

# Haplogroup R1a, Its Subclades and Branches in Europe during the Last 9000 Years

Igor L. Rozhanskii, Anatole A. Klyosov\*  
The Academy of DNA Genealogy, Newton, USA  
Email: \*aklyosov@comcast.net

Received May 5<sup>th</sup>, 2012; revised June 11<sup>th</sup>, 2012; accepted June 25<sup>th</sup>, 2012

This study identifies and describes 38 branches of the haplogroup R1a STR haplotypes which currently exist in Europe or which migrated from Europe to areas in the east, south, and southeast between 6000 and 4500 years before the present (ybp). The study is based on 2471 haplotypes which have been tested for either 67- or 111-markers; it essentially creates a unified robust system, which assembles dozens of R1a-SNPs and thousands of STRs and assigned haplotypes to branches, some of which do not have SNP assignments as yet. The assembled system consists of base (deduced ancestral) haplotypes, one for each STR branch and for each SNP-assigned subclade, each with its characteristic (ancestral) set of alleles, arranged in the chronological space from ~9000 ybp to 1300 ybp. We found that the most ancient R1a subclades (R1a1-M198- and R1a1a-M198+/M417-) bearers of which currently live in Europe (the present day haplotypes are scattered between England and the Balkans) appeared in Europe at least 7300 ybp, and possibly 9000 ybp. R1a's three principal downstream subclades, L664 (North-Western branch), Z93 (South-Eastern branch), and Z283 (Eurasian branch), split from their common European ancestor at about the same time, around 7000 - 6000 ybp. L664 apparently stayed in North-Western Europe; its lineage recovered and began expanding ~4575 ybp. The Z93 subclade began to expand during the Aryan migrations, on the Aryan's journey to India and the Middle East in the 3rd-2nd millennia BC. The Z283 subclade split ~5500 ybp into three branches. One of them, Z280 (the Central Eurasian branch) moved east to the Russian Plain in 4800 - 4600 ybp, and formed at least 16 sub-branches there and in the course of the later westward repopulation of Europe in the 1st millennium BC—1st millennium CE. Some of the older branches, like the Russian Plain branch, largely stayed in the present Russia-Ukraine-Belarus-Poland-Baltic countries region, and were described by early historians as the Scythians, Antes, Veneti, and a multitude of different proto-Slavic tribes (though many of them belonged to haplogroups other than R1a, primarily I1 and I2). Those R1a branches which are "older" than 3000 years, such as the Russian Plain branch (4600 ybp), the Western Eurasian (4300 ybp), and the Balto-Carpathian (4300 ybp), did not move *en mass* to Europe but stayed behind at the Russian Plain. In the middle of 1st millennium CE, the time of the collapse of the Roman Empire, multiple migrations of R1a were taking place eastward and westward; these migrations gradually formed the current landscape of R1a in Europe. All 38 branches and their datings are listed in the Appendix of this paper; current distribution maps are shown in the body of the paper.

**Keywords:** Y Chromosome; Mutations; Haplotypes; Haplogroups; TMRCA; STR; SNP; R1a1

## Introduction

The preceding article in this journal (Klyosov & Rozhanskii, 2012) provided data in support of the theory that haplogroup R1a arose in Central Asia around 20,000 years before the present (ybp), and described the prior history of the haplogroup, which is directly related to the appearance of Europeoids (Caucasoids) ~58,000 years ago.

At some point, apparently between 20 and 15 thousand ybp, the bearers of R1a began a migration to the west, through Tibet and over the Himalayas. They arrived in Hindustan no later than 12,000 ybp. They apparently continued their way across the Iranian Plateau, along East Anatolia and the rest of Asia Minor between 10,000 and 9000 ybp. By ~9000 ybp they arrived in the Balkans and spread westward over Europe and to the British Isles. At that point, R1a still had DYS392 = 13 in most haplotypes, as did its sister haplogroup R1b. This marker is very slow to change, and mutates on average once in 3500

conditional generations (that is once in 87,500 years in one DNA-lineage; the concept of conditional generations of 25 years in each is described in "Materials and Methods" section). More than 99% of the R1a haplotypes known today bear the mutation giving DYS392 = 11. Somewhere along this extended timescale, bearers of R1a or its upstream haplogroups developed the Pre-Proto-Indo-European (Pre-IE) language and carried it along during their journey from Central Asia to Europe. Linguists date the earliest signs of Pre-IE in Anatolia at 9600 ± 600 ybp (Gamkrelidze & Ivanov, 1995; Gray & Atkinson, 2004; Renfrew, 2000), which coincides with the data of DNA genealogy for the migration of R1a bearers across Asia Minor that is described in Klyosov and Rozhanskii (2012).

Some known archaeological cultures in the Balkans and Central/Eastern Europe, dated 8000 - 7000 ybp (Bug-Dniester, Vinča, Starčevo, LBK, etc.), can be attributed, as least in part, to bearers of R1a. Yet they also can be attributed to other ancient haplogroups, such as I, J, E, G.

As the bearers of haplogroup R1b1a2 began to populate

\*Corresponding author.

Europe after 4800 ybp (the Bell Beakers and other R1b1 migratory waves, including, perhaps the Kurgan people, though their identification and haplogroup assignment remain unclear, and likely represented both R1b and R1a bearers in different time periods), haplogroup R1a had moved from Central and North-Western Europe to the Russian Plain around 4800 - 4600 ybp (Klyosov, 2011). From there, R1a migrated to the south, east, and south-east as the historic Aryans (Klyosov & Rozhanskii, 2012).

We have to emphasize that the definition of the Aryans in DNA genealogy is broader than the definition of Aryans in linguistics and archaeology. According to linguists and archaeologists, Aryans were the ancient people who spoke Indo-European Aryan or “steppe” languages; this early group split at some point into the Indo-Aryan and Iranian languages. When and where that split occurred is a matter of debate; linguists generally claim that the split occurred immediately before the Aryans moved to India and Iran, while some archaeologists believe that it happened a thousand years before that.

In DNA genealogy, with its emphasis on genetic heritage, the Aryans, as the putative bearers of Indo-European dialects, also clearly carried haplogroup R1a, which apparently was brought to India and Iran coincident in time with the spread of Indo-European (IE) dialects in those same regions. That time is generally recognized around 3500 ybp, since studies of Gordon Childe (Childe, 1926) and even earlier. In other words, R1a likely represented a substantial portion of the Aryan linguistic and cultural migration, the roots of which we can trace back to Europe and to earlier migrations, following R1a ancient movements. As soon as we accept such a definition, the time of the split of the Aryans into “Indo-Aryans” and “Iranians” becomes evident: migration is swift but language change is slow. The “language split” must have occurred before the Aryan entrance into Anatolia and northern Syria (Mitanni). This reasoning suggests that the split was around 4500 ybp—as the archaeologists predicted.

Today’s linguists find IE languages in Anatolia and Mitanni, the South Urals, Iran, India, beyond the Ural Mountains, and in South Siberia. All these languages have the same Aryan roots. These peoples developed common horse-breeding terminology and shared essentially the same vocabulary for household items, and for gods and religion, although sometimes, apparently, the shared words ended up with contrasting meaning, as in India and Iran, where some ancient Aryan good and evil gods sometimes had opposite meanings.

Currently, most of descendants of the ancient peoples who carry European R1a live in Eastern Europe, primarily in Russia (up to 63% of the population) and Poland, Ukraine, Belarus (up to 57% of the population in the last three countries) (Underhill et al., 2009; Balanovsky et al., 2008; Kayser et al., 2005; Kharkov et al., 2004, 2005).

Haplogroup R1a and its regional distribution and history were a subject of many studies, which have been reviewed in the preceding paper (Klyosov & Rozhanskii, 2012). R1a was identified in excavated DNA in Germany with the dating of 4600 ybp (Haak et al., 2008), in Siberia, with the dating of 3800 - 3400 ybp (Keyser et al., 2009), in Tarim Basin, Xinjiang, China, with the dating of ~4000 ybp (Li et al., 2010). Its distribution was described in various parts of Eurasia (e.g., Wells et al., 2001; Behar et al., 2003; Cinnioglu et al., 2004; Kharkov et al., 2004; Sengupta et al., 2006; Balanovsky et al., 2008; Sharma et al., 2009; Underhill et al., 2009; Abu-Amero et al.,

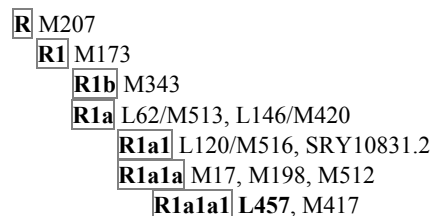
2009). However, only in the last few years, thank mainly to efforts of Dr. Peter Underhill’s laboratory at Stanford University, Dr. Thomas Krahn of Family Tree DNA’s Genomics Research Center, and researchers of 1000 Genomes Project, many new SNPs in haplogroup R1a have been discovered; as indicated by beginning letters M, L, and Z, respectively, in the SNPs below. Their base haplotypes, mutation patterns and history have not been described in the literature as yet. Apparently, only one attempt was made regarding R1a-M458 subclade (Underhill et al., 2009), however, the authors employed the erroneous “evolutionary mutation rate” (see Klyosov & Rozhanskii, 2012, and references therein) and obtained a grossly overestimated date for M458 appearance as 10,700 ybp in Eastern Europe (Poland), about 250% higher than the figure obtained in our studies. Commonly, 250% - 300% exaggeration in historical datings is a “feature” of the “evolutionary mutation rate” method (ibid.).

This study focuses on DNA lineages, or branches, of haplogroup R1a in Europe, from the Atlantic to the Ural Mountains, and on the history of their appearance and migrations across and from Europe in the last ~9000 years (Appendix 1). The methodology of the study, based largely on considerations of extended 67- and 111-marker haplotypes, and involving principal methods of DNA genealogy, is described in detail in the preceding papers in this journal (Rozhanskii & Klyosov, 2011; Klyosov & Rozhanskii, 2012) and in the Materials and Methods section of this article. Available databases containing tens of thousands of 67-marker haplotypes are listed in Rozhanskii and Klyosov (2011) and in this paper (Appendix 2).

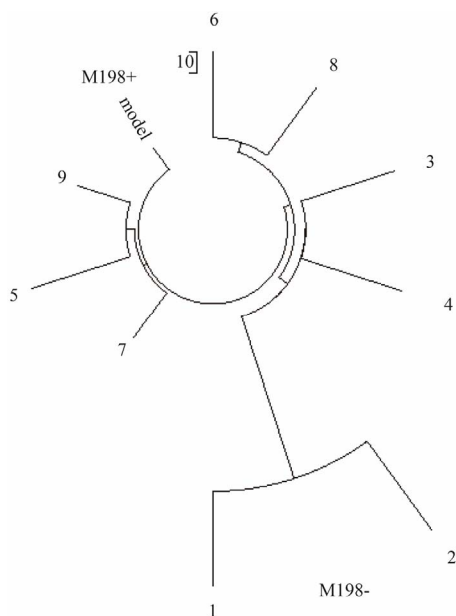
### The Old European Branch ( $\geq 9000$ ybp) (R1a-SRY10831.2+, M417-)

The Old European Branch of R1a in Europe has a distinct ancient DYS392 = 13, as does its sister haplogroup R1b (see the diagram below). However, while in R1b this allele is by far the predominant one, DYS392 = 13 occurs in R1a in only 0.4% of all haplotypes. All others have DYS392 = 11. Actually, only twelve DYS392 = 13 haplotypes have been identified thus far in the 67-marker haplotype format, and nine of them form the Old European branch (the haplotype tree of this branch is shown in **Figure 1**).

The following diagram shows the upper part of the haplogroup tree with respect to R1a and its most ancient subclades (R1b is shown for a comparison) (ISOGG-2012, a fragment, [http://www.isogg.org/tree/ISOGG\\_HapgrpR.html](http://www.isogg.org/tree/ISOGG_HapgrpR.html)) in its current classification:



The two haplotypes at the bottom of the tree in **Figure 1** are from Belarus and the US, both have positive SNP SRY10831.2 and negative M198, which firmly places them into the R1a1 subclade on the haplotype tree above. They two have their combined haplotype (here and throughout the paper haplotypes are given in the FTDNA format):



**Figure 1.**

The 67-marker 9 haplotype tree for the Old European Branch in Europe. This and all other haplotype trees in this paper are composed from haplotypes listed in the IRAKAZ database (Rozhanskii et al., 2012). The upper 7-haplotype branch has SNP M417-, M198+ (R1a1a), the lower 2-haplotype branch has M198-, SRY-10831.2+ (R1a1).

13 25 15 10/11 12 12 12 13 12 13 13 31/33—16 9 9 11 11  
24/25 16 20/21 29/34 12/13 16 16 16—10/11 11/12 19 23 15  
15/16 17/20 17/18 35 38 11/12 11—11 8 15 17 8 11 10 8 11/12  
10 12 22 22 17 10 12 12 15 8 12 22/23 22/23 13 12 11 13/14  
10/11 11 12 12/13

The 25 mutations between them are translated to  $25/.12 = 208 \rightarrow 262$  conditional generations (25 years each), or 6550 years. If they belong to the same DNA-lineage, their most recent common ancestor lived ~3300 ybp. If they belong to different lineages, their common ancestor would have lived much earlier.

The upper, rather tight series of haplotypes in **Figure 1** all have positive M198, but negative M417, which places them in the R1a1a subclade. Their base haplotype is as follows:

13 25 15 11 13 14 12 12 10 14 13 31—16 9 10 10 11 25 14  
19 31 12 15 15 15—10 11 19 23 16 16 17 19 36 38 11 11—11  
8 17 17 8 12 10 8 12 10 12 22 22 15 10 12 12 13 8 13 23 22 12  
12 11 13 11 11 12 12

All seven haplotypes have 78 mutations from the base haplotype, which gives  $78/.12 = 93 \rightarrow 103$  conditional generations, or  $2575 \pm 390$  years to their common ancestor. However, this date actually points to a time when the population expanded. In reality, the base haplotype shown above differs from the two M198- haplotypes by 40 and 41 mutations respectively; if they belonged to the same subclade, that would set them apart by at least 12,300 and 12,800 years. Since they belong to different subclades (R1a1 and R1a1a), their common ancestors lived more than 6400 years before the present, and not less than the age of the R1a1a-M17 subclade (at least 9000 ybp).

We do not know exactly where their common ancestor lived, since members of this branch are scattered in Europe from the Balkans to the British Isles (**Map 1**). It may be that they are

descendants of those R1a bearers who came to the Balkans ~9000 ybp at the European end of their migration route from Central Asia (Klyosov & Rozhanskii, 2012).

In this paper, unless we specifically describe a particular subclade with its particular SNP, we will use R1a as a common overall designation of all downstream subclades of the haplogroup.

## Principal European R1a Subclades

As was described above and in Klyosov and Rozhanskii (2012), R1a arose around 20,000 ybp in Central Asia, migrated (in part) to Europe over the following 11,000 years, and arrived in Europe, as R1a1 and R1a1a subclades, more than 7 thousand years or, more probably, ~9000 ybp. There are only a few descendants of these ancient subclades which are currently identified in Europe.

The next downstream subclade in Europe was R1a1a1-M417, which arose  $\geq 7600$  ybp (see below). Its principal downstream subclades are as follows:

**R1a1a1** M417  
**R1a1a1i** L664- (the Tenth, NW branch)  
**R1a1a1i** L664+  
**R1a1a1h** Z93 (South-Eastern branch)  
**R1a1a1g** Z283 (Eurasian branch)  
**R1a1a1g1** M458 (European branch)  
**R1a1a1g2** Z280 (Central Eurasian)  
**R1a1a1g3** Z284 (Scandinavian)

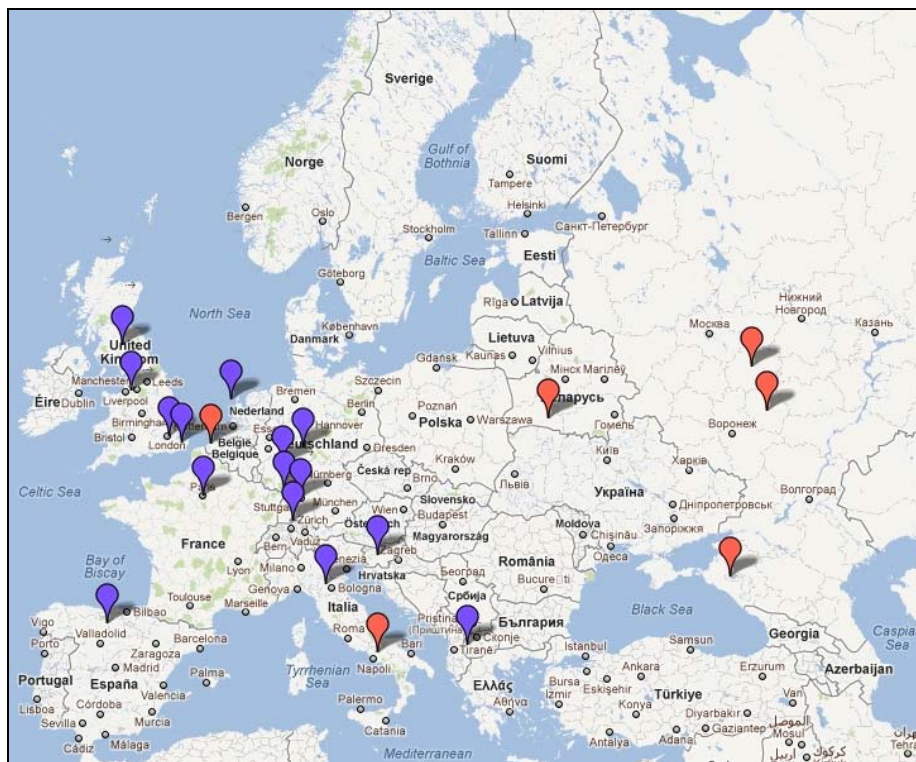
The Tenth, or the North-Western branch, is the oldest subclade of R1a-M417, identified thus far. It arose at least 6300 ybp (see below) and has a distinct mutation DYS388 = 10 as distinct from the DYS388 = 12 in majority of R1a haplotypes. Subclade Z93 separated from the rest of R1a around 5700 ybp. It migrated from Europe or from the Russian Plain no later than 4800 ybp. The migration was to the south (to the Middle East) and south-east (Iran and India), where bearers of Z93 are the predominant R1a bearers today. Subclade Z283 arose ~5500 ybp and populated Eastern Europe, part of Central Europe, and further east to the Ural Mountains.

The overall 111-marker R1a haplotype tree is shown in **Figure 2**. The purpose of the Figure is not in details of positions of individual haplogroups, but to show the overall shape, configuration of the tree, and relative positions of its subclades. A few of M420 and M417 haplotypes, available in the 111-marker format, sit at the very base/core of the tree, since they descended from the most ancient European R1a ancestors. The principal subclades of the Z283 Eurasian super-branch are located around the tree—Z280, Z284, M458. Some of them consist of two or several more sub-branches, which are described below.

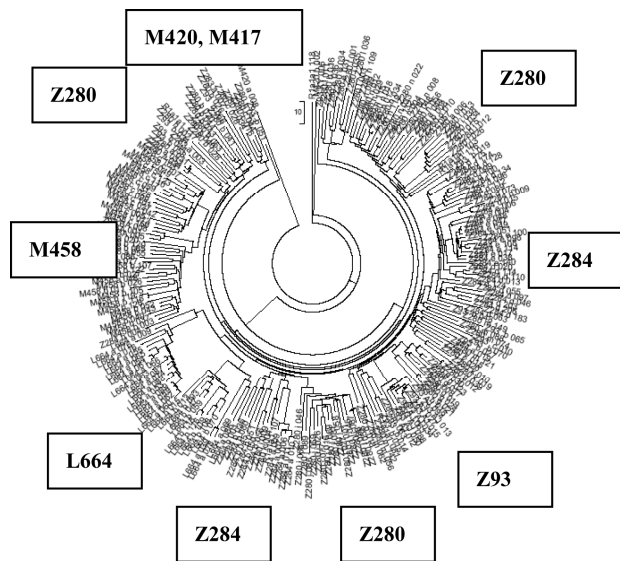
The main purpose of presenting the overall tree in **Figure 2** is to show that the SNP (single nucleotide polymorphism) subclades and STR (short tandem repeat) branches form a robust mutually supporting system. They are two sides of the proverbial coin; both are necessary for a justified, meaningful, and advanced analysis of haplotype datasets.

### North-Western Branch (the Tenth), 4575 ybp

This is one the most distinct of R1a branches. All haplotypes



**Map 1.**  
The Old European Branch, referred to birthplaces of the furthestmost ancestors recorded in available databases. Red pins denote a branch with SNP M198-, SRY10831.2+. Blue pins mark a branch with M417-, M198+.



**Figure 2.**  
An 111-marker haplotype tree of R1a haplogroup, composed of 251 haplotypes. Most of the tree is taken by the Eurasian super-branch Z283, of which branches Z280 (Central Eurasian), Z284 (Scandinavian), and M458 (European) are shown. Another two principal branches are Z93 (South-Eastern) and L664 (North-Western branch, the Tenth). The L664 branch shown at 7 o'clock.

of this branch have DYS388 = 10 (hence, the Tenth), and two-thirds of them are located in England, Ireland and Northern Germany (**Map 2**), and almost the same number of their bearers

live in the USA (with roots mainly in England, Ireland, Scandinavia, Germany).

The haplotype tree (**Figure 3**) splits into four major lineages, or sub-branches. One is the old branch (with a common ancestor of ~4300 ybp), which in turn splits to subclades L664- (4100 ybp; though, this SNP seems to be rather unstable) and L664+ (4300 ybp), then the young branch, with L664+ (2250 ± 250 ybp), and two DYS448 null-mutated branches, of 1500 ± 325 ybp and its downstream of 260 ± 60 ybp, the last marks the arrival of the Tenth to the USA (Klyosov, 2010a, 2010b, 2012).

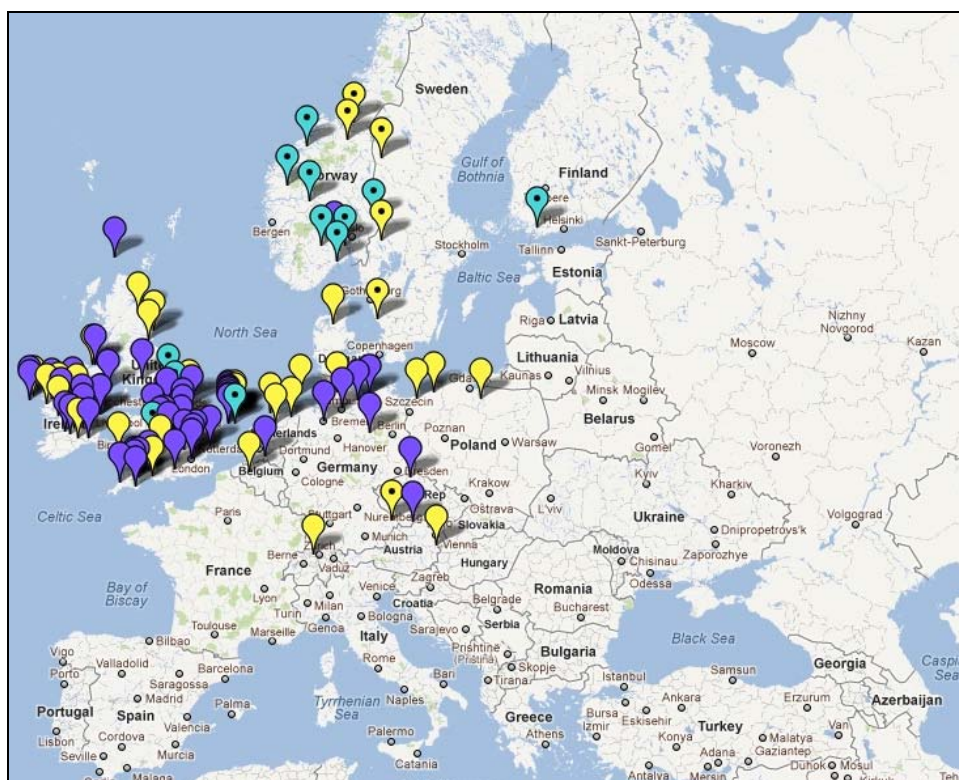
The base 111-marker haplotype of the old branch splits into two sub-branches as follows:

13 25 16 10 11 14 12 10 10 13 11 30—15 9 10 11 11 24 14  
19 31 12 15 15 16—11 11 19 24 16 15 19 20 33 39 12 11—11  
8 17 17 8 12 10 8 11 10 12 22 22 16 10 12 12 14 8 14 23 22 12  
13 11 13 11 11 12 13—31 15 9 16 12 26 27 18 13 12 1212 11  
10 12 11 10 11 11 31 12 15 24 12 9 10 19 15 20 12 23 15 12 15  
25 12 23 19 10 15 18 9 11 11

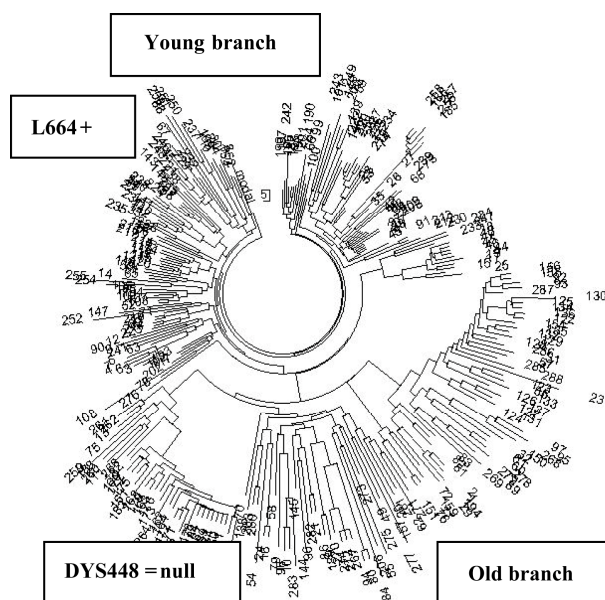
13 25 15 10 11 14 12 10 10 13 11 31—15 9 10 11 11 25 14  
19 31 12 14 14 17—12 11 19 23 15 16 18 19 34 37 12 11—11  
8 17 17 8 12 10 8 11 10 12 22 22 15 10 12 12 13 8 14 23 22 12  
12 11 13 11 11 14 13—32 15 9 15 12 26 27 19 12 121212 10 9  
12 11 10 11 11 31 12 14 24 13 9 10 19 15 20 12 24 15 12 16 24  
12 24 19 10 15 19 9 11 11

which differ by as much as 33 mutations. One branch (shown first) is likely L664-, with a common ancestor of 2125 ± 280 ybp (.139 mutations per marker), another, likely L664+, is 2025 ± 260 years "old" (.133 mutations per marker). The distance between them sets these two lineages apart by  $33/.198 = 167 \rightarrow$





**Map 2.** The map of the North-Western branch, depicting birthplaces of the furthest ancestors recorded in available databases. Yellow pins denote an old branch. Deep-blue and light-blue pins mark younger branches.



**Figure 3.** A haplotype tree of the Tenth's (DYS388 = 10), composed of 285 of 67-marker haplotypes. The dataset was provided by Martin Voorwinden and updated with the Prendergast family series in January 2012. The younger branch on the left all has SNP L664+, where determined; the branch on the 3 o'clock (the Prendergast family with a common ancestor ~1000 ybp) has both L664+ and L664-, which raises a question on technical (methodological) stability of the L664-. The old branch has both L664+ and L664-.

200 conditional generations, or 5000 years. This placed their common ancestor by  $(2125 + 2025 + 5000)/2 = 4575$  ybp.

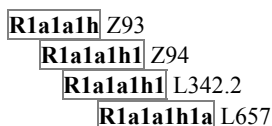
An analysis of a more extended dataset of 67-marker haplotypes of the Tenth's gave 4300 years to their common ancestor, which is the same value within the margins of error, and showed that the divergence of L664+ and L664- branches occurred around 4100 ybp, or almost from the common ancestor of currently living Tenth's (Klyosov, 2012).

The robustness of the calculations despite the small number of extended haplotypes is clear. Twenty-one 67-marker haplotypes show 191 mutations, which give  $191/21/12 = 76 \rightarrow 83$  generations, or 2075 years to a common ancestor. The same haplotypes in the 111-marker format give 324 mutations, that is  $324/21/198 = 78 \rightarrow 85$  generations, or 2125 years to the common ancestor. Obviously, the margins of error of  $2075 \pm 260$  and  $2125 \pm 240$  ybp may be somewhat excessive for the calculations.

The combined 67-marker base haplotype of the Tenth's differs from the Old European base haplotype (see above) by 29 mutations, which places their common ancestor at  $\geq 7600$  years before the present. Since they belong to different subclades, M417- and M417+, their common ancestor lived 7600 ybp only if the "depth" of the M417+ subclade is no more than 7600 years. In that case, the Tenth's actually initiated the M417+ subclade.

### South-Eastern Branch (Z93), ~5700 ybp

According to the haplotype tree (ISOGG-2012)



subclade Z93 has three downstream subclades, Z94, L342.2 and L657. All four subclades populate mainly the Middle East, India, and some locations in the eastern part of the Russian Plain, hence, it is called the South-Eastern branch (**Map 3**). It arose in Europe ~5700 ybp and migrated along with the future Aryans across the Russian Plain to the south, east, and south-east between 5000 and 3500 ybp.

The haplotype tree of the R1a-Z93 clade is shown in **Figure 4**. Its Jewish R1a1 branch includes more than half of the whole tree (104 haplotypes out of 203 haplotypes); all the Jewish haplotypes are related to the downstream subclade L342.2. The Bashkir and the Kyrgyz branches also have the same L342.2. In fact, L342.2 is the predominant subclade in the Z93 clade, embracing 98% of its haplotypes. Only a few haplotypes on the tree and one English lineage of closely related individuals are L342.2 negative, which tells us the initial state of Z93.

Bearers of Z93 are spread from England through Armenia, Oman, and Kuwait, all the way to Khakassia in South Siberia; the latter (or his ancestors) apparently was a migrant from the Russian Plain. Six currently known representatives have the following base haplotype

13 24 16 11 11 15 12 12 10 13 11 30—15 9 9/10 11 11 24 14 20 31 12 15 15 15—11/12 11 19 22/23 16 16 18 17 35 40 13 11—11 8 17 17 8 12 10 8 11 10 12 22 22 15 10 12 12 13 8 13/14 23 21 12 12 11 13 10 11 12 13

All six have 111 mutations from the base haplotype which gives  $111/6 \cdot 12 = 154 \rightarrow 177$  generations, or  $4425 \pm 630$  years to their common ancestor. In fact, they belong to different mi-

nor sub-branches, which split from the trunk at the very beginning of the Z93 subclade, and their actual TMRCA is expected to be 5000 years (or more) before the present.

Most of the haplotype tree is L657 negative. An example of L657- base haplotype of an available 20-haplotype Asian branch in the 111 marker format is given below:

13 25 16 11 11 14 12 12 10 13 11 30—15 9 10 11 11 24 14 20 32 12 15 15 16—11 12 19 23 16 16 18 19 34 38 13 11—11 8 17 17 8 11 10 8 11 10 12 22 22 15 10 12 12 13 8 14 23 21 13 12 11 13 11 11 12 13—32 15 9 15 12 26 27 19 12 12 13 12 10 9 12 11 10 11 11 30 12 13 24 13 9 10 19 15 19 11 23 15 12 15 24 12 23 19 10 15 17 9 11 11

An example of L657+ base haplotype of an available Asian branch is as follows:

13 25 16 10 11 14 12 12 10 13 11 30—16 9 10 11 11 24 14 20 32 12 15 15 17—12 11 19 23 15 17 17 19 36 40 14 11—11 8 17 17 8 11 10 8 11 10 12 22 22 15 10 12 12 13 8 12 23 21 12 12 11 13 11 11 12 12

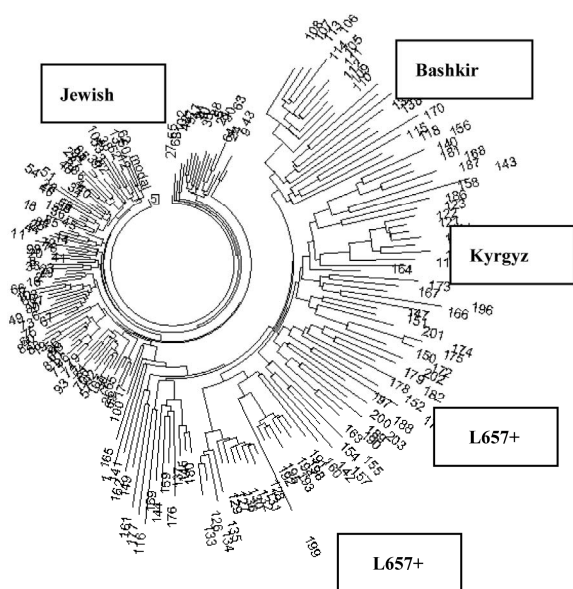
The first base haplotype is apparently ancestral to the branch, consisting of haplotypes from Turkey, Saudi Arabia, Qatar, the United Arab Emirates, Palestine, India, Kazakhstan, Iraq, Pakistan, Poland, Germany, and one from Tatar Russia. 20 of 67-marker haplotypes in the branch contain 354 mutations, which place their common ancestor at  $354/20 \cdot 12 = 148 \rightarrow 174$  generations, or  $4350 \pm 490$  ybp. Fourteen 111-marker haplotypes in the branch contain 381 mutations, which place their common ancestor at  $381/14 \cdot 198 = 139 \rightarrow 162$  generations, or  $4050 \pm 460$  ybp.

The second sub-branch is non-symmetrical, and contains six haplotypes from Saudi Arabia with a common ancestor of only  $315 \pm 110$  ybp, and one very different haplotype from India, with an apparent common ancestor for all L657+ haplotypes of 2700 ybp. There are 16 mutations between the two L657+ and



**Map 3.**

The map of the South-Eastern branch, depicting birthplaces of the furthermost ancestors recorded in available databases. Magenta pins denote a parent Z93 branch, red and green pins correspond to Bashkir and Kyrgyz branches, respectively. The rest of the branch (L342.2 positive) is marked by deep-blue pins. The most represented Ashkenazy Jewish branch (not shown) covers densely an area of historical Polish-Lithuanian Commonwealth.



**Figure 4.** A haplotype tree of the South-Eastern branch (Z93), composed of 203 of 67-marker haplotypes. Practically all haplotypes (except four L342.2-) are assigned to L342.2+ subclade. There are two L657+ branches, of 7 and 17 haplotypes each. The tight, numerous young branch on the left is the Jewish R1a1 branch. The Bashkir (10 haplotypes) and Kyrgyz (7 haplotypes) branches are also shown.

L657- branches in the 67-marker format, which places a common ancestor of the two base haplotypes at 5300 years ago. It might relate to the appearance of the L342.2 subclade.

The R1a-L342.2 Bashkir branch of 10 haplotypes points to their common ancestor of about  $1125 \pm 190$  ybp. Their 111-marker base haplotype is as follows:

13 24 16 11 11 15 12 12 13 11 31—15 9 10 11 11 24 14 20 31 12 15 15 15—11 11 19 23 16 15 18 19 35 39 14 11—11 8 17 17 8 12 10 8 11 10 10 22 22 15 10 12 12 13 8 14 23 21 13 12 11 13 11 11 12 13—33 15 9 15 11 26 28 19 11 12 12 12 10 9 12 11 10 11 12 30 13 14 24 13 9 10 19 15 19 12 24 14 13 15 24 12 23 19 10 15 17 9 11 11

It has 28 mutations compared with the 111-marker predominantly Middle Eastern L657- base haplotype, which places their common ancestor (L342.2) to 4800 years before the present.

The Kyrgyz branch is based mostly on the SMGF database ([www.smgf.org](http://www.smgf.org)), which employs another format, and shares only 34 with the 67 markers reported by FTDNA. Eight Kyrgyz haplotypes are known in the 67- and 111-marker formats, but six of them belong to a rather recent lineage ( $650 \pm 150$  ybp), with possible origin in Altay. Since some alleles in the base haplotype of the entire branch are uncertain, they are marked by “x”:

13 25 16 1111 14 12 12 10 14 11 32—15 9 11 11 23 14 2131 12 1515 16—11 11 19 23 16xxxx15 11—11 8 17 17 8 x 10 8 xx 12 xxx 10 12 12x 8 14 xx 12 12 11 xx 11 12 12

A common ancestor of the Kyrgyz branch lived  $2100 \pm 250$  ybp; this and the other branches of the L342.2 share a common ancestor within the range of  $4400 \pm 500$  ybp, which corresponds to the ancestor of L342.2 itself.

Finally, the large Jewish R1a-L342.2 branch with its base 111-marker haplotype has the following alleles:

13 25 16 10 11 14 12 12 10 13 11 30—14 9 11 11 11 24 14 20 30 12 12 15 15—11 11 19 23 14 16 19 20 35 38 14 11—11 8 17 17 8 12 10 8 11 10 12 22 22 15 10 12 12 14 8 14 23 21 12 12 11 13 10 11 12 13—32 15 9 17 12 27 27 19 12 12 12 10 9 12 11 10 11 11 30 12 12 25 13 9 10 20 15 20 11 23 15 12 15 25 12 23 19 10 15 17 9 11 11

This group descends from a common ancestor who lived only  $1300 \pm 150$  ybp. It arose in Germany, but as part of a lineage with a common ancestor living  $4475 \pm 400$  ybp, or what has been called the “Abraham times” (Klyosov & Rozhanskii, 2012).

It is of interest to compare which base haplotype is closer to the Jewish base haplotype—the Middle Eastern or the Bashkir base haplotype. It would seem that the Middle Eastern haplotype should be closer. Indeed, the mutational distances in the 111-marker format are 30 and 40, respectively. The common ancestor of the Jewish base and the Middle Eastern base, and of the Jewish base and the Bashkir base lived  $4900 \pm 500$  and  $4400 \pm 500$  ybp, respectively—or at the same time, within the margin of error. It fits well with the 4800 ybp obtained above for the L342.2 subclade. The common ancestor of all these L342.2 branches lived at the time of the Aryan migrations from Eastern Europe across the Russian Plain on their way to the east and south.

### Eurasian Branch (Z283), ~5500 ybp

The Eurasian R1a branch Z283 includes haplotypes from Europe and from the Asian part of Russia which spreads to Pacific Ocean on the east. It is also found in low frequencies in Asia Minor and among ethnic Armenians. The Eurasian branch is the most extended one among R1a1 haplotypes, both in number of haplotypes and in its territory. It embraces 86% of all European R1a1 haplotypes, while the Tenth (North-Western branch) take only 12%. The rest are minor subclades, such as the Old Europeans. The Z93 subclade, with its mainly Asian haplotypes, encompasses 10% of the European R1a1 haplotypes, and it was not included in the above figures. Three countries—Poland, Russia and Germany—embrace 55% of all the Eurasian branch, with the rest mainly in Ukraine, Belarus, the Czech Republic, Slovakia, Slovenia, Lithuania, Finland, Hungary, and Italy.

- R1a1a1g** Z283 (Eurasian branch)
- R1a1a1g2** Z280 (Central Eurasian branch)
  - R1a1a1g2a** P278.2 (West Carpathian)
  - R1a1a1g2b** L365 (North European)
  - R1a1a1g2c** L366 (Central Eurasian-3)
  - R1a1a1g2d** Z92 (North Eurasian)
  - R1a1a1g2d** (North Carpathian)
  - R1a1a1g2e** L784
- R1a1a1g3** Z284 (Scandinavian branch)
  - R1a1a1g3a** L448 (Young Scandinavian)
  - R1a1a1g3a1** L176.1 (Scottish Highland)
  - R1a1a1g3b** Z287 (Old Scandinavian)
- R1a1a1g1** M458 (European branch)
  - R1a1a1g1a** Central European branch
  - R1a1a1g1a1** L1029
  - R1a1a1g1b** L260 (West Slavic branch)

The above diagram shows the downstream subclades of



Eurasian branch Z283. The most populous one is the Central Eurasian branch (Z280) with its subclades, which include almost half of all Eurasian (Z283) haplotypes. It is followed first by the Scandinavian branch (Z284 and its subclades), with one-third of the Eurasian haplotypes; and then by M458 (and its subclades) with 21% of the Eurasian haplotypes. A few haplotypes from Canada, Germany, Switzerland, England, Poland and Turkey have been proven to belong to the parent Z283 branch.

### Central Eurasian Z280 Subclade (4900 ± 500 ybp)

Figure 5 shows the 67-marker 801-haplotype tree of the central Eurasian branch (Z280); Figure 6 presents the 111-marker tree of 96 haplotypes, currently available in this extended format. The trees are shown here not for a detailed analysis but for their overall shape. All four available SNPs have their own rather distinct branches on the trees, along with other branches which do not have an assigned SNP as yet (see legend to Figure 5 and 6).

Overall, there are 13 currently identified branches within subclade Z280. A common ancestor of Z280 lived 4900 ± 500 ybp, and had a base (ancestral) 111-marker haplotype as follows:

13 25 16 11 11 14 12 12 10 13 11 30—15 9 10 11 11 24 14  
20 32 12 15 15 16—11 12 19 23 16 16 18 19 35 38 14 11—11  
8 17 17 8 12 10 8 11 10 12 22 22 15 10 12 12 13 8 14 23 21 12  
12 11 13 11 11 12 13—32 15 9 15 12 26 27 19 12 12 12 10  
9 12 11 10 11 11 30 12 13 24 13 9 10 19 15 20 11 23 15 12 15  
24 12 23 19 10 15 17 9 11 11

Some of the Z280 subclade bearers can be assigned, at present, to a rather loose parent branch, the Central Eurasian group proper, descending directly from the above base haplotypes. Base haplotypes of the most populous branches are shown below, along with the timespans to their common ancestors.

### The Russian Plain (RP) Branch, or Central Eurasian-1 (CEA-1) Branch (4600 ± 500 ybp)

13 25 16 11 11 14 12 12 10 13 11 30—15 9 10 11 11 24 14  
20 32 12 15 15 16—11 11 19 23 16 16 19 19 35 38 14  
11—11 8 17 17 8 12 10 8 11 10 12 22 22 15 10 12 12 13 8  
14 23 21 12 12 11 13 11 11 12 13

The RP base haplotype differs from the Z280 base haplotype by only two mutations (425 “lateral” years, that is an average time required for these two mutations to occur), as marked; it places their common ancestor at  $(4900 + 4600 + 425)/2 = 4960$  ybp, which is Z280 itself (4900 ± 500 ybp) within the margin of error.

### Central Eurasian-2 (CEA-2) Branch (3500 ± 400 ybp)

13 25 16 11 11 14 12 12 10 13 11 30—15 9 10 11 11 24 14  
20 31 12 15 15 16—11 11 19 23 16 16 18 19 34 38 14 11—11 8  
17 17 8 12 10 8 11 10 12 22 22 15 10 12 12 13 8 13 23 21 12  
12 11 13 11 11 12 13

The Central Eurasian-2 branch differs from the Russian Plain base haplotype by 4 mutations (marked above). Four mutations separate the common ancestors by  $4/12 = 33 \rightarrow 34$  generations, or 850 years, and their common ancestor lived  $(4600 + 3500 +$

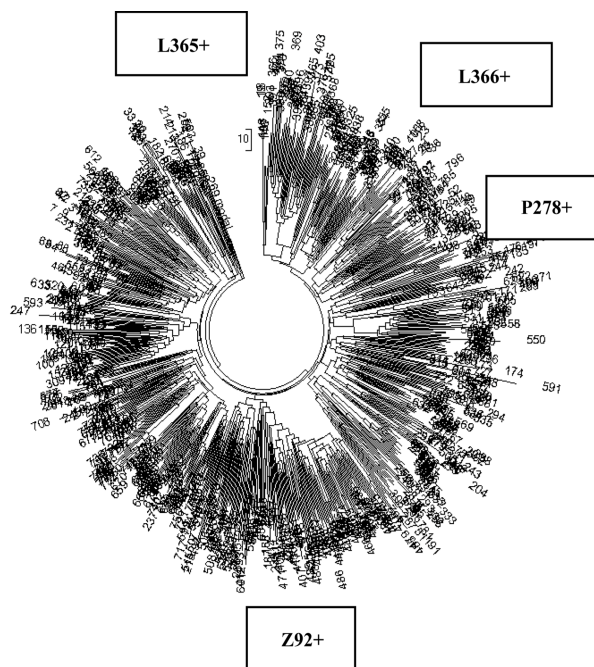


Figure 5. A haplotype tree of the Central Eurasian branch (Z280), composed of 801 of 67-marker haplotypes. The North European (L365), Central Eurasian-3 (L366), West Carpathian (P278.2), and North Eurasian (Z92) branches have their identified SNPs. The tree contains other branches which do not have their SNPs identified as yet, and some of them are as follows, clock-wise: North Carpathian (NC), Balto-Carpathian-1 (BC-1), West Eurasian-3 (WEA-3), Central Eurasian-1 (CEA-1), WEA-1, CEA-2, Eastern Carpathian (EC), BC-3, BC-2.

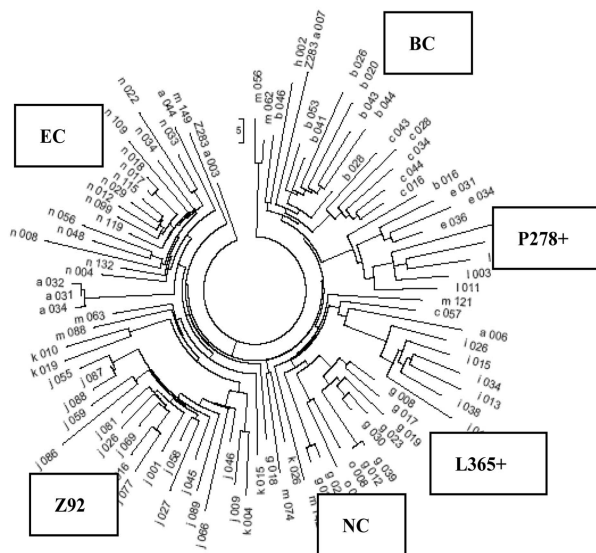


Figure 6. A haplotype tree of the Central Eurasian branch (Z280), composed of 96 of 111-marker haplotypes. The North European (L365), West Carpathian (P278.2), and North Eurasian (Z92) branches are indicated, as well as (clock-wise): Balto-Carpathian (BC), North Carpathian (NC), Eastern Carpathian (EC).

$850)/2 = 4475$  ybp. This is the time of the Russian Plain common ancestor (4600 ± 500 ybp) within the margin of error.



### Central Eurasian-3 (CEA-3) Branch (2500 ± 350 ybp)

13 25 16 11 11 14 12 12 10 13 11 30—15 9 10 11 11 24 14  
20 **31 13** 15 15 16—11 **12** 19 **22** 16 16 **18** 19 **33 39 13** 11—11  
8 17 17 8 12 10 8 11 10 12 22 22 15 10 12 12 **12 8 13** 23 21 12  
12 11 13 11 11 12 13

The Central Eurasian-3 is a minor branch, characterized by SNP L366. It differs from the parent base haplotype (RP branch) by 11 mutations. This places the common ancestor at ~4825 ybp, which is within the margin of error for both the Russian Plain (RP) and the Central Eurasian (Z280) common ancestors.

### Western Eurasian (WEA) Branch (4100 ± 500 ybp)

13 25 16 11 11 14 12 12 **11** 13 11 30—15 9 10 11 11 24 14  
20 **33** 12 15 15 16—11 12 19 23 16 16 18 19 **34** 38 14 11—11 8  
17 17 8 12 10 8 11 10 12 22 22 15 10 12 12 13 8 14 23 21 12  
12 11 13 11 11 12 13

The Western Eurasian base haplotype differs from the Z280 base haplotype by three mutations (650 years), hence, their common ancestor lived 4825 ybp. This is Z280 itself within the margin of error. The WEA branch has a distinct sub-branch (WEA-1) with a common ancestor of 2300 ± 300 ybp. It deviates by 10 mutations from the parent WEA base haplotype, or by 2275 “lateral” years. This places their common ancestor with the parent branch at 4300 ybp, i.e., the WEA parent branch itself (4100 ± 500 ybp).

### Eastern Carpathian (EC) Branch (2600 ± 300 ybp)

13 25 16 11 11 14 12 12 10 13 11 30—15 9 10 11 11 24 14  
20 32 12 15 15 16—11 **11** 19 23 16 16 **17 20 36 39 12** 11—11  
8 17 17 8 12 10 8 **10** 10 12 22 22 15 10 12 12 13 8 14 23 21 12  
12 11 13 11 11 12 13

The base haplotype differs from that of Z280 by 8 mutations, which corresponds to 1800 years, and translates to a common ancestor at 4650 ybp. Again, it is well within margin of error for TMRCA of the entire Central Eurasian branch.

### Western Carpathian (WC) Branch (2600 ± 300 ybp)

13 25 **15 10** 11 **15** 12 12 **11** 13 11 **29**—15 9 10 11 11 24 14  
20 **31** 12 15 15 16—11 12 19 23 **17** 16 **17** 19 35 **39** 14 11—11  
8 17 17 8 12 10 8 11 **11** 12 22 22 15 10 12 12 13 8 **13** 23 21 12  
12 11 13 **12** 11 12 13

Twelve mutations (2775 years) from the base Z280 haplotype place the common ancestor of the Western Carpathian branch at 5100 ybp. Its assignment is also confirmed by SNP P278.2.

The above branches are spread almost uniformly over most of Central and Eastern Europe; however, they are virtually absent in Scandinavia (Map 4).

### Balto-Carpathian (BC) Branch (4300 ± 500 ybp)

13 25 16 **10** 11 14 12 12 10 13 11 30—15 9 10 11 11 24 14  
20 32 **13** 15 15 16—11 12 19 23 16 16 18 **18 34 37 13** 11—11



**Map 4.** The map of the parent Central Eurasian, Western Eurasian, Eastern and Western Carpathian branches, depicting birthplaces of the furthestmost ancestors recorded in available databases. Yellow pins denote the parent branch, pins of other colors correspond to junior branches.

8 17 17 8 12 10 8 11 10 12 22 22 15 10 12 12 13 8 14 23 21 12  
12 11 13 11 11 12 13

The Balto-Carpathian branch deviates by 6 mutations (1325 “lateral” years) from the Z280 base haplotype. The BC branch has two distinct sub-branches (**BC-1** and **BC-2**) with common ancestors of  $2550 \pm 300$  and  $2200 \pm 250$  ybp, respectively. They deviate by 4 and 9 mutations from the parent (BC) base haplotype, or by 850 and 2025 “lateral” years respectively. This places their common ancestor with the parent branch at 3850 and 4300 ybp, or the BC parent branch itself ( $4300 \pm 500$  ybp).

In contrast to the “core” branches of Z280, the Balto-Carpathian shows distinct geographical distribution, as reflected by its name (**Map 5**). Its **BC-2** sub-branch is characteristic for Slovenians and Croatians, being their “trademark” genealogical lineage.

### Northern Eurasian (NEA) Branch (4450 ± 450 ybp)

This branch is characterized by SNP Z92. It consists of two major sub-branches. The older branch (**NEA-1**) ( $3600 \pm 400$  ybp) has the following base haplotype:

13 25 16 11 11 14 12 12 **11** 13 11 30—15 9 10 11 11 24 14  
20 **33** 12 **14** 15 16—11 12 19 23 16 16 18 **18 34 40** 14 11—11  
8 17 17 8 12 10 8 11 10 12 22 22 15 10 12 12 13 8 **13 23 22** 12  
12 11 13 11 11 12 13

Its base haplotype deviates from that of Z280 by 9 mutations (2025 years), which places their common ancestor at 5260 ybp, within the margin of error from  $4900 \pm 500$  ybp for Z280 itself.

The base haplotype of the younger branch (**NEA-2**) ( $2350 \pm 300$  ybp) is:

13 25 16 11 11 **15** 12 12 10 13 11 30—15 9 10 11 **10 25** 14  
20 32 12 **14 14** 16—**12** 12 19 23 **15** 16 18 **20 34** 38 **13** 11—12  
8 17 17 8 12 10 8 11 10 12 22 22 15 10 12 12 13 8 **13 23 22** 12  
12 11 13 11 11 12 13

It has 12 mutations (2775 years) from the base Z280 haplotype and 13 mutations (3025 years) from the NEA-1 base haplotype. The common ancestor of these two sub-branches of Z92 lived  $4450 \pm 450$  ybp.

Present-day bearers of both **NEA** branches populate Lithuania, Belarus, and the north-eastern part of Poland (**Map 6**). It is the most widespread genealogical lineage among ethnic Russians.

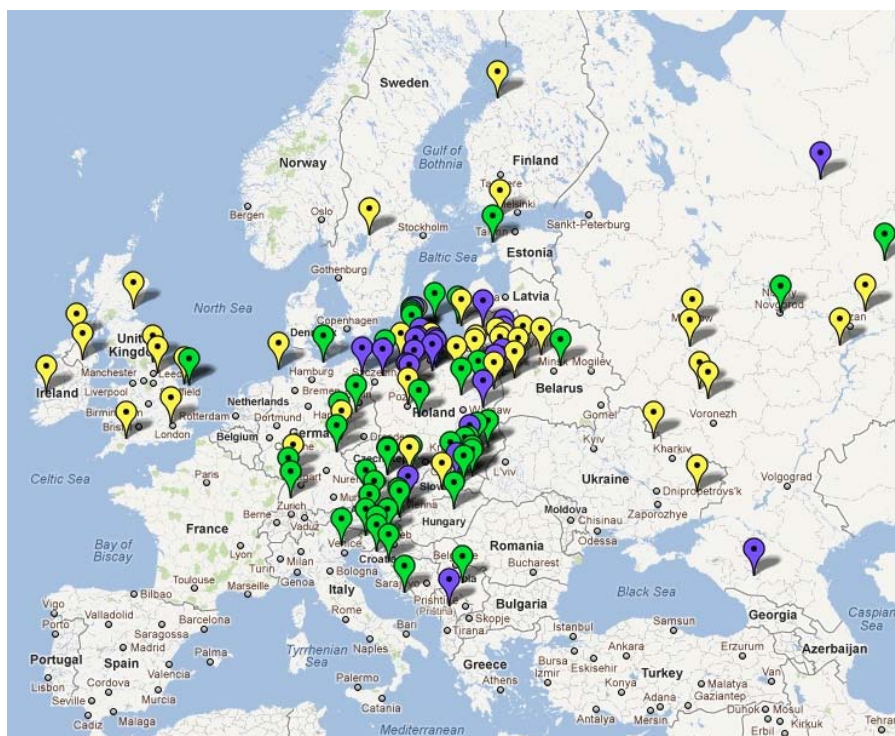
### Northern European (NE) Branch (2600 ± 350 ybp)

13 25 **15** 11 11 14 12 12 10 13 11 30—**17** 9 10 11 11 **23** 14  
20 **33 13** 15 15 16—11 **11** 19 23 16 **15** 18 19 **34** 38 14 11—11  
8 17 17 8 12 10 8 11 10 12 **21** 22 **16** 10 12 12 13 8 14 **24 21 13**  
12 11 13 1111 12 13

The Northern European base haplotype differs by 13 mutations (3025 “lateral” years) from that of Z280, and is independently defined by SNP L365. It again places its common ancestor at 5260 ybp, which is Z280 itself.

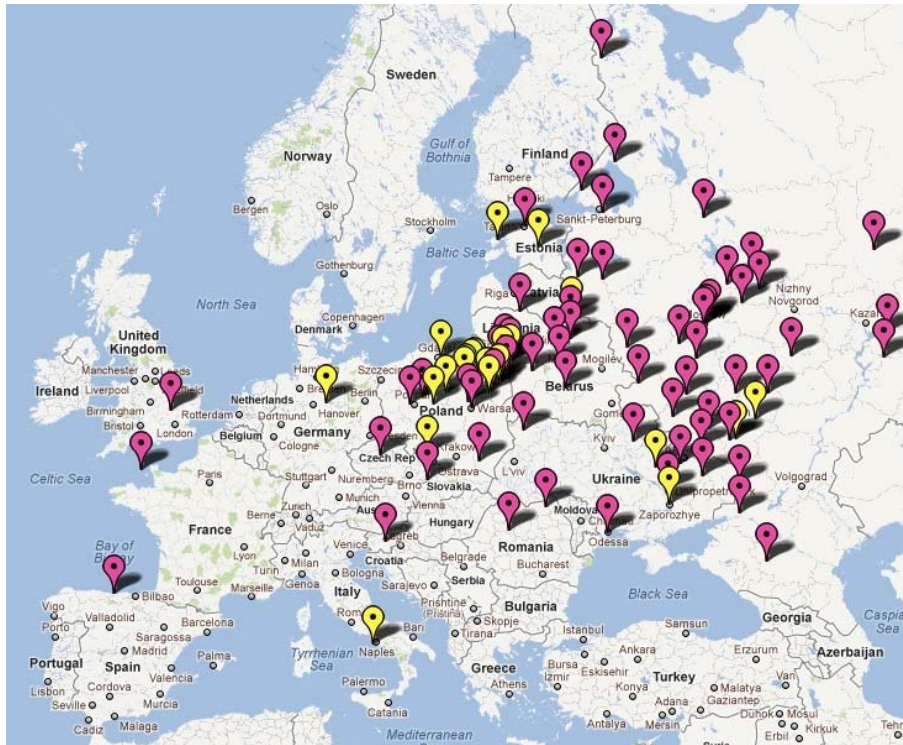
### Northern Carpathian (NC) Branch (2150 ± 300 ybp)

13 25 **17** 11 11 14 12 12 10 13 11 **31**—**14** 9 10 11 11 24 14  
20 32 **13** 15 15 16—11 12 19 23 16 16 18 19 35 **40** 14 11—11  
**9** 17 17 8 12 10 8 11 10 12 22 22 15 10 12 12 13 8 14 23 21 12



**Map 5.** The map of the Balto-Carpathian branch, depicting birthplaces of the furthermost ancestors recorded in available databases. Yellow pins denote the parent branch, deep-blue and green pins correspond to BC-1 and BC-2 sub-branches, respectively.





**Map 6.** The map of the Northern Eurasian branch, depicting birthplaces of the furthest ancestors recorded in available databases. Yellow pins denote the older branch (NEA-1), magenta pins correspond to NEA-2.

12 11 13 11 11 12 13

This base haplotype differs by 7 mutations (1550 years) from the base Z280 haplotype.

There is also a minor branch, which is defined by subclade **L784** (2500 ± 450 ybp):

13 24 17 11 11 14 12 12 10 14 11 30—15 9 10 11 11 24 14  
20 31 13 15 15 16—11 12 19 23 16 17 17 19 34 39 12 12—11  
8 17 17 8 12 10 8 11 10 12 22 22 15 10 12 12 13 8 14 23 21 12  
12 11 13 11 11 12 13

The latter branches seem to be geographically limited to present day Poland and Eastern Germany, and—in lower frequencies—the Baltic countries, Ukraine, and Russia (Map 7).

### Overall Z280 Base Haplotype and Its “Age”

The eight base haplotypes of the main branches (Russian Plain, West Eurasian, Balto-Carpathian, North European, East Carpathian, North Eurasian, North Carpathian, and West Carpathian) differ among themselves by 68 round-up or 64 full mutations (some mutations are fractional), resulting in the Z280 base haplotype, listed above. It gives  $64/8/12 = 67 \rightarrow 72$  generations, or 1800 years down from their average “age” (3100 years). This results in 4900 years from their common ancestor, which is Z280 itself (4900 ± 500 ybp, see above).

### European Subclade M458 (4200 ± 450 ybp)

This subclade consists of two principal branches, the Central European (CE) and West Slavic (WS). A division between them is clearly seen on their haplotype tree (Figure 7).

The CE branch, in fact, consists of two sub-branches, both arose 2900 ± 400 ybp. One of them (CE-1) has the following

base haplotype:

13 25 16 10 11 14 12 12 11 13 11 29—16 9 10 11 11 23 14  
20 32 12 15 15 16—11 11 19 23 17 16 18 19 34 38 14 11—11  
8 17 17 8 11 10 8 12 10 12 21 22 15 10 12 12 13 8 14 25 21 13  
12 11 13 11 11 12 13

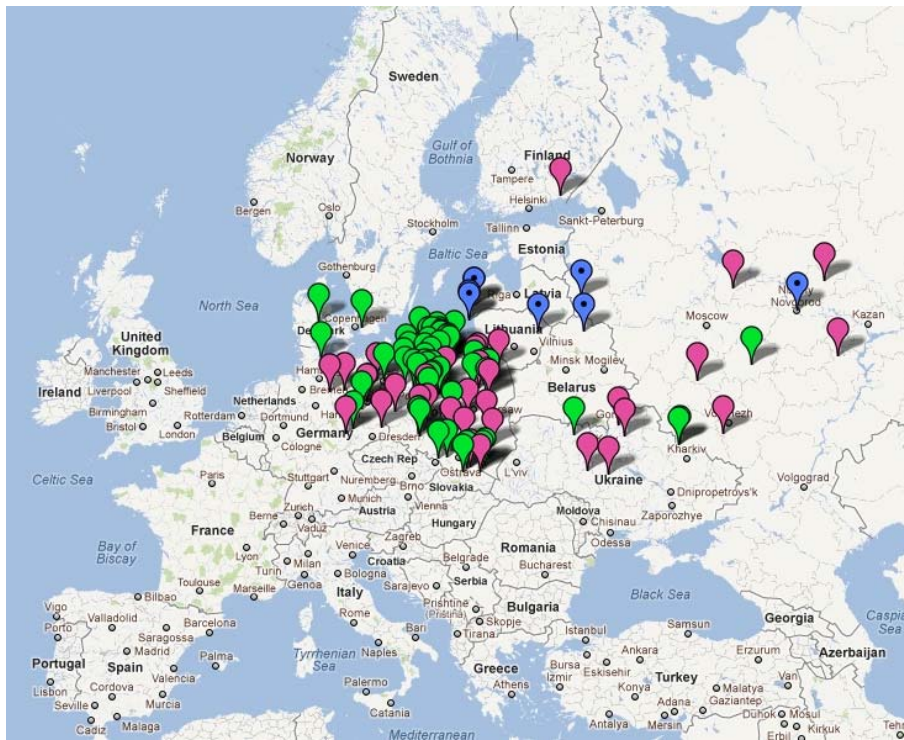
It differs by 14 mutations (3325 years) from the Z280 base haplotype (see above), which places their common ancestors at ≥5600 ybp. The “equal or higher” sign here indicates that the difference between them cannot be less than the “age” of subclades Z280 and M458.

The second branch (CE-2) differs from CE-1 in only two loci, having there three mutations: a multi-copy duplication 12-15-15-16 → 12-12-15-15-15-16 in DYS464 (presumably a one-time event), and a two-step mutation 38 → 40 in CDYb. These three mutations place a common ancestor of CE-1 and CE-2 at 3100 ybp, if both branches arose independently from their common ancestor. It might be thought that CE-2 is a CE-1 downstream branch, which arose immediately after CE-1, within the indicated margin of error in their dating. This is likely, since even the 67 marker haplotype tree cannot cleanly separate the CE branches (see Figure 7). Their haplotypes are nearly identical statistically, except that they descended from slightly different ancestral (base) haplotypes. Also, there is a downstream subclade L1029 within the CE branch which embraces the most of bearers of both the sub-branches. Apparently, it arose at the same time as the CE branch itself.

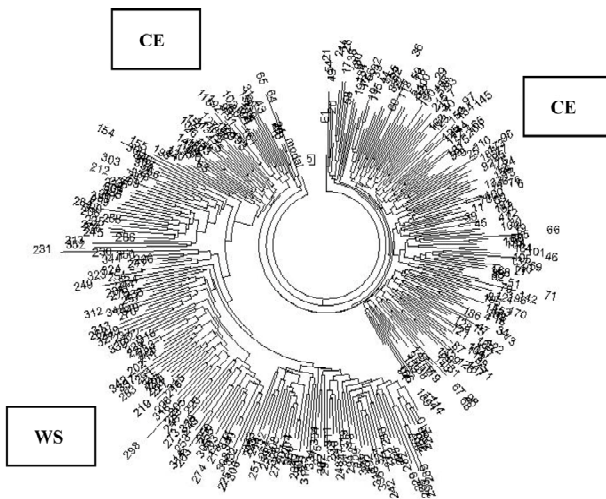
In contrast, the Western Slavic branch (L260) is quite distinct. It does not include any CE haplotype on the tree (Figure 7). It arose 2700 ± 300 ybp, and its base haplotype is as follows:

13 25 17 10 10 14 12 12 10 13 11 30—16 9 10 11 11 23 14  
20 31 12 15 16 16—11 11 19 23 16 16 18 19 34 39 13 11—11





**Map 7.** The map of the Northern European (green pins) and Northern Carpathian (magenta pins) branches, depicting birthplaces of the furthest ancestors recorded in available databases. Blue pins denote the minor branch, which is characterized by SNP L784.



**Figure 7.** A 67-marker haplotype tree of the R1a-M458 subclade, composed of 352 haplotypes. The Central European (CE) branch takes the right-hand side and the upper left-hand side. The distinct Western Slavic (WS) branch is on the left.

8 17 17 8 12 10 8 **12** 10 12 22 22 15 10 12 12 13 8 14 **25** 21 12 12 11 13 11 11 12 13

It differs by 14 mutations (3325 years) from the Z280 base haplotypes, which places their common ancestor at the same  $\geq 5600$  years (see the explanation above). There are 12 mutations between the CE-1 base haplotype and that of the Western Slavic branch, which makes the M458 subclade 4200 years

“old” at least based on available haplotypes.

The European subclade M458 shows nearly the same geographical distribution as the Central Eurasian Z280, but statistically it is more densely represented on its western edge (see also Underhill et al., 2009), reaching up to 70 % of all R1a1 bearers in the Czech Republic (**Map 8**). On the other hand, it is relatively rare among ethnic Lithuanians and Slovenians, who belong predominantly to Z280 subclade.

### Scandinavian Subclade Z284 (4300 ± 500 ybp)

This subclade includes four principal branches, the Old Scandinavian (OS, Z287/Z288, which arose 3700 ± 400 ybp), Young Scandinavian (YS, L448, which arose 2700 ± 350 ybp), the Scottish Highland (SH, L176.1, which arose 1850 ± 290 ybp), and the parent branch, which is negative to the known downstream SNPs (which arose 4300 ± 500 ybp).

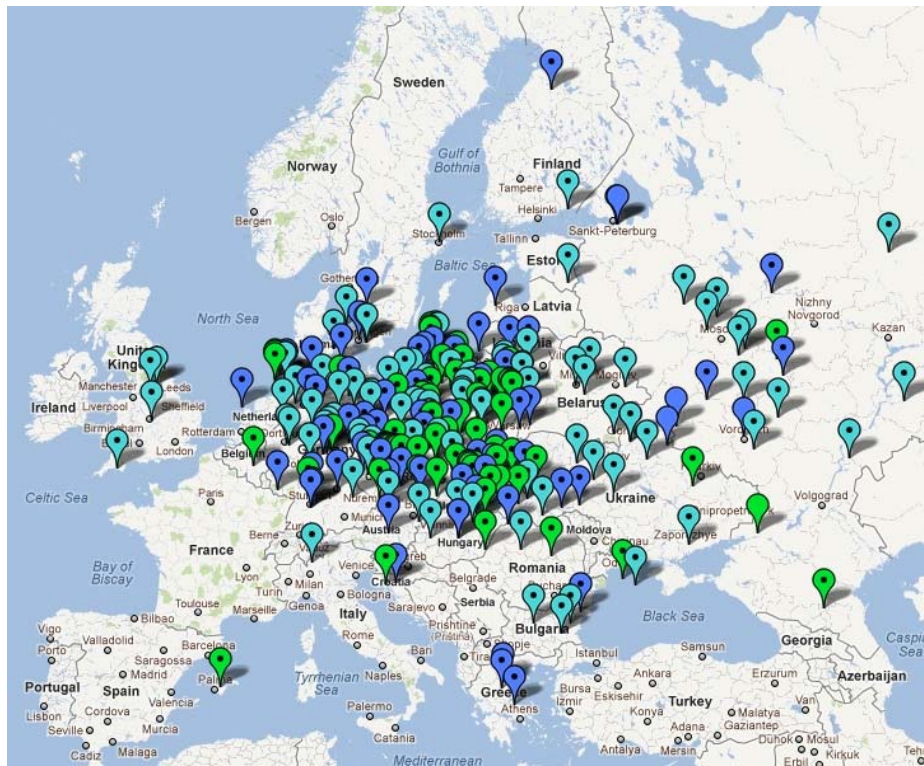
Additionally, the R1a nomenclature also includes the L175 subclade as downstream of L176.1, however, it is most probably a private SNP. The haplotype tree is shown in **Figure 8**.

The lower part of the tree is represented by the Old Scandinavian branch with its base haplotype as follows:

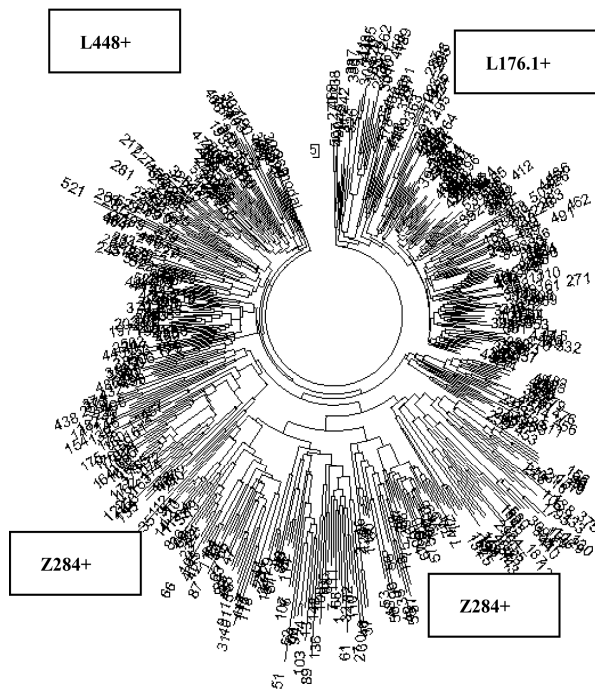
13 25 16 **10** 11 14 12 12 10 13 11 30—15 9 10 11 11 24 14 20 **31** 12 15 15 16—11 12 19 23 16 16 18 19 35 **39 12** 11—**12** 8 17 17 8 12 10 8 11 10 12 22 22 **16 11** 12 12 13 8 14 23 21 12 12 11 13 11 11 12 13

It differs by 8 mutations from the Z280 base haplotypes (1800 years), which places their common ancestor at  $\geq 5200$  years (see the explanation above regarding the “equal or higher” sign).

Since there are 17 identical (base) 12-marker haplotypes



**Map 8.** The map of the M458 subclade, depicting birthplaces of the furthest ancestors recorded in available databases. Blue pins denote the Central European branch (CE-1 deep-blue, CE-2 light-blue), green pins correspond to the West Slavic branch.



**Figure 8.** A 67-marker haplotype tree of the R1a-Z284 (Scandinavian) subclade, composed of 540 haplotypes. Three principal subclades/branches are shown: the Old Scandinavian (Z284), the Young Scandinavian (L448), and the Scottish Highland (L176.1).

among 178 OS haplotypes in the dataset, the logarithmic method gives  $[\ln(178/17)]/.12 = 117 \rightarrow 133$  generations, or  $\sim 3325$  years to the common ancestor; this is within the above  $3500 \pm 400$  ybp, as determined by the linear method. The good fit (within 5%) means that practically all 178 OS haplotypes actually descended from one common ancestor (Klyosov, 2009a).

The Young Scandinavian branch (L448+) has the following base haplotype:

13 25 **15** 11 11 14 12 12 10 **14** 11 31—15 9 10 11 11 **23** 14  
20 32 12 15 15 16—11 12 19 **21** 16 16 **17 18 34** 38 **12** 11—11  
8 17 17 8 12 10 8 11 10 12 22 22 15 **11** 12 12 13 8 14 23 21 12  
12 11 13 11 11 12 **12**

Since the YS base haplotype is younger than the OS base, it is more distant from the Z280 base haplotype on the same haplogroup tree; indeed, they deviate by 12 mutations (2775 years); this places their common ancestor at  $\geq 5300$  ybp, which is practically identical with the date obtained above with the OS base haplotype.

The same figure is reproduced with a rather loose set of haplotypes, positive at Z284 and negative at L448 and Z287/Z288 ( $4200 \pm 500$  ybp), with the base haplotype as follows:

13 25 **15 10** 11 14 12 12 10 13 11 30—15 9 10 11 11 24 14  
20 32 12 15 15 16—11 **11** 19 23 16 16 **17 18 35 39 12** 11—**12**  
8 17 17 8 12 10 8 11 10 12 22 22 15 **11** 12 12 13 8 14 23 21 12  
12 11 13 11 11 12 13

Its 10-mutation difference (2275 years) with the base Z280 results in  $\geq 5700$  years to their common ancestor. The comparison of the above base haplotype with those for OS and YS suggests that it is identical or nearly identical to the base haplotype of the entire Z284 subclade, with its estimated “age” of



~4300 years.

The haplotype of the Scottish Highland branch (L176.1) is

13 25 **16** 11 11 14 12 12 10 14 11 31—15 **8** 10 11 11 23 14  
20 **29** 12 15 15 **15**—**10** 12 19 21 16 16 17 **19** 34 38 12 11—11  
8 17 17 8 12 10 8 11 10 12 22 22 15 11 12 12 13 8 14 23 21 12  
12 11 13 11 11 12 12

Seven mutations (1550 years) with the YS base haplotype confirms their relative position of the subclade tree.

The known sub-branches of Z284 were found mostly in Scandinavian countries and on the Isles (**Map 9**). Norwegians and Swedish-speaking Finns seem to be the typical bearers of this genealogical DNA-lineage.

### Chronology of the Main M417+ Subclades

The combined data of SNP mutations and extended STR haplotypes allowed us to make a step-wise reconstruction of the base haplotypes of the major subclades of the R1a1 haplogroup. A consensus tree was calculated from these data (**Figure 9**). Its topology follows closely that of SNP phylogeny, while the dates of its principal nodes are quite robust, as it is shown by the examples above.

### Conclusion and Some Historical Considerations

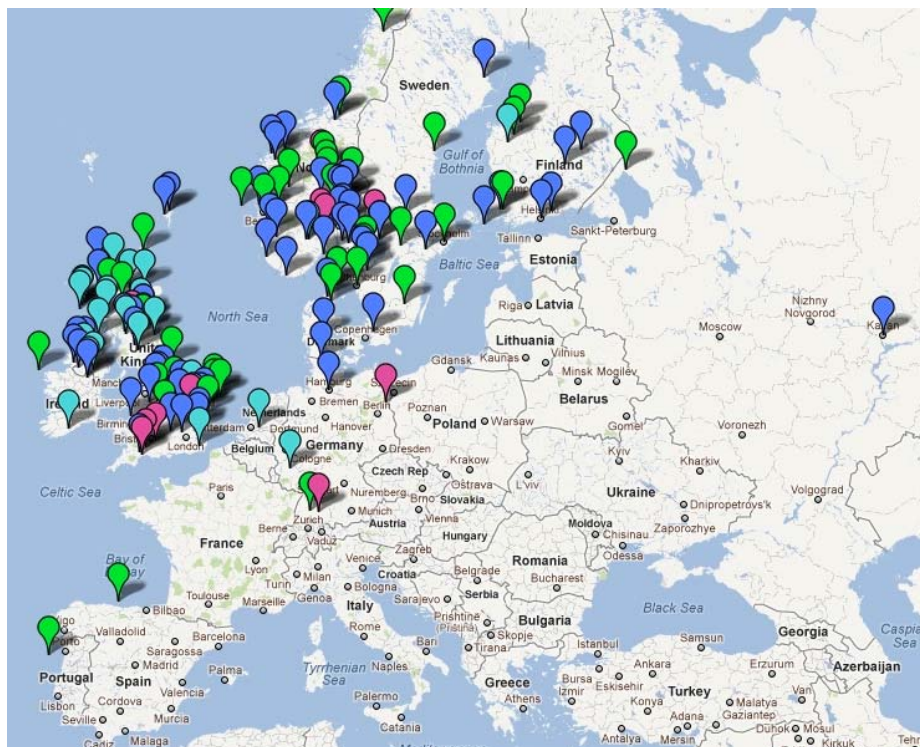
It follows from the data presented in this study, that all three principal downstream R1a subclades, L664, Z283 and Z93, split from their common European ancestor at about the same time, around 7000 - 6000 ybp, and that all three went through a population bottleneck, and then expanded 1000 - 1500 years

later. L664 apparently stayed in North-Western Europe; its lineage recovered and began expanding ~4575 ybp.

The Z280 subclade (descendants of Z283) moved east to the Russian Plain ~4800 - 4600 ybp when the Arbins, bearers of haplogroup R1b, were arriving in Europe from different directions—from North Africa to the Pyrenees (4800 ybp); then—as the Bell Beakers—to the continent; and finally from Asia Minor and the Middle East to the Apennines and the Balkans (4800 - 4500 ybp). The Z280 subclade and its Russian Plain branch formed at least 16 sub-branches from the middle of the 3<sup>rd</sup> millennium BC and later in the course of the westward repopulation of Europe in the 1<sup>st</sup> millennium BC-1<sup>st</sup> millennium CE.

The Z93 subclade began to expand during the Aryan migrations, on their way to India and the Middle East in the 3<sup>rd</sup>-2<sup>nd</sup> millennia BC. The eastward migration of the Aryans continued to the east (the Ural Mountains and beyond, to the Altay), south east (Iran and Hindustan), south (Anatolia, Mitanni and further south to the Middle East) between 4500 and 3500 ybp. The Jewish R1a common ancestor lived  $4525 \pm 400$  ybp; his descendants also almost vanished, but their long population bottleneck ended around the 8<sup>th</sup> century CE, in the Diaspora, apparently in Germany.

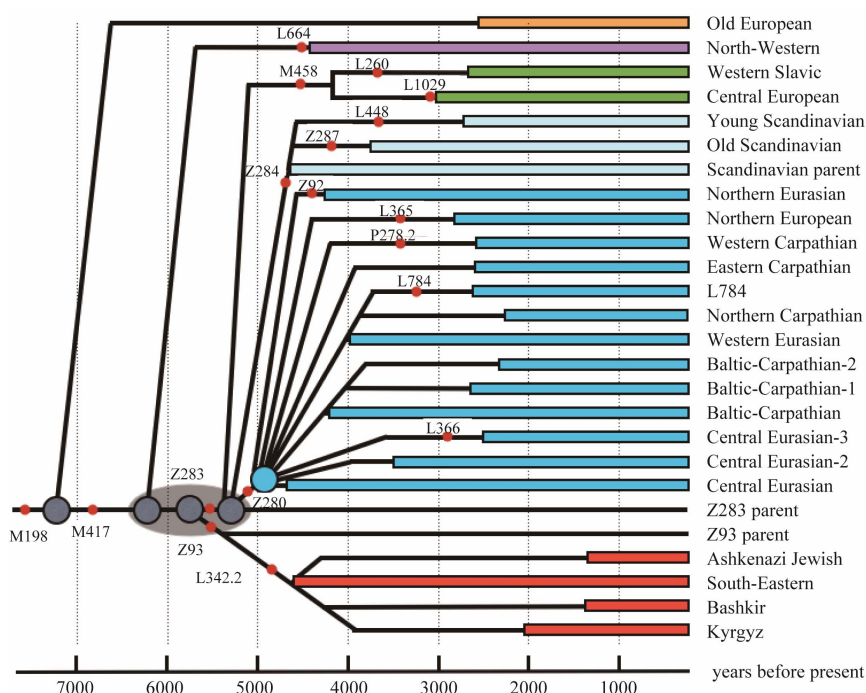
During the 1<sup>st</sup> millennium BC bearers of R1a started to repopulate Europe west of the Carpathian Mountains. This led to the appearance of the Northern European branch (L365, 2600 ybp), Western Carpathian (P278.2, 2600 ybp), Eastern Carpathian (2600 ybp, SNP is unknown), Northern Carpathian (2150 ybp, SNP is unknown), two Balto-Carpathian branches (2550



**Map 9.**

The map of the Scandinavian Z284 subclade, depicting birthplaces of the furthest ancestors recorded in available databases. Magenta pins denote the parent branch, green pins mark the Old Scandinavian one (Z287/Z288), deep-blue and light-blue pins correspond to the young Scandinavian branch (L448) and its Scottish Highland sub-branch (L176.1), respectively.





**Figure 9.**

The R1a1 haplogroup tree according to the data on extended haplotypes from available databases. The lengths of colored bars correspond to TMRCA of present-day bearers of particular branches. Big circles denote the principal nodes. Red circles mark the assignments by SNP, wherever possible. Designation of branches follows the same notation as in the text.

and 2200 ybp, SNPs are unknown), two Northern Eurasian branches of Z92 subclade (3600 and 2350 ybp; the Z92 subclade itself arose 4450 ybp, apparently on the Russian Plain), Western Eurasian-1 (2300 ybp). The same sequence of population events can be traced in two subclades which are downstream from Z283: M458 (European) and Z284 (Scandinavian). The former arose 4200 ybp in a yet unknown region, and expanded as two Central European branches of subclade M458 (both arose 2900 ybp) and the Western Slavic branch (L260, 2700 ybp). Z284 arose probably in 4300 ybp, somewhere on the Scandinavian peninsula, and produced three principal Scandinavian subclades—Old Scandinavian (Z287, 3700 ybp), Young Scandinavian (L448, 2700 ybp), and Scottish Highland (L176.1, 1850 ybp).

Some of the older branches, e.g. the Russian Plain branch, stayed in the present Russia-Ukraine-Belarus-Poland-Baltic region; ancient historians described these peoples as Scythians, Antes, Veneti, and various other proto-Slavic tribes (though many of them belonged to other haplogroups, primarily I1 and I2). R1a branches which are “older” than 3000 years, such as the Russian Plain branch (4600 ybp), the Western Eurasian branch (4300 ybp), the Balto-Carpathian (4300 ybp), did not move *en mass* to Europe and stayed behind at the Russian Plain.

In the middle of the 1<sup>st</sup> millennium CE, at the times of the collapse of the Roman Empire, many populations, including R1a, were migrating westward and eastward; these migrations gradually formed the current landscape of R1a in Europe. All 38 branches of R1a and their datings are listed in the Appendix of this paper; current distribution maps are shown in the body of this paper.

Presently, the largest R1a populations live in Russia (up to 63% of the total population in the South of historical Russia), in Poland, Ukraine and Belarus (up to 57% of the population). Outside of Eastern and Central Europe, R1a accounts for to 15%-20% of the population of India (some estimates place it up to 30% in North India), up to 9% of the population in the Middle East belongs to R1a (Aby-Amern et al., 2009); in the rest of the world, the presence of R1a is minor (Klyosov & Rozhanskii, 2012, and references therein).

## Materials and Methods

Two thousand four hundred and seventy-one (2471) of 67- and 111-marker R1a1 haplotypes were considered in this study. Together, they constitute the IRAKAZ R1a database (Rozhanskii et al., 2012), which was collected from FTDNA, YSearch, and private lists, kindly provided for public usage by Martin Voorwinden, Andrew MacEachern and others, to whom we are deeply indebted.

The methodology of haplotype dataset analysis was described in Rozhanskii and Klyosov (2011). In this study the linear and the logarithmic method were used, the latter when mutation counting was not required, as described in Klyosov (2009a). The mutation rate constants are listed in Klyosov (2009a); Rozhanskii and Klyosov (2011), and for a number of cases are given in the text of this paper.

The concept of a “conditional generation” is an important part of DNA genealogy, and it is not related to a common “generation”. This is because a generation length is a “floating” figure which depends on a society, culture, traditions, and particular circumstances (wars, famine, cataclysms, epidemics,

etc.). It means that conversions of “generations” into years should be done by calibrating using known historical events. The calibration was performed on many examples, as described in the preceding paper in *Advances in Anthropology* (Rozhanskii & Klyosov, 2011).

As it follows from basic rules of chemical kinetics, the experimentally determined ratio  $n/N$  in the equation

$$n/N = kt$$

ultimately determines a product  $kt$ .

In this equation:

$n$  = a number of mutations (from the base haplotype) in a series of  $N$  haplotypes,

$N$  = a number of haplotypes, descending from one common ancestor (having the base haplotype) in the series,

$k$  = the average mutation rate constant for the haplotypes (per generation) of the given format,

$t$  = a number of generations (TMRCA).

Therefore, if we set the length of the generation, for example as 25 years (making it “the conditional generation”, which is a mathematical value rather than a floating “generation”), then the value of  $k$  can be calibrated with respect to a certain timespan during which  $n$  mutations occurred per  $N$  haplotypes. For example, in 84 of 67 marker haplotypes of the Clan Donald (the common ancestor of which, John, Lord of the Isles, died in 1386, with the TMRCA equaling to 650 years before present, which is 26 “conditional generations” of 25 years each), there were 262 mutations. This gives  $262/84 = 26 k$ , and hence  $k = .12$  mutations per the conditional generation (25 years) per the 67 marker haplotype.

As it is shown in (Rozhanskii & Klyosov, 2011), the obtained mutation rate constant of .12 is rather accurately reproducible for many series of 67 marker haplotypes from many haplogroups from many tested populations around the world. In fact, it is reproducible for ANY tested population in the world for which 67 marker haplotypes are available (Rozhanskii & Klyosov, 2011). This “calibration” connects a number of mutations in a series of haplotypes (in a given format and length, from the 6 marker haplotypes to the 111 marker haplotypes) to a known number of years to a known historical event or a known genealogy, setting a length of generation as 25 years, and verifying the mutation rate constant related to this “conditional generation”.

It should be noted that we could have set the timespan to the common ancestor of the Clan Donald as 13 generations of 50 years each and it would not have changed a thing. There would still be 650 years to the common ancestor. The mutation rate constant would simply be twice as high.

The mutation rate constant for the 111-marker haplotypes, also employed in this study, equals to .198 mutation per haplotype per conditional generation (Rozhanskii & Klyosov, 2011).

The above equation is essentially the basis for the “linear” method in DNA genealogy. It is based on mutation counting in haplotypes in the dataset compared with the “base” haplotype, which is (or the best available approximation to) the ancestral haplotype for the dataset. The “base” haplotype is the ancestral haplotype in an ideal case. However, since those base haplotypes often are deduced ones, it would be inaccurate to call them “ancestral” when that might not necessarily be true. Hence, we call them “base” haplotypes. Base haplotypes in the dataset were determined by minimization of mutations; by definition, the base haplotype is one which has the minimum col-

lective number of mutations in the dataset having one common ancestor.

Another principal method in DNA genealogy is the logarithmic method, which does not involve mutation counting in haplotypes. It is based on the assumption (which largely or practically always holds true) that a transition of the base haplotypes into mutated ones (or, rather, belonging to descendants inherited mutated haplotypes) is described by the first-order kinetics:

$$N = Ae^{kt}$$

that is

$$\ln(N/A) = kt$$

where:

$N$  = a total number of haplotypes in a set,

$A$  = a number of unchanged (identical, not mutated) base haplotypes in the set,

$k$  = an average mutation rate,

$t$  = a number of generations to a common ancestor.

One can see the logarithmic method does not consider a number of mutations in haplotypes; only mutated and non-mutated (base) haplotypes are considered.

The above equation is the fundamental one in chemical kinetics. Radioactive decay, for example, is described by the same equation, though it does not involve an analogy with “back mutations”, since it is truly irreversible. Back mutations (or reverse mutations), though, are quite common in haplotypes, and their analysis and corrections for back mutations are given in (Klyosov, 2009a; Rozhanskii & Klyosov, 2011).

Haplotype trees were composed using PHYLIP, the Phylogeny Inference Package program (Felsenstein, J., 2004; for multiple examples of usage and calculations see Klyosov, 2009a, 2009b and references therein). Corrections for back mutations were introduced as described in Klyosov (2009a); Rozhanskii and Klyosov (2011). Margins of error were calculated as described in Klyosov (2009a).

## Acknowledgements

The authors are indebted to Dr. Judith Remy Leder for her valuable help with the preparation of the manuscript.

## REFERENCES

- Abu-Amero, K. K., Hellani, A., Gonzalez, A. M., Larruga, J. M., Cabrera, V. M., & Underhill, P. A. (2009). Saudi Arabian Y-chromosome diversity and its relationship with nearby regions. *BMC Genetics*, 10, 59. doi:10.1186/1471-2156-10-59
- Balanovsky, O., Rootsi, S., Pshenichnov, A., Kivisild, T., Churmosov, M., Evseeva, I., Pocheshkhova, E. et al. (2008). Two sources of the Russian patrilineal heritage in their Eurasian context. *American Journal of Human Genetics*, 82, 236-250. doi:10.1016/j.ajhg.2007.09.019
- Behar, D. M., Thomas, M. G., Skorecki, K., Hammer, M. F., Buluygina, E., Rosengarten, D., Jones, A. L. et al. (2003). Multiple origins of Ashkenazi Levites: Y chromosome evidence for both Near Eastern and European Ancestries. *American Journal of Human Genetics*, 73, 768-779. doi:10.1086/378506
- Childe, G. (1926). *The Aryans: A study of Indo-European origins*. Dorchester: Dorset Press.
- Cinnioglu, C., King, R., Kivisild, T., Kalfoglu, E., Atasoy, S., Cavalleri, G. L., Lillie, A. S. et al. (2004). Excavating Y-chromosome haplotype strata in Anatolia. *Human Genetics*, 114, 127-148. doi:10.1007/s00439-003-1031-4
- Felsenstein, J. (2004). PHYLIP (Phylogeny Inference Package). Ver-

- sion 3.6. Seattle: Department of Genome Sciences, University of Washington.
- Gamkrelidze, T. V., & Ivanov, V. V. (1995). *Trends in linguistics 80: Indo-European and the Indo-Europeans*. Berlin: Mouton de Gruyter.
- Gray, R. D., & Atkinson, Q. D. (2003). Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature*, *426*, 435-439. doi:10.1038/nature02029
- Haak, W., Brandt, G., de Jong, H. N., Meyer, C., Ganslmeier, R., Heyd, V., Hawkesworth, C. et al. (2008). 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 of the United States of America*, *105*, 18226-18231. doi:10.1073/pnas.0807592105
- Kayser, M., Lao, O., Anslinger, K., Augustin, C., Barger, G., Edlmann, J., Elias, S. et al. (2005). Significant genetic differentiation between Poland and Germany follows present-day political borders, as revealed by Y-chromosome analysis. *Human Genetics*, *117*, 428-443. doi:10.1007/s00439-005-1333-9
- Keyser, C., Bouakaze, C., Crubezy, E., Nikolaev, V. G., Montagnon, D., Reis, T., & Ludes, B. (2009). Ancient DNA provides new insight into the history of south Siberian Kurgan people. *Human Genetics*, *126*, 395-410. doi:10.1007/s00439-009-0683-0
- Kharkov, V. N., Stepanov, V. A., Borinskaya, S. A., Kozhekbaeva, Zh. M., Gusar, V. A., Grechanina, E. Ya., Puzyrev, V. P. et al. (2004). Gene pool structure of Eastern Ukrainians as inferred from the Y-chromosome haplogroups. *Russian Journal of Genetics*, *40*, 326-331. doi:10.1023/B:RUGE.0000021635.80528.2f
- Kharkov, V. N., Stepanov, V. A., Feshchenko, S. P., Borinskaya, S. A., Yankovsky, N. K., & Puzyrev, V. P. (2005). Frequencies of Y-chromosome binary haplogroups in Belarusians. *Russian Journal of Genetics*, *41*, 928-931. doi:10.1007/s11177-005-0182-x
- Klyosov, A. A. (2009a). DNA Genealogy, mutation rates, and some historical evidences written in Y-chromosome. I. Basic principles and the method. *Journal of Genetic Genealogy*, *5*, 186-216.
- Klyosov, A. A. (2009b). DNA Genealogy, mutation rates, and some historical evidences written in Y-chromosome. II. Walking the map. *Journal of Genetic Genealogy*, *5*, 217-256.
- Klyosov, A. A. (2010a). DNA lineages and origin of the "Tenths", the North-Western European subfamily R1a1 with YS388 = 10. *Proceedings of the Russian Academy of DNA Genealogy*, *2*, 983-1028.
- Klyosov, A. A. (2010b). DNA lineages and origin of the Tenths, the R1a1 subfamily with DYS388 = 10: The story in progress. *Proceedings of the Russian Academy of DNA Genealogy*, *3*, 2205-2217.
- Klyosov, A. A. (2011). Biological chemistry as a foundation of DNA genealogy: The emergence of "molecular history". *Biokhimiya (English translation)*, *76*, 517-533.
- Klyosov, A. A. (2012). An emerging pattern of L664 SNP mutation in the Tenths (bearers of DYS388=10), North-Western branch of R1a1 haplotypes. *Proceedings of the Russian Academy of DNA Genealogy*, *5*, 30-37.
- Klyosov, A. A., & Rozhanskii, I. L. (2012). Haplogroup R1a as the Proto Indo-Europeans and the legendary Aryans as witnessed by the DNA of their current descendants. *Advances in Anthropology*, *2*, 1-13. doi:10.4236/aa.2012.21001
- Li, C., Li, H., Cui, Y., Xie, C., Cai, D., Li, W., Mair, V. H. et al. (2010). Evidence that a West-East admixed population lived in the Tarim Basin as early as the early Bronze Age. *BMC Biology*, *8*, 15. doi:10.1186/1741-7007-8-15
- Renfrew, C. (2000). 10,000 or 5000 year ago? Questions on time depth. In Renfrew, C., McMahon, & A., Trask, L. (Eds.), *Time depth in historical linguistics* (pp. 413-439). Cambridge: The McDonald Institute for Archaeological Research.
- Rozhanskii, I. L., & Klyosov, A. A. (2011). Mutation rate constants in DNA genealogy (Y chromosome). *Advances in Anthropology*, *1*, 26-34. doi:10.4236/aa.2011.12005
- Rozhanskii, I., Klyosov, A., & Zolotarev, A. (2012). IRAKAZ R1a database. *Proceedings of the Academy of DNA Genealogy*, *5*, 553-559.
- Sengupta, S., Zhivotovsky, L. A., King, R., Mehdi, S. Q., Edmonds, C. A., Chow, C. E. T., Lin, A. A. et al. (2006). Polarity and temporality of high-resolution Y-chromosome distributions in India identify both indigenous and exogenous expansions and reveal minor genetic influence of central Asian pastoralists. *Human Genetics*, *78*, 202-221. doi:10.1086/499411
- Sharma, S., Rai, E., Sharma, P., Jena, M., Singh, S., Darvishi, K., Bhat, A. K. et al. (2009). The Indian origin of paternal haplogroup R1a1\* substantiates the autochthonous origin of Brahmins and the caste system. *Journal of Human Genetics*, *54*, 47-55. doi:10.1038/jhg.2008.2
- Underhill, P. A., Myres, N. M., Rootsi, S., Metspalu, M., Zhivotovsky, M. A., King, R. J. et al. (2009). Separating the post-Glacial coancestry of European and Asian Y chromosomes within haplogroup R1a. *European Journal of Human Genetics*, *18*, 479-484. doi:10.1038/ejhg.2009.194
- Wells, R. S., Yuldasheva, N., Ruzibakiev, R., Underhill, P. A., Evseeva, I., Blue-Smith, J., Jin, L. et al. (2001). The Eurasian Heartland: A continental perspective on Y-chromosome diversity. *Proceedings of the National Academy of Sciences of the United States of America*, *98*, 10244-10249. doi:10.1073/pnas.171305098



**Appendix 1**

A list of R1a branches in Europe (dates in ybp).

1	R1a1-M198- (or SRY10831.2*)	≥9000
2	R1a1a-M198+ M417- (or M17*)	≥9000
3	The Tenth, NW branch,	~4575
4	The Tenth, L664+	~4575
5	The Tenth, DYS448 null mutation	1500 ± 325 and 260 ± 60 (USA)
6	Z93+ started to expand	~5700, ~4050
7	Z93-L342.2+	4900 ± 500
8	Z93-L342.2-L657+	4050 ± 460
9	Z93-L342.2, Jewish branch started to expand	4475 ± 400, 1300 ± 150
10	Z93-L342.2, Bashkir branch	1125 ± 200
11	Z93-L342.2, Kyrgyz branch	2100 ± 250
12	Z283 (Eurasian branch)	~5500
13	M458 (European branch)	~4200
14	M458 CE parent	~3100
15	M458 CE-1	2900 ± 400
16	M458 CE-2	2900 ± 400
17	M458-L260 WS	2700 ± 300
18	Z284 (Scandinavian branch)	4300 ± 500
19	Z284-Z287, Old Scandinavian	3700 ± 400
20	Z284-L448, Young Scandinavian	2700 ± 350
21	Z284-L448-L176.1, Scottish Highland	1850 ± 290
22	Z280, Central Eurasian (CEA)	4900 ± 500
23	Z280, CEA-1, Russian Plain	4600 ± 500
24	Z280, CEA-2	3500 ± 400
25	Z280-L366, CEA-3	2500 ± 350
26	Z280-L365, North European (NE)	2600 ± 350
27	Z280-P278.2, Western Carpathian (WC)	2600 ± 300
28	Z280-Z92, North Eurasian (NEA)	4450 ± 450
29	Z280, NEA-1	3600 ± 400
30	Z280, NEA-2	2350 ± 300
31	Z280, Western Eurasian (WEA)	4100 ± 500
32	Z280, WEA-1	2300 ± 300
33	Z280, Eastern Carpathian (EC)	2600 ± 300
34	Z280, Balto-Carpathian (BC)	4300 ± 500
35	Z280, BC-1	2550 ± 300
36	Z280, BC-2	2200 ± 250
37	Z280, Northern Carpathian (NC)	2150 ± 300
38	Z280, L784	2500 ± 450

**Appendix 2**

The following DNA projects were selected as primary haplotype databases:

<http://www.familytreedna.com/public/R1aY-Haploroup/default.aspx?vgroup=R1aY-Haplogroup&section=yresults>  
<http://www.familytreedna.com/public/R1a/default.aspx?section=yresults>  
<http://www.familytreedna.com/public/r-arabia/default.aspx?section=yresults>  
[http://www.familytreedna.com/public/Hungarian\\_Magyar\\_Y-DNA\\_Project/default.aspx?section=yresults](http://www.familytreedna.com/public/Hungarian_Magyar_Y-DNA_Project/default.aspx?section=yresults)  
<http://www.familytreedna.com/public/India/default.aspx?section=yresults>  
<http://www.familytreedna.com/public/Turkic/default.aspx?section=yresults>  
<http://www.familytreedna.com/publicwebsite.aspx?vgroup=germany&section=yresults>  
<http://www.familytreedna.com/public/Finland/default.aspx?section=yresults>  
<http://www.familytreedna.com/public/Czech/default.aspx?section=yresults>  
<http://www.familytreedna.com/public/slovakia/default.aspx?section=yresults>  
<http://www.familytreedna.com/public/polish/default.aspx?section=yresults>  
[http://www.familytreedna.com/public/russian\\_impire.aspx/default.aspx?section=yresults](http://www.familytreedna.com/public/russian_impire.aspx/default.aspx?section=yresults)  
<http://www.familytreedna.com/public/russiadna/default.aspx?section=yresults>

Reference data were selected according to SNP assignment from YSearch database: (<http://www.ysearch.org>) and public projects of FTDNA (<http://www.familytreedna.com>).