research is that they are distinct by the non-recombining parts of Y-chromosome DNA, and this draws a patrilineal phylogenetic tree, and also the age of certain mutations is identifiable. This way their statistical presence in certain human populations can provide indications on origins and migrational backgrounds of certain human populations. The aim of the paper is to analyze apparent spatial and chronological coincidences between Y-DNA haplogroups and certain prehistoric cultures in Western Eurasia, in an attempt to identify prehistoric migrations in the analyzed area. The analysis makes an attempt to attach certain prehistoric cultures to certain YDNA haplogroups, thus identifying prehistoric migration routes. For this purpose, now the present paper

analyzes the geographic spread of subclades of haplogroup R1a1a, a haplogroup widely present in Northern Europe, Eastern Europe, and Central Asia. The paper attempts to test a main hypothesis that based on certain matches in three different factors:

- present day geographic distribution of Y-DNA haplogroups including the geographic distribution of microsatellite variance within distinct haplogroups and subclades that suggests the geographic origin of the group in question,
- the age of distinct Y-DNA haplogroups and subclades, that has already been determined by biologists specialized on the issue
- the age and geographic spread of certain prehistoric cultures in the region

A fourth scale where we should also see matches is ancient DNA, ie. DNA samples taken from sites of prehistoric cultures. This fourth factor however is problematic in a sense that as of now, ancient DNA is still sporadic, with only few, sporadic available samples from some cultures, and no samples at all from some other cultures, where in many cases even the samples that are available are characterized by uneven geographic, chronologic and demographic distribution. Therefore, hitherto aDNA samples were not available regarding most prehistoric cultures in an abundance and representative distribution enough to draw a clear picture, and often the sporadic data of different cultures seem to contradict each other in some cases, that will certainly be clarified as soon as data will be abundant, but what is still the case as of now. In the second part of my paper I will still attempt to test the hypothesis in scope of ancient DNA, and take a view on alternate scenarios as well.

### **Migrations unveiled? – Hypothesis**

As described in the introduction, the main hypothesis of the paper tries to identify geographical and chronological matches between the age and present geographic distribution of certain haplogroups and subclades, and the geographic range of certain prehistoric cultures. The earliest split in the spread of R1a1a in Europe seems to be the split between the subclade marked by the SNP L664, and the rest of the group. This split between L664 the rest of R1a1a-M417 around 5000BC<sup>100</sup> matches in time with the emergence of the Rössen and Cucuteni cultures around 4600 BC and 4800 BC. Also the present day geographical distribution of L664 in Central Europe largely coincides with the range of the Rössen culture along the Danube river. Subclades of non-L664 R1a1a-M417 are much younger west

---

<sup>100</sup> R1a1a & Subclades – Y SNP & STR Classification System and project Sections, Family Tree DNA 2014 <http://www.familytreedna.com/public/R1a>

of the line between Rössen and Cucuteni, and are related to older R1a1a subclades of Eastern Europe. Also, as we will see later in this paper, both the range and age of those matches with the Corded Ware Culture, that appeared at this part of Europe much later than the estimated time of the L664-nonL664 split. This geographical and chronological match would indicate a common ancestor for the Rössen and Cucuteni cultures. This seems to suggest some location on the northern periphery of the Balkans, and the western periphery of the Eurasian steppes, somewhere around present day Hungary, Romania and Serbia, but this part is hard to specify.

The next match in chronological order is the split between R-Z282 and R-Z93, that around 4000 BC<sup>101</sup> matches both in space and in time with the emergence of the Eastern European Yamna and the Central Asian Afanasevo cultures, as both emerged around 3500 BC. The geographical match is also apparent, with the core area of R-Z282 being in Eastern Europe, as that of Yamna, and the core area of R-Z93 being in Central Asia, same as that of Afanasevo. In the latter case, present day microsatellite variance within the geographical distribution of Z93 even specifically suggests a region as the very origin of the subclade that is at the same time the core area of Afanasevo findings.<sup>102</sup>

The third phase of matches is the split of R-Z282 into younger subclades. Most of this coincides with the spread of the Corded Ware Culture, which is also confirmed by recent findings of ancient YDNA from that culture, with most of the samples being R1a1.<sup>103</sup>

So if we put these pieces of the puzzle together, all this suggests a prehistoric migration, dominated by Haplogroup R1a1a-M417 that seemingly originated from prehistoric Balkans. The hypothetic prehistoric migration suggested by these facts seems to be the following: At first, a population carrying M417 would have spread and built up its' demographic momentum somewhere on the borderlands between the northern edge of the Balkans, and the southwestern edge of the Pontic Steppe. Then, around 4800-4600 BC, two flanks of migration would have started from this area, one westwards, along the Danube river, dominated by R1a1a-L664, establishing the Rössen culture, and an other one eastwards, establishing the Cucuteni culture. After an other one thousand years, two more eastwards waves of migration seems to have spread from the Cucuteni area: One to Central Asia,

---

<sup>101</sup> R1a1a & Subclades – Y SNP & STR Classification System and project Sections, Family Tree DNA 2014 <http://www.familytreedna.com/public/R1a>

<sup>102</sup> Underhill, Peter A., et al. "The phylogenetic and geographic structure of Y-chromosome haplogroup R1a." *European Journal of Human Genetics* (2014).

<sup>103</sup> <http://www.nature.com/nature/journal/v522/n7555/full/nature14507.html>

marked by Z93,<sup>104</sup> and establishing the Afanasevo culture, and an other one to the Pontic Steppe of present day Ukraine and Southern Russia, dominated by Z282,<sup>105</sup> and establishing the Yamna culture. The next possible phase may have been a migration northwards from Yamna, covering most of Europe north of the Carpathians and east of the Rhine, forming the Corded Ware Culture.

One more interesting issue in this aspect is the case of the Andronovo culture. Although most of the area covered by the Andronovo culture are regions where Z93 is predominant within R1a, the historical core area from where Andronovo spread, was in the Ural-Idel region, ie. closer to Yamna and Corded Ware territory than to Afanasevo, and in a region, where Z282 make up the vast majority of indigenous R1a today. This would suggest a Z282 superstrate and a Z93 substrate forming Andronovo. Among later subclades, the geographical range of the Volga-Carpathian subclade suggests an origin linked not to the Corded Ware culture, but to the Scythian cultures originating from Andronovo, as it is spread along the Pontic steppes, in Southern Russia, Ukraine, Hungary, and Poland. At the same time, the Volga-Carpathian subclade, as a subclade of Z282, suggests Z282 presence in the Scythian population thus, in the Andronovo culture as well.

### **Problematic issues, and alternate scenarios**

The age, and present day geographic range of certain subclades of R1a seem to clearly support the hypothesis presented in this paper, ie. a prehistoric migration, starting from the Northern Balkans, then splitting to Rössen and Cucuteni, then from Cucuteni splitting to Yamna and Afanasevo, and from Yamna, splitting to Corded Ware and Andronovo. Yet, as stated in the introduction, one more important criteria to verify the hypothesis is ancient DNA, ie. DNA from archeological sites of the named prehistoric cultures. Once ancient DNA from these cultures will be abundant, that will give the final word to decide this question once and for all. Yet, as of now, ancient Y-DNA samples are still sparse and sporadic and fragmental from some of the cultures in question (Yamna, Corded Ware, Andronovo) and almost entirely absent from the rest (Vinča, Rössen, Cucuteni, Afanasevo).

The main issue that causes the problem here is that aDNA data regarding the origins of R1a1a in Europe is so fragmental, that it makes it difficult to interpret. Ancient samples do match present day distribution as

---

<sup>104</sup> R1a1a & Subclades – Y SNP & STR Classification System and project Sections, Family Tree DNA 2014 <http://www.familytreedna.com/public/R1a>

<sup>105</sup> R1a1a & Subclades – Y SNP & STR Classification System and project Sections, Family Tree DNA 2014 <http://www.familytreedna.com/public/R1a>

far back in time, as the Corded Ware<sup>106107</sup> and Andronovo<sup>108</sup> cultures, with ancient DNA in both of these cultures matching the distribution of R1a subclades. Before that however, the picture becomes unclear.

Issue starts to become complicated with Yamna, as among the ancient YDNA samples from this culture, there is not a single R1a1a.<sup>109</sup> At first look, we could conclude that this would put into doubt the entire hypothesis of this paper. If we take a closer look however, the case of Yamna is not that clear at all. First, most of these samples are of a limited number, and all from a single location near Samara, in Russia, what was the eastern geographic periphery of the Yamna culture, and no samples are available so far from the core area of the culture, what was in present day Ukraine. Second, ancient autosomal DNA analysis do show clear descent of the Corded Ware population from Yamna.<sup>110</sup> So although not by YDNA but by autosomal DNA, the Yamna-piece of the puzzle also clearly matches the rest of the pieces, at least in a sense that the R1a dominated population of Corded Ware clearly originates from Yamna.

Where the issue really becomes unclear, is the issue of cultures before Yamna. Basically, there are no more, than four R1a samples altogether from cultures before the era of Yamna, ie. before 3500 BC, in a distribution that puts the geographical origins of European R1a1a in question. These are samples from the subclade R1a1 from Mesolithic Russia, with one sample from Karelia, from the Kunda culture from between 5500-5000 BC,<sup>111</sup> and three samples from the Narva culture from a location near Serteya, along the Western Dvina river (not to be confused with the Northern Dvina), one of them from 4000 BC, and the other two from 2500 BC.<sup>112</sup>

---

<sup>106</sup> Allentoft, Morten E., et al. "Population genomics of Bronze Age Eurasia." *Nature* 522.7555 (2015): 167-172.

<sup>107</sup> Haak, Wolfgang, et al. "Ancient DNA, Strontium isotopes, and osteological analyses shed light on social and kinship organization of the Later Stone Age." *Proceedings of the National Academy of Sciences* 105.47 (2008): 18226-18231.

<sup>108</sup> Hollard, Clémence, et al. "Strong genetic admixture in the Altai at the Middle Bronze Age revealed by uniparental and ancestry informative markers." *Forensic Science International: Genetics* 12 (2014): 199-207.

<sup>109</sup> Haak, Wolfgang, et al. "Massive migration from the steppe was a source for Indo-European languages in Europe." *Nature* (2015).

<sup>110</sup> Haak, Wolfgang, et al. "Massive migration from the steppe was a source for Indo-European languages in Europe." *Nature* (2015).

<sup>111</sup> Haak, Wolfgang, et al. "Massive migration from the steppe was a source for Indo-European languages in Europe." *Nature* (2015).

<sup>112</sup> Chekunova E.M., Yarceva N.V., Chekunov M.K., Mazurkevich A.N.: Te first results of genetic typing of local population and ancient humans in Upper Dvina region, in: THE state hermitage museumrussian academy of sciencesinstitute for the history of material cultureherzen state universityumr 8215 cnrs. Trajectoires archaeologyof lake settlementsiv–ii mill. Bc: chronology of cultures,environmentand palaeoclimatic rhythms materials of

These samples represent R1a1, the parent subclade of R1a1a, the R1a subclade most common in Europe, but are not part of R1a1a itself. There is one more single sample that could still be R1a. This fifth sample is an R1 sample from thousands of kilometers away of these, from the Funnelbeaker-related Baalberge culture, from a site near Esperstedt in Germany, from between 3887-3797 BC.<sup>113</sup> The sample is of the haplogroup R1, the parent group of both R1a and R1b. This sample has already been tested for the SNP M515 with a negative result, therefore it is not part of R1a1a. It hasn't been tested for the SNP M420 yet, therefore it could still be R1a. What is interesting about this sample, that it is the earliest R1 sample from Central Europe, and no other R1 sample, neither R1b nor R1a has been found from before a 1000 year later than this sample. With the absence of any R1a or R1b sample from within 1000 kilometers and 1000 years, it is hard to predict, how this single sample of R1 connects to the history of R1 and R1b in Europe.

Four or five samples of course, don't really mean anything on their own, since uncharacteristic haplogroups in small numbers are commonly present in human populations, and early diverged small subclades often drifted away geographically from their parent haplogroup. Therefore, the presence of both the R1a1 samples from the Narva and Kunda Cultures and the R1 sample from the Baalberge Culture could have very well been the result of such phenomena. The problem is the hitherto absence of R1a even in samples from prehistoric cultures where data is relatively abundant, and where its' present day microsatellite variance would suggest its' origins. Present day microsatellite variance would suggest the origin of European R1a somewhere around present day Turkey and Iran.<sup>114</sup> This, according to its' present distribution in Europe, with higher microsatellite variance in Central Europe than in Eastern Europe or Central Asia, would clearly suggest a route through Anatolia and the Balkans. However among ancient YNDA data gained from 26 samples of the Anatolian Neolithic, R1a (as well as R1b) is entirely absent,<sup>115</sup> and it is also absent from Neolithic cultures of the Balkans, such as the Stračevo,<sup>116</sup> Vinča,<sup>117</sup> and Lengyel<sup>118,119</sup> cultures.

---

international conference dedicated the semi-centennial anniversary of the researches of lake dwellings in north-western russia saint-petersburg, 13–15 november 2014

<sup>113</sup> Haak, Wolfgang, et al. "Massive migration from the steppe was a source for Indo-European languages in Europe." *Nature* (2015).

<sup>114</sup> Underhill, Peter A., et al. "The phylogenetic and geographic structure of Y-chromosome haplogroup R1a." *European Journal of Human Genetics* (2014).

<sup>115</sup> Mathieson, Iain, et al. "Eight thousand years of natural selection in Europe." *bioRxiv* (2015): 016477.

<sup>116</sup> Szécsényi-Nagy, Anna, et al. "Tracing the genetic origin of Europe's first farmers reveals insights into their social organization." *Proceedings of the Royal Society of London B: Biological Sciences* 282.1805 (2015): 20150339.

Even if the number of samples is relatively few from these cultures, the fact that R1a is missing from all of them seems to be too much for a mere coincidence. The only other possible geographic route could be through Central Asia and Eastern Europe, a scenario of which the presence of the four R1a1 samples from the Kunda and Narva cultures may be signs. This scenario is however also contradicted by some ancient DNA data, as the Comb Ware Culture seems to have been dominated by haplogroup N1c, that likely arrived from Siberia or Central Asia the same time when that culture appeared in the region,<sup>120</sup> and autosomal DNA data also defines the population of Corded Ware originating from Yamna,<sup>121</sup> and not from Comb Ware. Although R1a1 samples from the Narva and Kunda cultures suggest R1a1 presence in the region from before the arrival of N1c, it seems that by the height of Comb Ware Culture, N1c was already predominant in the region, and therefore, Comb Ware Culture seems to be a highly unlikely candidate to be the source of R1a1a predominance in the Corded Ware and Andronovo Cultures. R1a is also absent from the Funnelbeaker culture,<sup>122,123</sup> that geographically was the direct predecessor of the Corded Ware in its' own region, so not only Comb Ware, but also Funnelbeaker seems to be an unlikely candidate for that role.

Since R1a is not only absent from Neolithic Balkan cultures, but it is also absent from all cultures in Central Europe before Rössen, and Cucuteni ie. before 4800 BC, (most notably from western LBK<sup>124,125</sup>) little space seem to remain in Early Neolithic Europe for the role of the original homeland of

---

<sup>117</sup> Szécsényi-Nagy, Anna. *Molecular genetic investigation of the Neolithic population history in the western Carpathian Basin*. Diss. Mainz, Univ., Diss., 2015, 2015.

<sup>118</sup> <https://genetiker.wordpress.com/2015/08/23/y-snp-calls-for-another-neolithic-hungarian-genome/>

<sup>119</sup> Szécsényi-Nagy, Anna. *Molecular genetic investigation of the Neolithic population history in the western Carpathian Basin*. Diss. Mainz, Univ., Diss., 2015, 2015.

<sup>120</sup> Reconstructing Genetic History of Siberian and Northeastern European Populations; Anton Valouev, Emily

HM Wong, Andrey Khrunin, Larissa Nichols, dmitrypushkarev, Denis Khokhrin, Dmitry Verbenko, Oleg Evgrafov, James Knowles, John Novembre, Svetlana Limborska  
 biorxiv doi: <http://dx.doi.org/10.1101/029421>

<sup>121</sup> Haak, Wolfgang, et al. "Massive migration from the steppe was a source for Indo-European languages in Europe." *Nature* (2015).

<sup>122</sup> Haak, Wolfgang, et al. "Massive migration from the steppe was a source for Indo-European languages in Europe." *Nature* (2015).

<sup>123</sup> <https://genetiker.wordpress.com/2015/08/23/y-snp-calls-for-another-neolithic-hungarian-genome/>

<sup>124</sup> Ancient, D. N. A. "Ancient DNA from European early neolithic farmers reveals their near eastern affinities." (2010).

<sup>125</sup> Haak, Wolfgang, et al. "Massive migration from the steppe was a source for Indo-European languages in Europe." *Nature* (2015).

R1a1a. So while R1a1a is predominant in ancient DNA samples from the Corded Ware and Andronovo cultures, in accordance to its' present day geographical distribution, it seems to have been almost entirely absent from everywhere from before that. To put it sharply, it seems to have arrived from nowhere, to dominate Bronze Age Eastern Europe and Central Asia.

If our approach is to find cultures with R1a abundance, the only space that remains for R1a are cultures from which ancient YDNA samples are as of now unavailable, and that are at the same time the direct predecessors of the Corded Ware and Andronovo cultures, where R1a1a is already proven to have been predominant. Cultures fulfilling these criteria are the Rössen, the Cucuteni, the Dnieper-Donets cultures. Those neighbors of theirs, where ancient YDNA is available (Vinča, Comb Ware) seem to be dominated by different haplogroups than R1a, and so are their direct predecessors, from where data is available (LBK<sup>126127</sup> for Rössen) and at the same time they are already followed by cultures marked by R1a predominance. Therefore Rössen, Cucuteni and Dnieper Donets cultures seem to be the only possible candidates for the origins of R1a abundance in Europe.

Out of the Rössen, Cucuteni, and Dnieper Donets Cultures, Rössen can be interesting only in the context of the subclade R1a-L664. As described in the hypothesis, both the age and present day geographic distribution of L664 matches with Rössen culture. Regarding this issue, we have two questions: One is whether if really Rössen is the most likely candidate for L664, and the other one is on what route did it take before arriving to its' present location.

First, let us have a look of the scenario, if L664 spread via Rössen. What is problematic with this scenario is that if the split between L664 and the rest of haplogroup R1a1a was the split between Rössen and Cucuteni, then there should be a common source for Cucuteni. However, Vinča culture, that seems to be the most likely possible candidate for that, lacks R1a from its' ancient YDNA samples.<sup>128</sup> According to its' geographic location one more possible candidate for that role could be the Eastern Linear Pottery Culture from present day Hungary and Romania. No ancient YDNA is available yet from this culture, so R1a predominance is still possible for it.

This however, seems to be a weak form of evidence, since no data is available from this cultures yet, and the lack of R1a from Vinča samples

---

<sup>126</sup> Ancient, D. N. A. "Ancient DNA from European early neolithic farmers reveals their near eastern affinities." (2010).

<sup>127</sup> Haak, Wolfgang, et al. "Massive migration from the steppe was a source for Indo-European languages in Europe." *Nature* (2015).

<sup>128</sup> Szécsényi-Nagy, Anna. *Molecular genetic investigation of the Neolithic population history in the western Carpathian Basin*. Diss. Mainz, Univ., Diss., 2015, 2015.



seem to put the option of Rössen being R1a rather unlikely. So we should also take a view on the scenario of L664 not being from Rössen. Since already relatively large numbers of ancient YDNA samples are available from LBK proper, and Funnelbeaker Cultures, and R1a seem to be almost entirely absent from both (except for the described R1 sample from the Baalberge Culture, but since Baalberge is geographically the southwestern periphery of the Funnelbeaker Culture, adjacent to Rössen, so that sample can rather indicate R1 presence in Rössen, as R1 is entirely absent from all other regions of the Funnelbeaker horizon.) but at the same time, aDNA samples from Corded Ware already shows the present day distribution of R1a1a, then in case if L664 did not come with Rössen, it could have only arrived with Corded Ware. What is against this scenario however is that in this case, L664 should be at least sporadically present in Eastern Europe, along the route of the Yamna and Corded Ware migrations. If L664 had spread from Yamna through the migration that formed the Corded Ware culture, then it should be present among the Corded Ware ancient YDNA samples as well. The fact that it is not present, suggests that it had to split from the rest of R1a1a, and migrated to Central Europe before the Corded Ware culture had emerged. One more option for the spread of L664 could be a bottleneck, ie. small number of people carrying L664 moving as a tiny minority within Corded Ware, and then going through a demographic boom after arriving to present day Germany. In this case however, this should be visible in the microsatellite variance of L664 today in a way, that most of its' subclades should originate from a boom that happened in the age of the Corded Ware, and at the same time, its' microsatellite variance should be significantly higher in present day Poland, Belarus and Ukraine, than in present day Germany and Western Europe. This is however, not the case, as we know today.

This suggests that even despite all the other problematic issues, based on all the data that is available to us as of now, the option of L664 having spread with Rössen is still most likely scenario.

We must note, that due to the low and thus unrepresentative number of samples, aDNA in many cases can be misleading, therefore present day geographic patterns can not be left out of consideration. One apparent example for that is the case of the Urnfield culture. The Urnfield Culture covered much of present day Germany and France, a region where subclades of haplogroup R1b are predominant today. Among the 14 pieces of ancient YDNA samples from that culture however, the vast majority is from haplogroup I2a2b, and only one single sample is R1b.<sup>129</sup> This could make us

---

<sup>129</sup> Schilz, Felix. "Molekulargenetische Verwandtschaftsanalysen am prähistorischen Skelettkollektiv der Lichtensteinhöhle." (2006).

to draw a conclusion that the population of the Urnfield Culture, ie. the population of South Germany and Eastern France was not yet dominated by R1b yet at that time, and that abundance of R1b originates from some mass migration that arrived to the region only after the Urnfield era. Such a conclusion would however be clearly wrong, since more numerous and diverse aDNA samples from the Bell Beaker Culture, that dominated the very same region before Urnfield, already shows the clear predominance of R1b, same as it is in the human population of the region today.<sup>130131132</sup> This means that R1b already became predominant in the region by the era of the Bell Beaker Culture, exactly how it is suggested by the age and present day geographic distribution, as well as microsatellite variance of R1b subclades in the region, and the aDNA samples from the Urnfield Culture were misleading due to their low, thus unrepresentative number and geographical diversity. In our case, the fact, that only as few as six samples are available from the Vinča culture,<sup>133134</sup> (what would be the most likely candidate for a common origin of L664 and the rest of R1a according to the present day geographic distribution and microsatellite variance of these groups) but none of those are R1a. However the fact that R1a is absent not only from Vinča, but from samples from all those cultures that preceded it on the Balkans, still make the Balkan route for R1a arriving to Europe in the Neolithic unlikely, at least according to data that we have now. An other possible source of confusion could be the bottleneck phenomenon, when people carrying a certain SNP, represents only a tiny fraction of a certain population, but in a certain phase of the migration of that population of a part of it, their number suddenly explodes, and becomes dominant. In our case Yamna culture of present day Ukraine and southern Russia seem to have been proven to be such. Ancient DNA samples from the Yamna culture, as well as from cultures following it on the Pontic Steppe seem to largely lack R1a until as far as as the Srubnaya Culture, that appeared only around 1900BC, almost one millennium after the end of Yamna.<sup>135</sup> On the other hand however, autosomal DNA analysis showed the population of Corded Ware Culture, the

---

<sup>130</sup> Allentoft, Morten E., et al. "Population genomics of Bronze Age Eurasia." *Nature* 522.7555 (2015): 167-172.

<sup>131</sup> Haak, Wolfgang, et al. "Massive migration from the steppe was a source for Indo-European languages in Europe." *Nature* (2015).

<sup>132</sup> Lee, Esther J., et al. "Emerging genetic patterns of the European Neolithic: perspectives from a late Neolithic Bell Beaker burial site in Germany." *American journal of physical anthropology* 148.4 (2012): 571-579.

<sup>133</sup> Szécsényi-Nagy, Anna. *Molecular genetic investigation of the Neolithic population history in the western Carpathian Basin*. Diss. Mainz, Univ., Diss., 2015, 2015.

<sup>134</sup> <http://www.y-str.org/p/ancient-dna.html>

<sup>135</sup> Mathieson, Iain, et al. "Eight thousand years of natural selection in Europe." *bioRxiv* (2015): 016477.

largest source of R1a dominance in Europe to clearly originate from Yamna.<sup>136</sup> This suggests a bottleneck for R1a in the Yamna population, after which it went through a demographic explosion during the Corded Ware migrations. Such scenarios are still possible for R1a in other regions as well, such as the Cultures of the Balkans, since a large portion of the Yamna population is also proven by autosomal analysis to originate from the Middle East,<sup>137</sup> and Neolithic techniques from that region most likely arrived to the Yamna area via the Balkans. The presence of some R1a1 samples in the Kunda and Narva cultures could also suggest such an R1a bottleneck in the otherwise likely N1c dominated Comb Ware culture, but since autosomal analysis shows the Corded Ware population originating from Yamna, and not from Comb Ware, this scenario for the origins of R1a in present day Poland and Germany seems to be unlikely, except for a scenario where an early migration from Central Asia acts as the common source of R1a in all the Comb Ware, Yamna, Dniepr-Donets, and Eastern LBK Cultures, then submerging into a bottleneck there.

## Conclusion

As a conclusion to this paper, we can state that the present geographic range of different R1a haplogroups strongly suggests the prehistoric migration described in the hypothesis of this paper, and even in the scope of ancient DNA samples, it still seems to be the most likely, or at least the least unlikely scenario of those presented. This suggests a prehistoric migration most likely starting from the Eastern LBK culture of present day Hungary and Romania. Then in the next step, splitting into a western and an eastern flank around 5000-4500 BC, with the western flank marked by the L664 subclade of R1a1a, and founding the Neolithic Rössen culture (4600-4300 BC) along the Danube river as far west as Germany, and an eastern flank marked by non-L664 R1a1a, founding the Chalcolithic Cucuteni culture (4800-3000 BC) in present day Romania, Moldova and Ukraine. Then in the third phase, a new wave of migration starting from Cucuteni, and spreading throughout the Eurasian steppes, establishing the early Bronze Age Yamna and Afanasevo cultures on the pontic steppe, and in Central Asia during the 4<sup>th</sup> millennium BC, supposedly marked by the Z282 and Z93 subclades of R1a1a. In the fourth phase, two flanks of Bronze Age migration seem to have spread out from Yamna culture during the 3<sup>rd</sup> millennium BC : the Corded Ware culture Northwards, into the temperate forest belt of Northern and Eastern Europe, and the Andronovo culture

---

<sup>136</sup> Haak, Wolfgang, et al. "Massive migration from the steppe was a source for Indo-European languages in Europe." *Nature* (2015).

<sup>137</sup> Haak, Wolfgang, et al. "Massive migration from the steppe was a source for Indo-European languages in Europe." *Nature* (2015).

eastwards, shading over the Afanasevo culture in Central Asia. available evidence strongly suggest the occurrence of this complex and large scale chain-migration, but final word can only be said when ancient YDNA will be abundant from all these prehistoric cultures, but such conditions for research are still to be achieved. Of course, if such a migration did occur, that raises questions regarding the linguistic identity of the certain waves of this migration, to identify of which present day language family, were these waves the ancestors. Many would suggest the ancestors of the Indo-European language family, but certain facts suggest that the Uralic family could have also played a major role in early R1a migrations.<sup>138</sup> This however, could be the topic of an other paper.

---

<sup>138</sup> Horváth, csaba barnabás. "the story of two northward migrationsorigins of finno-permic and balto-slavic languages in northeast europe, based on human y-chromosome haplogroups337." *euopean scientific journal* 10.10 (2014).