

Haplogroup R1a as the Proto Indo-Europeans and the Legendary Aryans as Witnessed by the DNA of Their Current Descendants

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

Received November 15th, 2011; revised December 20th, 2011; accepted January 10th, 2012

This article aims at reconstructing history of R1a1 ancient migrations between 20,000 and 3500 years before present (ybp). Four thousand four hundred sixty (4460) haplotypes of haplogroup R1a1 were considered in terms of base (ancestral) haplotypes of R1a1 populations and timespans to their common ancestors in the regions from South Siberia and northern/northwestern China in the east to the Hindustan and further west across Iranian Plateau, Anatolia, Asia Minor and to the Balkans in Europe, including on this way Central Asia, South India, Nepal, Oman, the Middle East, Comoros Islands, Egypt, etc. This study provides a support to the theory that haplogroup R1a arose in Central Asia, apparently in South Siberia and/or neighboring regions, around 20,000 ybp. Not later than 12,000 ybp bearers of R1a1 already were in the Hindustan, then went across Anatolia and the rest of Asia Minor apparently between 10,000 and 9000 ybp, and around 9000 - 8000 ybp they arrived to the Balkans and spread over Europe east to the British Isles. On this migration way or before it bearers of R1a1 (or the parent, upstream haplogroups) have developed Proto Indo-European language, and carried it along during their journey to Europe. The earliest signs of the language on passing of bearers of R1a1 through Anatolia were picked by the linguists, and dated by 9400 - 9600 - 10,100 ybp, which fairly coincides with the data of DNA genealogy, described in this work. At the same time as bearers of the brother haplogroup R1b1a2 began to populate Europe after 4800 ybp, haplogroup R1a1 moved to the Russian Plain around 4800 - 4600 ybp. From there R1a1 migrated (or moved as military expeditions) to the south (Anatolia, Mitanni and the Arabian Peninsula), east (South Ural and then North India), and south-east (the Iranian Plateau) as the historic legendary Aryans. Haplotypes of their direct descendants are strikingly similar up to 67 markers with contemporary ethnic Russians of haplogroup R1a1. Dates of those Aryan movements from the Russian Plain in said directions are also strikingly similar, between 4200 and 3600 ybp.

Keywords: Y Chromosome; Mutations; Haplotypes; Haplogroups; TMRCA; STR; SNP; Indo-European; India; Aryans; R1a1

Introduction

This study focuses on the origin of Indo-Europeans and the Aryans who entered India (the Hindustan), Iran (Iranian plateau), and Anatolia (Mesopotamia) approximately 3500 years ago.

The research findings, described in this study, demystify the origin of the Aryans. For nearly two centuries, the “Aryan problem” (essentially—Who were the Aryans? Where did they come from? Where did they disappear? Were they a particular human race, different from others?) has posed many challenges, often controversial and conflicted, for researchers, archeologists and linguists; however, this study opens new ground for our consideration and is based on the data provided by DNA genealogical test results.

The methodology of DNA genealogy, including considerations of extended 67 marker haplotypes, is described in detail in the preceding paper in this journal (Rozhanskii & Klyosov, 2011) and in Materials and Methods section of this article. The 67 marker haplotypes have been introduced to the scientific domain and personal usage several years ago, and available databases containing tens of thousands of 67 marker haplotypes are listed in (Rozhanskii & Klyosov, 2011) and in this paper (Appendix).

First, the following two 67 marker haplotypes of haplogroup R1a1 are presented, belonging to the two authors of this paper:

13 24 16 11 11 15 12 10 13 11 30—16 9 10 11 11 24 14
20 34 15 15 16 16—11 11 19 23 15 16 17 21 36 41 12 11—11
9 17 17 8 11 10 8 10 10 12 22 22 15 10 12 12 13 8 15 23 21 12
13 11 13 11 11 12 13

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

Next, the following are two Indian R1a1 haplotypes, taken arbitrarily from the “Indian FTDNA Project” (the references to Projects are at the end of this paper):

13 24 16 11 11 14 12 12 10 13 11 31—16 9 10 11 11 24 14
20 33 12 15 15 16—10 12 19 23 15 17 18 18 35 41 15 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 10 11 12 12

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

Both Indian haplotypes contain 24 and 21 mutations with the

first haplotype (mutations are shown in bold, and rules of their counting are explained in the paper cited above), and 28 and 36 mutations with the second one. This produces on average 27.25 ± 6.60 pairwise cross-mutations between all four haplotypes; that is, $27.25/.12 = 227 \rightarrow 292 \pm 55$ conditional generations (25 years each) = 7300 ± 1400 years between two haplotypes on average, or 3650 ± 700 years to a common ancestor of all the four haplotypes (.12 here is the mutation rate constant for 67 marker haplotypes, see the preceding paper cited above). According to all historical accounts, the Aryans arrived in India in the middle of the 2nd millennium BC, which is approximately 3500 years ago.

This simplified calculation is based on these four haplotypes that belong to different subclades of R1a1 haplogroup (Z280, M458 and L342.2). However, considering each of the four haplotypes, the first two are from the current Russian-Ukrainian (Indo-European) group, and the second two are from the Indian (Indo-European) group. Both are similar and belong to the same R1a1 haplogroup. Currently, up to 72% of the upper castes in India belong to bearers of the same R1a1 haplogroup (Sharma et al., 2009).

This simplified calculation is given here for illustrative purposes, though four 67 marker haplotypes contain as many as 268 markers, which is quite statistically informative in a first approximation. A much more detailed analysis of Indian and ethnic Russian extended series of R1a1 haplotypes is given in (Klyosov, 2009b, 2011b), and principally the same results were obtained with respect to patterns of mutations in haplotypes, migration routes, and their chronology.

This brings closure to the question of the Aryans' DNA-related origin and who entered India during the middle of the 2nd millennium CE. They belonged to the R1a1 haplogroup, which is the prevalent one in the present-day Eastern Europe (Russia, Poland, Ukraine, Belarus, in the first one up to 62% of total male population, in the latter three up to 55% of total male population [Klyosov, 2009b, 2011b and references therein]).

There is merit in comparing the Indian haplotypes with the R1b1a2 haplotypes—a group who populates ~60% of Europe, living primarily in the British Isles, Spain, France, Belgium, Germany, the Netherlands, and other Central and Western European countries. The typical ancestral haplotype in R1b1a2 haplogroup, dated about 4,800 years before present, is as follows (Klyosov, 2011a):

13 24 14 11 11 14 12 12 13 13 29—17 9 10 11 11 25 15
19 29 15 15 17 17—11 11 19 23 15 15 18 17 36 38 12 12—11
9 15 16 8 10 10 8 10 10 12 23 23 16 10 12 12 15 8 12 22 20 13
12 11 13 11 11 12 12

There are 48 and 44 mutations between the above and the Indian R1a1 haplotypes shown earlier. This formally places their common ancestor at more than 10,000 years before present and, in fact, much earlier, at least 15,000 years ago. R1b1a2 bearers were not among the Aryans coming to India, and it is very likely that they were not Indo-Europeans then. Specifically, there is no supporting evidence that 4000 years before present (ybp) bearers of R1b1a2 spoke Indo-European (IE) languages. On the other hand, Central Europe was likely populated by R1b1a2 speakers of non-IE languages. Moreover, there are very few bearers of R1b haplogroup in India, mostly on its Arabian Sea coast, and there were none of the R1b haplogroup among the 367 tested Indian Brahmins (Sharma et al., 2009). Therefore, it is highly unlikely that bearers of the R1b1 (as well as R1b1a2)

haplogroup were among the Aryans, and, hence, they were not among those carrying the Indo-European languages elsewhere in those times.

We are left holding two questions: first, from where did the R1a haplogroup arise and, second, what was their migratory route that brought them to 1) the Russian Plain (currently up to 62% R1a1, see above), 2) India and Iran (10% - 16% R1a1), 3) Anatolia (15% R1a1), 4) the Middle East (up to 7% - 13% R1a1), and (5) the Arabian Peninsula, where nowadays 2% - 10% of the population carries the R1a1 haplogroup (Abu-Amro, 2009; Underhill et al., 2009)?

As described in (Klyosov & Rozhanskii, 2011), Europeoids (Caucasoids) appeared ~58,000 ybp. They gradually branched to downstream haplogroups and migrated to the west, south and east. Haplogroup NOP, which was among them, arose ~48,000 ybp, and moved eastward, presumably towards South Siberia and/or adjacent regions. Haplogroup P arose ~38,000 ybp, apparently in South Siberia, and gave rise to haplogroup R and then R1 ~30,000 - 26,000 ybp (see the diagram in Klyosov & Rozhanskii, 2011). The timing of haplogroup R1a's appearance can be reconstructed from series of R1a haplotypes, made available from the databases (see the list in Materials and Methods and the Appendix). The most ancient common ancestors of this haplogroup lived in: northern and northwestern China (in particular, Xinjiang region, which is the south Altai area), in southern Siberia, in the Eastern Himalayas, India and Pakistan, the Comoros Islands, and in Europe, where their bearers apparently migrated from the east during both the remote past and later, for example, with the Scythians.

Northern China R1a Haplotypes

Apparently, the most ancient source of R1a1 haplotypes is provided by the people now living in northern China. It was shown (Bittles et al., 2007) that for a number of Chinese populations, such as Hui, Bonan, Dongxiang, Salars, a percentage of R1a1 haplotypes reached 18% - 32%. Their haplotypes were not provided in the paper, but the author, Professor Alan H. Bittles, kindly sent us a list of 31 of five-marker haplotypes typed as R1a1, the tree of which is shown in **Figure 1**. The haplotypes vary tremendously in their alleles, which already indicates that their common ancestor lived in ancient times. For example, values of DYS19 varied between 14 and 17, DYS388 between 12 and 14, and DYS393 between 10 and 13. It should be noted that mutations in the last two markers occurred on average once in 4,545 and 1,320 generations, respectively. With a correction for back mutations (Klyosov, 2009a) it occurs once in 8500 and 2400 generations. The 31 haplotypes contain 99 mutations from the deduced 5 marker base haplotype as shown here in the 12 marker FTDNA format with missing alleles indicated:

13 X 14 X X X X 12 X 13 X 30

This extent of mutations, which can be presented as $99/31/5 = .639 \pm .064$ mutation/marker, is a very high value. Actually, it is a measure of how ancient a common ancestor might be (for a comparison, contemporary European R1a1 and R1b1a2 haplotypes are separated by .250 - .270 mutations from their common ancestors, see Klyosov 2011a, 2011b). It can also be presented as $99/31/.00677 = 472 \rightarrow 683$ conditional generations; that is, it is $17,100 \pm 2,400$ years to the common ancestor of these 31 haplotypes (explanations and examples of calculations are given in Materials and Methods). The value of .00677 mutation/haplotype for conditional generation is the mutation rate

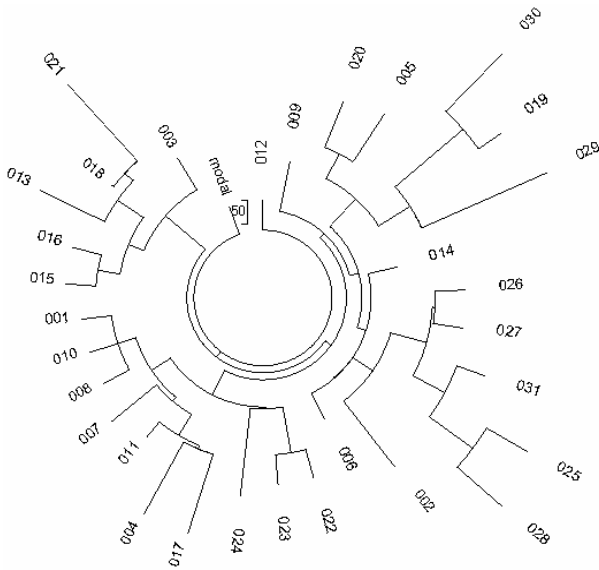


Figure 1. The 5-marker haplotype tree for R1a1 haplotypes in Northern China. The 31-haplotype tree was composed from data provided by Dr. A.H. Bittles and collected in ethnic communities Hui, Bonan, Dongxiang, and Salars (Bittles et al., 2007) (no haplotypes were provided in the referenced article). A fraction of R1a1 haplogroup in said populations is 18%, 25%, 32%, and 22%, respectively (ibid).

constant for the 5 marker haplotypes (Klyosov, 2009a). Since these haplotypes descend from such an ancient common ancestor and contain numerous mutations, this makes their deduced base (ancestral) haplotype rather uncertain. Therefore, the quadratic permutation method was employed for the same set of haplotypes (Klyosov, 2009a). This method does not require either a base haplotype or a correction for back mutations. The obtained timespan is $19,625 \pm 2,800$ years to a common ancestor (see Materials and Methods for calculations). This result is within the margin of error with that calculated by the above linear method.

Therefore, haplogroup R1a arose at approximately 20,000 ybp with the territory geographically belonging to Central Asia.

R1a1 Haplotypes from Altay

Thirteen Altay R1a1 haplotypes were listed in (Underhill et al., 2009), 12 of which showed a rather recent base haplotype (the last marker is DYS461):

13 26 16 11 X X X 12 11 14 11 31—10

These 12 haplotypes have only 7 mutations per 120 markers from the above base haplotype, which gives $7/120/0.0018 = 32 \rightarrow 33$ generations; that is, 825 ± 320 years to a common ancestor. The same set of the Altayan haplotypes in a different format is given in (Järve et al., 2009), with the base haplotype (not listed in the cited paper)

13 26 16 11 11 17 X X 11 14 11 31

and the same 7 mutations per 120 markers. Therefore, this exactly correlates to the same timespan to that of the common ancestor as given above. The same set of the Altayan haplotypes was given in (Järve et al., 2009) in a significantly more extended format, with the base haplotype (the second panel represents DYS 458, 437, 448, GATAH4, 456, 438, 594, 411S1

[two alleles], 596, 643, 645, 635, YPenta1, YPenta2):

13 26 16 11 11 17 X X 11 14 11 31—15 14 19 11 15 11 8 10 11 9 10 8 23 11 10

All 12 haplotypes in the extended format collectively contain the same 7 mutations. In other words, the added 15 “slow” markers did not produce mutations, and is an indicator of a quite recent common ancestor of the haplotype dataset. However, this base haplotype differs by 6 mutations from the base haplotype of the Russian Plain, which in the same format is

13 25 16 11 11 14 12 12 10 13 11 30—10

and by 12 mutations from the Russian Plain base haplotype in the extended format:

13 25 16 11 11 14 12 12 10 13 11 30— 15 14 20 12 16 11 10 10 11 10 10 8 23 11 10

(a more extended 67 marker base haplotype of the Russian Plain is shown below). These 6 and 12 mutations exactly fit the difference between the respective mutation rate constants for the two haplotype formats, equal to .020 and .0404 mutations per haplotype per generation, respectively (see Material and Methods). These mutation differences place a common ancestor of the Altayan and the Russian Plain haplotypes at 8100 ybp.

An additional Altayan haplotype from the list

14 24 17 11 11 15 X 12 12 10 13 11 31—15 14 20 13 15 11 10 10 12 10 10 9 24 11 8

differs by as much as 20 mutations from the above base Altayan haplotype, and by 12 mutations with the Russian Plain base haplotype. This places their common ancestors at 10,400 and 7300 ybp, respectively.

R1a1 Haplotypes in Tuva

Four R1a1 haplotypes from Tuva, the region which borders with Altay in Southern Siberia, north of Mongolia, were listed in (Underhill et al., 2009; Järve et al., 2009). Three of them are identical in all their 26 markers

13 26 15 9 11 14 X 12 10 13 11 30—18 14 19 12 16 11 10 10 11 10 10 8 23 11 8

and significantly differ, by as many as 20 mutations, from the fourth one, which belongs to the lineage found among the Altayans:

13 25 16 11 11 17 X 12 11 14 11 31—15 14 19 11 15 11 8 10 9 10 10 8 23 11 10

They also differ by 12 mutations from the Russian Plain base haplotype. This places a common ancestor of the Tuva R1a1 haplotypes at 10,000 ybp and also with the Russian Plain base haplotype by 7300 ybp.

What emerges from the analysis of the data is that the Altayan and the Tuva haplotypes have apparently the same ancient R1a1 common ancestor, who lived 10,000 - 10,400 ybp. However, the surviving DNA lineages, which “surfaced” only recently, particularly in Tuva, are different in Tuva and in Altay, though all coalescent to said ancient common ancestor.

R1a Haplotypes in the Eastern Himalayas

Five R1a1 haplotypes were listed in (Kang et al., 2011), which showed a rather recent base haplotype (the last two markers are DYS437 and DYS438):

13 25 15 10 11 14 13 14 10 13 11 30—14 11

These 5 haplotypes have only 4 mutations from the above base haplotype, which gives $4/5/.0215 = 37 \rightarrow 38$ generations; that is, 950 ± 480 years to a common ancestor. However, the above base haplotype has very unusual (for R1a haplogroup) alleles DYS426 = 13, and DYS388 = 14, and differs by 5 mutations with the Russian Plain base haplotypes. This places their common ancestor at 6650 ybp. This is clearly a separate branch of ancient R1a haplotypes in Eastern Himalayas.

R1a Haplotypes in India and Pakistan

There are two principal sources of haplotypes of haplogroup R1a in the Hindustan. One was brought by the Aryans in the middle of the 2nd millennium BC, as it was described above, and supported below with more extended series of Indian haplotypes. A timespan to the most recent common ancestor of these haplotypes varies between 4000 and 4600 ybp, and often around 4050 ybp, depending on a particular haplotype datasets. The base (ancestral) haplotype of the Aryan (Indo-European) haplotype in its 12 marker format is

13 25 16 10 11 14 12 12 10 13 11 30

This haplotype is nearly identical to that of the Russian Plain base (see below), except the latter came from a common ancestor who lived between 4,600 and 5,000 ybp as determined using different haplotype datasets (Klyosov, 2009a; Klyosov, 2011b).

A more ancient source is presumably the South Siberian and/or Central Asian haplotypes brought to the Hindustan during the westward migrations of R1a bearers between 20,000 and 10,000 ybp. Some studies alleged that the most ancient common ancestors of R1a haplotypes were Indian; however, the results were flawed by erroneous calculations of timespans using incorrect “population mutation rates” (see their description and discussion in Klyosov, 2009a, 2009c, and references therein), which routinely converted the actual 3600 - 4000 ybp (“Indo-European” R1a1 in India) into 12,000 - 15,000 ybp. This was erroneously claimed as the proof of “origin of R1a in India.” Furthermore, high percentages of R1a in some regions in India or in some ethnic and/or religious groups (such as Brahmins) were incorrectly claimed as the proof of the origin of R1a in India (Kivisild et al., 2003; Sengupta et al., 2006; Sahoo et al., 2006; Sharma et al., 2009; Thanseem et al., 2006; Fornarino et al., 2009). The application of the flawed approach resulted in confusion amongst researchers in the field of human population genetics over the last decade. The course of research is hopefully corrected by the application of today’s most recent developments of DNA genealogy, which utilizes a principally different methodology (Klyosov, 2009a, 2009b, 2009c; Rozhanskii and Klyosov, 2011; Klyosov, 2011b).

Forty-six of 6 marker R1a1 haplotypes of three different tribal population of Andra Pradesh, South India (tribes Naikpod, Andh, and Pardhan) listed in (Thanseem et al., 2006) and shown in the haplotype tree in **Figure 2**, contain 126 mutations; that is $.457 \pm .041$ mutations per marker (the mutation rate constant equals $.0123$ mutation/haplotype/generation, Klyosov, 2009a). It gives 7200 ± 960 years to a common ancestor (see Material and Methods for calculations). The base (ancestral) haplotype of those south Indian populations in the FTDNA format is as follows:

13 25 17 9 X X X X 14 X 32

This differs from the north-Indian “Indo-European” haplotype

(see above) by four mutations on six markers, which places their most recent common ancestor to approximately 11,600 ybp (see Materials and Methods for calculations).

The ancient north China R1a1 base haplotype (see above) differs from the Andra Pradesh R1a1 base haplotype by at least 5 mutations on the 5 available markers, which places their common ancestor at approximately 22,000 ybp. Within a margin of error, it can be deduced that this is the same common ancestor of the north China haplotypes. This mutational difference neatly fits the chronology and direction of the migration, which continues from the ancient (non Indo-European) Indian haplotypes to the Indo-European Indian haplotypes with their common ancestor (non-IE and IE) who lived approximately 11,600 ybp. Also, this data dovetails with the timing of the follow-up migration of R1a1 bearers from Hindustan via Asia Minor (with a detection of the proto Indo-European language in Anatolia with estimated divergence time of 9400 - 9600 - 10,100 ybp, see Gray & Atkinson, 2003; Renfrew, 2000; Gamkrelidze & Ivanov, 1995) to Europe (with the arrival 10,000 - 8000 ybp, see below) and then to the Russian Plain (5000 - 4800 ybp, see below).

The analysis of this data and of these findings essentially unites most, if not all, concepts of the “origin of Indo-European language” which have, at various times, placed the “origin” from India to Iran, Anatolia, the Balkans, to the Russian Steppes (Gimbutas, 1973, 1994; Mallory, 1989; Dixon, 1997; Anthony, 2007), except that they were related not to the “origin,” but to the passing areas of the R1a1 migration.

Population geneticists typically mix DNA lineages and branches in their analysis whereby “phantom common ancestors” emerge. This is exemplified with 110 of 10-marker R1a1 haplotypes of various Indian populations, both tribal and Dravidian and Indo-European castes, listed in (Sengupta et al., 2006). The resulting mixed haplotype tree is shown in **Figure 3**. It contains 344 mutations, which is $.313$ mutations per marker, and results in a “phantom” 5275 years to a “common ancestor,” just between the shown above 7180 ± 960 years for non-IE and 4050 ± 500 IE Indian haplotypes.

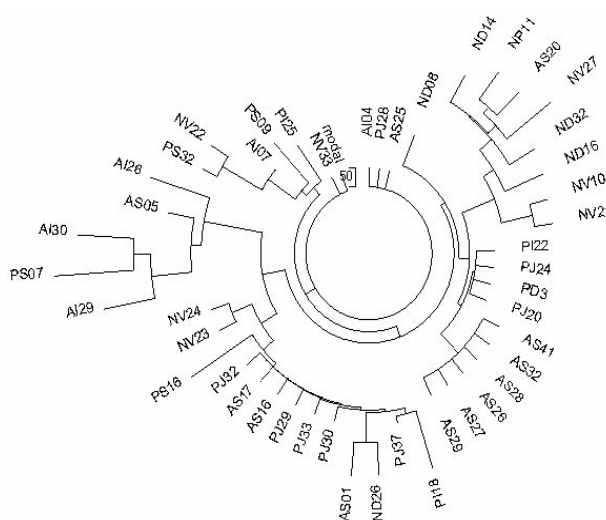


Figure 2. The 6-marker haplotype tree for R1a1 haplotypes in Andra Pradesh (tribes Naikpod, Andh, and Pardhan), South India. The 46-haplotype tree was composed from data listed in (Thanseem et al., 2006). The designations of haplotypes are those used in the article.

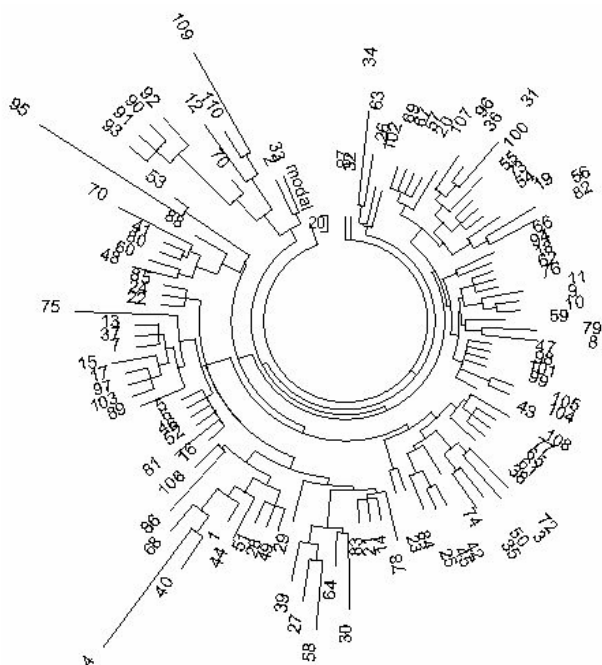


Figure 3. The 10 marker haplotype tree for R1a1 haplotypes in India (mixed population, including tribes and castes). The 110-haplotype tree was composed from data listed in (Sengupta et al., 2006). The article contains 114 Indian R1a1 haplotypes, however, four of them were incomplete.

For a comparison, consider the Pakistani R1a1 haplotypes listed in the Sengupta (2006) paper (**Figure 4**). Forty-two haplotypes contain 166 mutations, which give $.395 \pm .031$ mutations per marker, and 6800 \pm 860 years to a common ancestor. This value fits within margin of error to the “south-Indian” 7200 \pm 960 ybp; however, the base (ancestral) haplotypes differ significantly. The base Pakistani haplotype is as follows (in the FTDNA format plus DYS461):

13 25 17 11 X X X 12 10 13 11 30—9

It differs from the south Indian “non-IE” and the north Indian “IE” base haplotypes by four and two mutations on six markers, respectively. This places a common ancestor of the Pakistani and the south Indian “non-IE” R1a1 populations at approximately 12,980 ybp which is within margins of error with the 11,600 ybp reported above as the migration time through the Hindustan westward. The two mutations place a common ancestor of the Pakistani and the “Indo-European” Indian populations more recently, at 7800 ybp. This chronological trend might also point in the direction of the ancient migration of R1a1 westward.

A more detailed consideration of the Pakistani R1a1 haplotypes, including separate calculations of each of the four branches in **Figure 4**, results in a timespan of 8650 years to a common ancestor for all of these branches (Klyosov, 2010a). In all, it does not change the principal conclusions of this section.

R1a1 Haplotypes in Central Asia

Ten 10-marker Central Asian haplotypes were listed in (Sengupta et al., 2006). They contain 27 mutations from the base haplotype

13 25 16 11 X X X 12 10 13 11 31-- 9

which gives $.270 \pm .052$ mutations per marker, and 4300 \pm 940 years to a common ancestor. It is the same value that we have found for the Russian Plain and “Indo-European” Indian base R1a1 haplotypes.

Both the Central Asian base haplotype and the dating of a common ancestor described above are supported by the latest data on extended 67 marker haplotypes that were collected in November 2011 in the R1a1 FTDNA Project. The Central Asian base haplotype was as follows:

13 25 16 11 11 14 12 12 10 13 11 31—15 9 10 11 11 24 14 20 32 12 15 15 16 -- 11 11 19 23 15 15 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

This is identical to the above-mentioned Central Asian 10-marker base haplotype in all the given alleles. A common ancestor for the series of the extended haplotypes lived 3650 \pm 590 ybp. The Central Asian base haplotype differs from the Russian Plain base haplotype (Rozhanskii & Klyosov, 2009) by only 4 mutations in the 67 markers. This separates them in terms of the time of their common ancestors

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 18 19 34 38 13 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

by only $4/12 = 33 \rightarrow 34$ conditional generations; that is by ~850 years.

The result is compelling and provides an exact fit with the expected migration pattern of the R1a1 haplogroup from the Russian Plain (~4600 - 4400 ybp) to Central Asia (3650 \pm 590 ybp) on their way to the South Urals and to the Hindustan.

What these findings suggest is that there are two different subsets of the Indian R1a1 haplotypes. One was brought by European bearers known as the Aryans, seemingly on their way from the Russian Plain through Central Asia in the middle of the 2nd millennium BC. The other was much more ancient and migrated from South Siberia/northern China to India 12,000

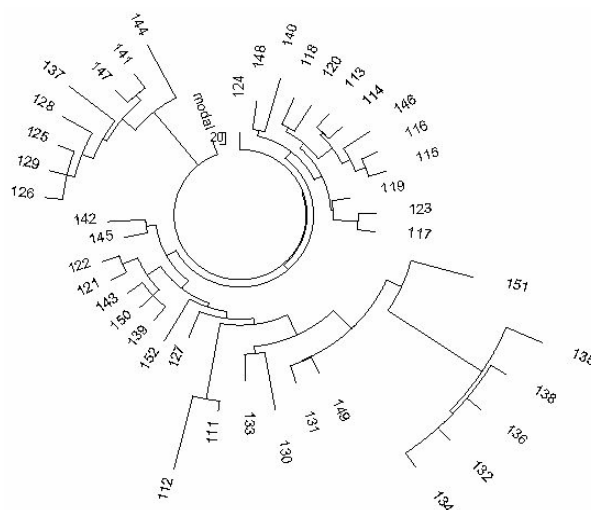


Figure 4. The 10-marker haplotype tree for R1a1 haplotypes in Pakistan. The 42-haplotype tree was composed from data listed in (Sengupta et al., 2006).

years ago. This migratory wave continued through the Iranian Plateau westward (via Anatolia and the rest of Asia Minor), to the Balkans and then further into the European continent.

R1a1 Haplotypes of the Comoros Islands

Fifteen R1a haplotypes have been found among 381 tested men on the Islands. Three of them were R1a*-SRY10831a, and twelve were R1a1 (Msaidie et al., 2011). The cited study did not generate any chronological estimates based on the haplotypes, and considered only 8 marker haplotypes (for a typical “population genetics” analysis without separation of haplogroups) while, in fact, determined 17 marker haplotypes.

The base haplotype for said 12 R1a1 haplotypes is as follows:

13 24/25 15 11 12 14 X X 10 13 11 18—16/17 14 19 12 15 11 23

All have 104 mutations; that is, $.51 \pm .05$ mutation per marker. This high value points at a significantly more ancient common ancestor compared with the that in the Russian Plain, Central Asia and the Indo-European Indian R1a1 populations (.28, .27, .24 mutations per marker, respectively). Furthermore, it is more ancient compared with the old south Indian and Pakistani R1a1 populations (.457 and .395 mutations per marker, respectively, see above). Indeed, a common ancestor of the Comoros R1a1 haplotypes lived $.51/.02 = 255 \rightarrow 340$ conditional generations; that is 8500 ± 1190 ybp.

For a comparison, the Russian Plain base haplotype in the same format is as follows:

13 25 16 11 11 14 X X 10 13 11 17—15 14 20 12 16 11 23

It differs by as many as 7 mutations from the Comoros base haplotype. This places their common ancestor at 9900 ybp. It is reasonable to suggest that this common ancestor was one of those R1a1 who were moving westward along the Iranian plateau and Asia Minor almost 10,000 ybp. Indeed, the dating around 9900 ybp is rather typical for archaeological settlements in Asia Minor with known dates of 10,200, 9900 and 9000 ybp (Myres et al., 2010). It is not necessarily true that the bearers of R1a1 were in the Comoros Islands 9900 ybp since it is known that the Persian traders had expanded their maritime routes to Madagascar by 700-900 AD (Msaidie, 2011).

R1a1 Haplotypes in the Arabian Peninsula

Sixteen R1a1 10 marker haplotypes from Qatar and United Arab Emirates were published (Cadenas et al., 2008). They split into two branches, and their base haplotypes

13 25 15 11 11 14 X Y 10 13 11 30

13 25 16 11 11 14 X Y 10 13 11 31

differed by only one mutation. Their common ancestor lived 3750 ± 825 years bp. Since a common ancestor of R1a1 haplotypes in Armenia and Anatolia lived 4500 ± 1040 and 3700 ± 550 years bp, respectively (Klyosov, 2008), the three dates do not conflict with each other. They were not part of the ancient migrations of 12 - 9 thousand ybp, but they were most likely on the military expeditions of the (Aryan) R1a1 from the Russian Plain southward through Anatolia, Mitanni, and to the Middle East and the Arabian Peninsula around 4000 - 3600 ybp. As mentioned earlier, today there are between 3% and 9% of R1a1 in those regions, among them members of famous tribes such as

Quraish/Quraysh (Muhammad, the founder of the religion of Islam, was born into the Quraysh tribe), Al Tamimi (Banu Tamim) and others.

Much more reliable data are obtained with extended 67 marker haplotypes from the Arabic FTDNA project. Twenty-seven haplotypes from Qatar, Kuwait, Saudi Arabia, UAE, Oman, Bahrain and Syria form a separate branch on the haplotype tree and result in the following base haplotype:

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 15 16 18 19 35 38 13 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

Therein 499 mutations exist and $499/27/.12 = 154 \rightarrow 182$ generations; that is, 4550 ± 500 years from a common ancestor of the Arabic haplotypes—practically the same as that for the Russian plain R1a1 common ancestor (see above). The two differ by only 1.4 mutations in all the 67 markers; that is, $1.4/.12 = 12$ generations apart, a ~300 year difference between the Russian Plane R1a1 common ancestor and the Arabic haplotypes common ancestor. The exception being that the Arabic haplotypes are typically coupled with the downstream L342 SNP mutation. The difference places their common ancestor at ~4825 ybp, which is the Russian Plain base (ancestral) haplotype. This is the same Aryan haplotype that was brought ~4500 ybp from the Russian Plain in a star-like manner to India, Iran, Anatolia, the Arabian Peninsula to arrive there a thousand years later, in the middle of the 2nd millennium BC.

Recent developments in the phylogeny of R1a1 haplotypes coupled with the DNA genealogy analysis have shown that the migrations of R1a1 from the Russian Plain in the described star-like manner were accompanied with the R1a1-L342 (around 4400 ybp) and then its downstream L657 subclade. The L342 subclade is almost absent on the Russian Plain, and it appears in the Bashkir population in the east, in Kazakhstan (L342 \rightarrow L657) south-east, in India (L342 \rightarrow L657), and in the Middle East (including the Arabian Peninsula, L342 \rightarrow L657). It shows primary directions of the Aryan (R1a1) migrations after ~4800 ybp.

The Arabian R1a1-L657 haplotypes along with all known Iranian, Indian and Kazakh L657 haplotypes have the following L657 base haplotype:

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

Its common ancestor lived 3000 ± 400 ybp. The above base haplotype differs by 9.85 mutations from the Russian Plain base haplotype (some mutations are fractional ones), which places the R1a1-L657 and R1a1 Russian plain common ancestor at 5000 ± 600 years bp.

R1a1-L342 Bashkir and Szekely/Seklers (Hungarian) Haplotypes

As noted in the preceding section, migrations of the ancient Aryans eastward (and in some cases westward, as illustrated below with the Szekely L342 R1a1 haplotypes) have resulted in the appearance of the downstream R1a1 subclades, such as L342, among the Bashkirs. The respective L342 base haplotype is as follows:

13 24 16 11 11 15 12 12 12 13 11 31—15 9 10 11 11 24 14
 20 31 12 15 15 15—11 10 19 23 16 15 19 19 35 38 14 11—11
 8 17 17 8 12 10 8 11 10 10 22 22 15 10 12 12 13 8 14 23 21 12
 12 11 13 11 11 12 13

Their common ancestor lived only 1300 ± 250 ybp; however, the base haplotype differ by as many as 14 mutations from the Russian Plain base haplotype. This places their common ancestor, for the Bashkirs and the Russian Plain, at 4700 ± 500 ybp. This is again the Aryan R1a1 common ancestor on the Russian Plain.

There is quite a distant L342 lineage among descendants of Hungarian Szekely servicemen, recorded in the first 1602 military census. The lineage is only 675 ± 260 years “old”. However, its base haplotype

14 23 17 11 11 14 12 12 10 14 11 32—17 9 10 11 11 24 14
 20 31 12 14 15 15—11 12 19 23 16 15 20 19 34 38 13 11

contains as many as 15 mutations in the first 37 markers from the respective L342 Bashkir base haplotype. This places their common ancestor to 3500 ± 400 ybp. It apparently reflects migrations of R1a1-L342 bearers from the Ural region westward to Transylvania along with Finno-Ugric migrations of those times.

R1a Haplotypes along the Ancient Migration Path from South Siberia to Europe

Three principal studies have been published recently, that contain hundreds of R1a1 haplotypes from all over the world (Underhill et al., 2009; Zhong et al., 2010; Shou et al., 2010). Analysis of those haplotypes and the chronology of their common ancestors have not been undertaken by the authors of these studies. **Figures 5-7** show general views of R1a1 haplotype trees, that were calculated from the data. The purpose for including pictures of these trees was not to analyze their fine structure in detail (Klyosov, 2010a, 2010b), but to demonstrate their complex multi-branch structure, hence, ancient origins. For example, relatively young trees (young “age” of their common ancestor) are often rather symmetrical and relatively uniform, such as the Russian Plain R1a1 haplotype tree with a common ancestor 4600 ybp (**Figure 8**).

Analysis of R1a1 haplotypes and their branches on the trees in **Figures 5-7** shows that their ancient common ancestors lived in south Siberia and Altay (belonging to both south Siberia and Central Asia). Their ancient descendants carried the R1a1 haplogroup while migrating from North and North-Western China, across Tibet and Hindustan, and then along the Iranian Plateau, from Asia Minor and finally into Europe. Some remnants of ancient R1a1 were left in Cambodia, Nepal, Oman, Israel, Iraq, Egypt, Crete, the Caucasus, Russia, Estonia (the respective haplotypes are recovered from data published in Underhill et al., 2009, Zhong et al., 2010, Shou et al., 2010). Results of the dynamics of mutation in these haplotypes significantly differ from those in the contemporary European R1a1, except one ancient and distinct lineage of R1a1 in Europe (see below). Their common ancestors as thusly reconstructed, lived from 20,000 ybp in south Siberia/northern China through 12,000 - 11,000 ybp in Hindustan and 6900 ybp in Uyghurs in north-western China.

Typically, ancient common ancestors are recognized by the distinct $DYS392 = 13$, unlike typical $DYS392 = 11$ in most of European (and elsewhere) R1a1 haplotypes. The study by Underhill et al. (2009) listed four Egyptian R1a1 haplotypes, two

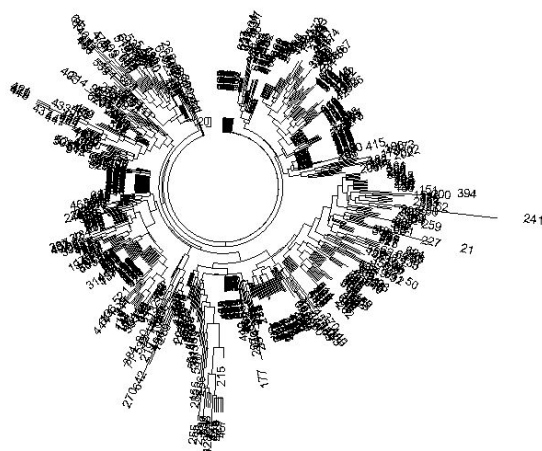


Figure 5. R1a1 10 marker 638-haplotype tree from all over the world, composed based on data published by Underhill et al. (2009).

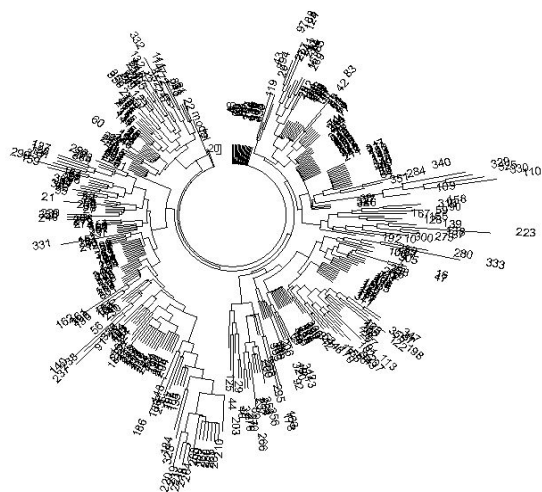


Figure 6. R1a1 8 marker 365-haplotype tree from all over the world, composed based on data published by Zhong et al. (2010).

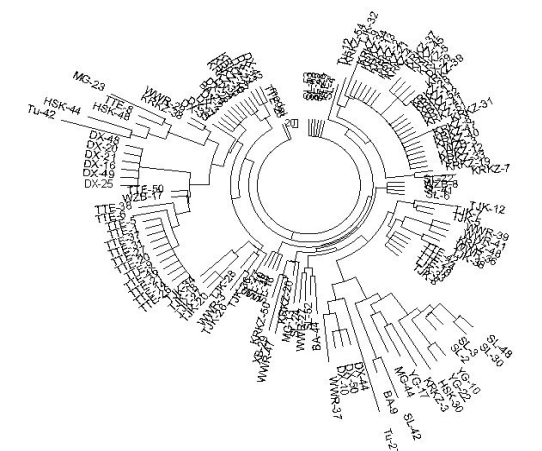


Figure 7. R1a1 8 marker 131-haplotype tree collected in North-Western China, composed based on data published by Shou et al. (2010). The lower right branch consists of R1⁺ haplotypes.

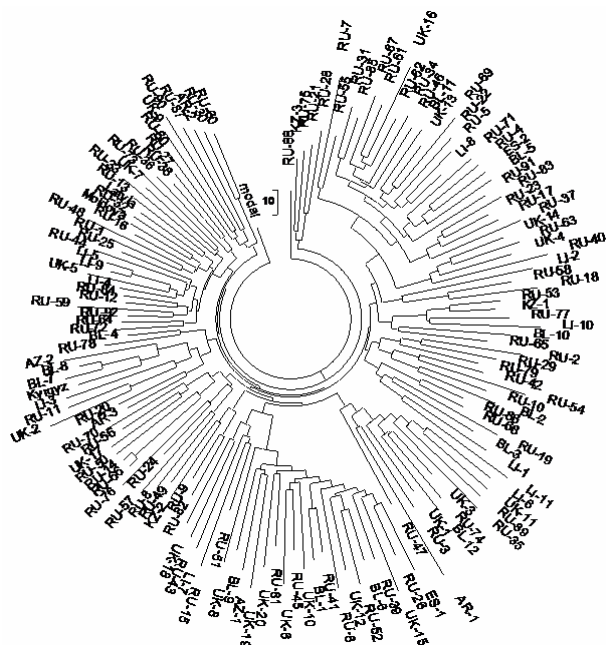


Figure 8. R1a1 67 marker 148-haplotype tree collected in Russia and Ukraine (the Russian Plain), published by Klyosov (2011b).

of them having DYS393 = 13, and two DYS393 = 11. These four haplotypes have their common ancestor of ~13,275 ybp.

The very top of the tree (Figure 6) contains 18 base haplotypes, which are identical to each other, and expressed in the 9 marker format as follows:

13 25 16 11 X X X 12 10 13 11 30

In this particular case these identical haplotypes are from Russia, Turkey, Ukraine, Slovakia, Iran, Nepal, India, and Hungary. The short haplotype format does not allow them to be resolved any further, but with the available 9 markers this base haplotype is an exact (albeit partial) reproduction of the base haplotype of the Russian Plain. Furthermore, the tree in Figure 8 produces exactly the same 67 marker base haplotype of the Russian Plain. The whole tree contains 148 haplotypes with 2748 mutations from the base haplotype. It produces $2748/148/67 = .277 \pm .005$ mutations per marker, and $.277/.12 = 155 \rightarrow 183$ generations; that is 4575 ± 470 years to a common ancestor of the Russian Plain base haplotype.

The whole pattern of ancient migrations of bearers of R1a1 haplotypes shows that after they had arrived to Europe via Asia Minor, as it is described above, between 11,000 and 8000 ybp (see below), they moved to the Russian Plain in the beginning of the 3rd millennium BC. It coincided time-wise with the arrival of bearers of R1b1a2 haplotypes in Europe. From there R1a1 split into three principal streams. One stream migrated south, over the Caucasus to Anatolia, the Middle East and the Arabian Peninsula. The second stream went eastward to South Ural, the Andronovo and Sintashta archaeological cultures in the 2nd millennium BC, between 4000 and 3000 ybp, and then split into two migration paths. One went south to India as the legendary Aryans, another went further east to Altay and the Northern China. This closed the loop of the ancient migrations of R1a1. Yet the third stream went south-east to the mountainous terrain of Middle Asia in ~4000 ybp, and some 500 years

later moved to Iran, otherwise known as the “Avesta Aryans.”

R1a1 Haplotypes of the Current Descendants of the Legendary Aryans in India

The R1a1 FTDNA Project in November 2011 contained 101 Indian haplotypes. Their base haplotype in the 25 marker format

13 25 16 10 11 14 12 12 10 13 11 30—16 9 10 11 11 24 14 20 32 12 15 15 16

contained only 1.4 mutations on average, compared with the Russian Plain base haplotype (see above). This translates into $1.4/.046 = 30 \rightarrow 31$ generations, or ~775 years between their common ancestors. In terms of time, this is a close distance between the Russian Plain and the Indian base haplotypes, and it fits with the time spans for the Russian Plain R1a1 common ancestor (4600 - 4800 ybp) and the Indian common ancestor (~4050 ybp), determined independently. This is the historical Aryan base haplotype.

Figure 9 shows that the Indian R1a1 haplotype tree contains five principal branches. Their base haplotypes in the 25 marker format are as follows (clockwise from the top):

13 24 16 10 11 14 12 12 10 13 11 30—15 9 10 11 11 23 14 20 32 12 15 15 16

13 25 16 11 11 14 12 12 10 13 11 30—16 9 10 11 11 24 14 20 32 12 15 15 16

13 25 15 10 11 14 12 12 10 14 11 30—16 9 10 11 11 25 14 20 32 12 15 15 16

13 25 16 10 12 14 12 12 10 13 11 31—16 9 10 11 11 24 14 20 32 12 15 15 16

13 25 16 10 11 14 12 12 10 13 11 31—15 9 10 11 11 24 14 20 32 12 15 15 16

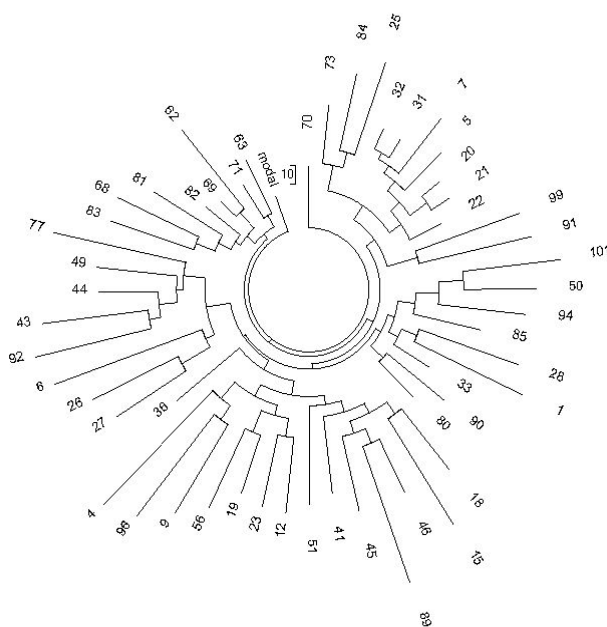


Figure 9. R1a1 53-haplotype 25-marker tree collected in the Indian FTDNA Project database (November 2011). The Project contained 101 of 12 marker haplotypes, but only 53 of them were in the 25 marker format.

A superposition of these five base haplotypes gives the above Indian “Indo-European”, the Aryan R1a1 base haplotype. All five base haplotypes differ collectively by 12 mutations from their ancestral (see above) base haplotype, which translates 4050 ± 500 years to their ancestral haplotype.

It should be noted that datasets of Indian R1a1 haplotypes are difficult to analyze, because they typically represent a superposition of haplotypes from various sources, including those from the ancient (pre-IndoAryan) ancestors, from the Russian Plain, Central Asia, the Middle East, etc. Since they all present in various amounts and proportions, only analysis of their haplotype trees can give meaningful results.

Other Scattered Ancient R1a1 Haplotypes in Asia

A “patchy” pattern of R1a1 was created by the territorially blending of ancestors from the very ancient R1a1 (from more than 10,000 - 15,000 ybp) to the rather recent, the Aryan migrations. The tree in **Figure 5** presents some haplotypes from Nepal, which differ by 5 mutations from the Russian Plain base haplotype, pointing to a common ancestor of 7200 ybp. Some Indian haplotypes show 7-mutation difference from the Russian Plain haplotype with a common ancestor of 10,200 ybp. A Cambodian haplotype makes 9 mutations, which places a common ancestor of the Russian Plain base haplotype and the Cambodian haplotype at 14,000 ybp. Some haplotypes from Pakistan, Iran, Oman and Arab Emirates show 5 - 6 mutations, pointing to a common ancestor of 7000 - 9000 ybp. A group of ethnic minorities from north-western China (Tu, Xibe, Tatars, Uyghurs, Yugurs, Salars, Bonan and others) typically have their collective R1a1 common ancestor of 6900 ybp (Klyosov, 2010b). All of them reportedly have the following base haplotype, obtained from the tree in **Figure 7**:

13 25 16 11 X X X 12 X 14 12 31

The value of $DYS392 = 12$ is a suspect, because it is an unusual one for R1a1, including those from Central Asia. However, almost all Asian haplotypes in (Shou et al., 2010) are reported as having this “12” allele. The difference in 2.85 mutations with the base Russian Plain or with the IE Indian base haplotype (if $DYS392 = 12$ is correct) or 1.85 mutation (if $DYS392 = 11$ is correct) places a common ancestor of the north-western Central Asian R1a1 haplotypes and the Russian/Indian haplotypes to either 9350 ybp or 7925 ybp, respectively. In any case they are significantly more ancient compared with the majority of European haplotypes.

A Presumably Ancient (10,000 - 7700 ybp) R1a1 Population in Europe

There is a distinct group of R1a1 bearers in Europe who have $DYS392=13$, while the overwhelming majority of R1a1 haplotypes in Europe and elsewhere has $DYS392 = 11$. The base haplotype of this group, coined “The Old European branch” (Rozhanskii and Klyosov, 2009) 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 17 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

This haplotype differs from the Russian Plain R1a1 haplotype by 6 mutations in the first 12 markers, 12 mutations in the first 37 markers, 20 mutations in the first 37 markers, and 24

mutations in all the 67 markers, which places a common ancestor of their haplotypes at 6,800 ybp. Besides, if compared to the so-called North-western base haplotype (Rozhanskii & Klyosov, 2009)

13 25 16 10 11 14 12 10 10 13 11 30—15 9 10 11 11 24 14 19 32 12 15 15 16—11 11 19 24 16 15 18 18 33 38 13 11—11 8 17 17 8 12 10 8 11 10 12 22 22 15 10 12 12 14 8 14 23 22 12 13 11 13 11 11 12 13

the difference is 9, 14, 24 and 29 mutations, respectively. Their common ancestor lived ~7400 ybp.

Lastly, there is an additional comparison worthy of consideration, and it regards the Central European branch (Rozhanskii & Klyosov, 2009) as follows:

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

Herein the difference is 9, 15, 25 and 34 mutations, respectively. Their common ancestor lived ~7700 ybp.

A series of 67 haplotypes of haplogroup R1a1 from the Balkans was published (Barac et al., 2003a,b; Pericic et al., 2005). In print, there were presented in the 9 marker format, and the respective haplotype tree is shown in **Figure 10**. Most of the tree contains typical European haplotypes with a common ancestor of ~4500 ybp. However, the left branch is distinctive since it contains R1a1 haplotypes with $DYS392 = 13$, such as

12 24 16 10 12 15 X X X 13 13 29

12 24 15 11 12 15 X X X 13 13 29

13 24 14 11 11 11 X X X 13 13 29

This branch was calculated using the both linear and the quadratic permutation methods. It obtained $.598 \pm .071$ mutations on average per marker, which resulted in $11,425 \pm 1780$ and $11,650 \pm 1550$ years to a common ancestor, respectively (Klyosov, 2009a).

Calculations with short haplotypes are subject to high margins of error compared with extended 67 marker haplotypes. The authors of this study consider the timespans to a common

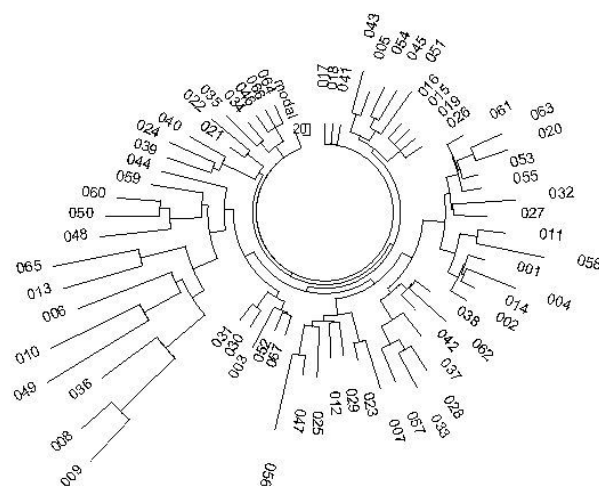


Figure 10. The 9-marker 67-haplotype tree for the Balkans, haplogroup R1a1. The tree was composed from data (Barac et al., 2003a, 2003b; Pericic et al., 2005).

ancestor of R1a1 haplotypes in Europe of 7400 - 7700 ybp to be more reliable in comparison to the 11,000 - 12,000 ybp. Haplotype databases continue to expand, and future studies will reveal the lower limit of “age” of haplogroup R1a in Europe. Thus, dismissing the data of **Figure 10** would be premature.

Materials and Methods

Four thousand four hundred sixty (4460) of R1a1 haplotypes were collected in databases from FTDNA, YSearch and SMGF (Sorenson Database), in the private databases by Martin Voorwinden (987 of the Tenth, DYS388 = 10 haplotypes) and the IRAKAZ (2018 of R1a1 haplotypes, regularly updated by Alexander Zolotarev and Igor Rozhanskii), and in peer review publications.

The methodology of haplotype datasets analysis was described in (Rozhanskii & Klyosov, 2011). In this study the linear and the quadratic permutation method were used, the latter when the base haplotype could not be decisively determined, as described in (Klyosov, 2009a). The mutation rate constants are listed in (Klyosov, 2009a; Rozhanskii & Klyosov, 2011), and for a number of cases are given in the text of this paper.

The mutation rate constant for the non-standard 25 marker haplotype discussed in the “Haplotype from Altay” was calculated using the respective father-son pair frequencies of transmissions listed in (Burgarella & Navascues, 2011). We have shown in (Rozhanskii & Klyosov, 2011) that for the 12 marker haplotype panel, the mutation rate constant equals .0200 mutation/haplotype/generation. Indeed, in the (Burgarella & Navascues, 2011) paper the sum of the respective ten frequencies (those for DYS385a, b were not determined) were equal to .0201 mutation/haplotype/generation.

In other words, our calibrated mutation rate constant for the 12-marker panel fits fairly well to the sum of average frequencies in father-son pairs. For the 15-marker second half of the panel (from DYS458 to YPenta2) the sum of the frequencies of father-son pairs (Burgarella & Navascues, 2011) was equal to .0204 mutation/generation, which makes the mutation rate constant for the whole 25 marker haplotype of .0404 mutation/haplotype/generation.

Haplotype trees were composed using software PHYLIP, Phylogeny Inference Package program (see Klyosov, 2009a, 2009b and references therein). Corrections for back mutations were introduced as described in (Klyosov, 2009a; Rozhanskii & Klyosov, 2011). Margins of error were calculated as described in (Klyosov, 2009a). Explanations and examples are given in this section.

Base haplotypes in the dataset were determined by minimization of mutations; by definition, the base haplotype is one which has the minimum collective number of mutations in the dataset. The base haplotype is the ancestral haplotype or the closest approximation to the latter.

A timespan to the common ancestor of two base haplotypes is determined as follows: 1) count the number of mutations between the two base haplotypes; 2) divide the obtained number by the mutation rate constant; 3) introduce a correction for back mutations, calculated using the following formula (Adamov & Klyosov, 2008; Klyosov, 2009a).

$$\lambda = \frac{\lambda_{obs}}{2} (1 + \exp(\lambda_{obs})) \quad (1)$$

where: λ_{obs} = observed average number of mutations per marker

in a dataset (or in a branch, if the dataset contains several branches/lineages), λ = average (actual) number of mutations per marker corrected for back mutations; 4) add the obtained value, multiplied by 25 (years), which represent the “lateral” timespan between times of appearance of the two base haplotypes, to TMRCA for the both base haplotypes and divide by 2. The result represents the TMRCA (time for the most recent common ancestor) for the two base haplotypes under study.

Example 1: Calculation of a timespan to a common ancestor when an average number of mutations per marker is equal to .395 (a series of Pakistani 10 marker haplotypes in this paper), and the mutation rate constant for the 10 marker haplotype is .018 mutation/haplotype/generation (25 years), or .0018 mutation/marker/generation. $.395/.0018 = 219$ generations without a correction for back mutations. Since the observed number of mutations per marker is .395, we employ formula (1) and obtain

$$\frac{1}{2}(1 + \exp(.395))$$

In order to calculate $\exp(.395)$, that is $e^{.395}$, we need to find a number, the natural logarithm of which is equal to .395. This number is 1.4844. Then we have

$$\frac{1}{2}(1 + 1.4844) = 1.2422$$

The obtained number of 1.2422 is the coefficient of the correction for back mutations. Therefore, by multiplying 219×1.2422 , we obtain that the corrected number of generations is 272, that it $272 \times 25 = 6800$ years. This is usually designates as $219 \rightarrow 272$ (generations). Since for 166 mutations in the dataset the margin of error is 12.66% (calculated as explained in Klyosov, 2009a), we at last obtain the timespan to a common ancestor of the Pakistani haplotypes is equal to 6800 ± 860 years.

Example 2: Forty-six (46) of 6 marker haplotypes of Andra Pradesh, South India contain 126 mutations; that is, $126/46/6 = .457 \pm .041$ mutations per marker (the mutation rate constant equals .0123 mutation/haplotype/generation; that is, .00205 mutation/marker/generation, Klyosov, 2009a). The number of generations to the common ancestor equals to $.457/.00205 = 223$ generations (25 years each), without a correction for back mutations. As explained in Example 1, since the observed number of mutations per marker is .457, formula (1) gives us the exponent equal to 1.5794, the coefficient of the correction equal to 1.2897, the number of generations to the common ancestor $223 \rightarrow 288$; that is 7200 years before present.

Example 3: Thirty-one (31) of 5 marker haplotypes of northern China, calculated by the quadratic permutation method. For the given series the sum of squared differences between each allele in each marker equals to 10,184. It should be divided by the square of a number of haplotypes in the series (961), by a number of markers in the haplotype (5) and by 2, since the squared differences between alleles in each marker were taken both ways. This gives an average number of mutations per marker of 1.060. After division of this value by the mutation rate of .00135 mutation/marker/generation (for 25 years per generation), $19,625 \pm 2800$ years to a common ancestor is derived. This result is within the margin of error with that calculated by the linear method: $99/31/5/.00135 = 472 \rightarrow 683$ conditional generations; that is, $17,100 \pm 2400$ years ago to the common ancestor of these 31 haplotypes. The values of .00677 and

.00135 are the mutation rate constants for the 5 marker haplotypes expressed in mutation/haplotype and mutation/marker, respectively, for conditional generation of 25 years (Klyosov, 2009a).

Margins of error were calculated as follows. First, a standard deviation (SD, one sigma) was calculated for an average number of mutations per marker, which is a reciprocal square root of a total number of mutations in the dataset (Klyosov, 2009a). A square root of the sum of SD^2 and $.10^2$ (the last figure corresponds to the square of the standard deviation of the mutation rate constant) gives the margin of error of the timespan to the common ancestor of the haplotypes in the dataset, provided that all of them are derived from the same the most recent common ancestor. For the dataset with 126 mutations (see above) the standard deviation is 8.91%, and the overall margin of error for the timespan of 7200 years is 13.4%; that is, 7200 ± 960 years.

A detailed examination of many “classical” genealogies has shown that the above procedural results are the best fit with actual data.

The same calculations are generated for two base haplotypes when a timespan to their common ancestor is sought. For example, the two Indian base haplotypes in **Figure 9** (two upper right-hand side branches)

13 24 16 10 11 14 12 12 10 13 11 30—15 9 10 11 11 23 14
20 32 12 15 15 16

13 25 16 11 11 14 12 12 10 13 11 30—16 9 10 11 11 24 14
20 32 12 15 15 16

differ by 4 mutations in the 25 markers (the mutation rate constant equals .046 mutation/haplotype/generation). Therefore, we obtain $4/.046 = 87 \rightarrow 96$ generations; that is, $96 \times 25 = 2400$ years from the average “age” of the base haplotypes to their common ancestor. Since we have determined in a separate set of calculations that the common ancestors of the both branches lived 1200 and 2900 ybp, their common ancestor lived $(1200 + 2900 + 2400)/2 = 3250$ ybp.

Assignments of haplotypes to haplogroups and subclades were based on their SNP classification, as provided in the databases. In some instances it was additionally supported calculating their position on the phylogenetic trees from their respective STR data.

Conclusion

The results of this study lend a support to the theory that haplogroup R1a arose in Central Asia, apparently in South Siberia or the neighboring regions, such as Northern and/or Northwestern China, around 20,000 years before present. The preceding history of the haplogroup is directly related to the appearance of Europeoids (Caucasoids) ~58,000 ybp, likely in the vast triangle that stretched from Western Europe through the Russian Plain to the east and to Levant to the south, as it was suggested in the preceding article (Klyosov & Rozhanskii, 2011). A subsequent sequence of SNP mutations in Y chromosome, with the appearance of haplogroups NOP ~48,000 ybp and P ~ 38,000 ybp in the course of their migration eastward to South Siberia, eventually gave rise to haplogroup R ~30,000 ybp and R1 ~26,000 ybp, and then to haplogroup R1a and R1a1 ~20,000 ybp. The timeframe between the appearance of R1a and R1a1 is uncertain.

At some point in time, the bearers of R1a1 began migration to the west, over Tibet and the Himalayas, and not later than

12,000 ybp they were in the Hindustan. They continued their way across the Iranian Plateau, along Anatolia and Asia Minor apparently between 10,000 and 9000 ybp. By 9000 - 8000 they arrived in the Balkans and spread westward over Europe and to the British Isles. At that point, R1a1 still had DYS392 = 13 in their haplotypes, as did their brother haplogroup R1b1. This marker is very slow, and mutates on average once in 3500 conditional generations. Somewhere on this extended timescale, bearers of R1a1 (or the parent, upstream haplogroups) developed Proto-IE language and carried it along during their journey from Central Asia to Europe. The earliest signs of the language in Anatolia were detected by linguists, and dated by 9400 - 9600 - 10,100 ybp, which coincides with the data of DNA genealogy that is described in this paper.

The arrival of R1a1 in Europe might be marked by known archaeological cultures in the Balkans and Central/Eastern Europe, dated back to 9000 - 7000 ybp. 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 Europe after 4,800 ybp (the Bell Beakers and other R1b1 migratory waves to Europe, including perhaps the Kurgan people, though their identification and haplogroup assignment remains unclear), haplogroup R1a1 had moved to the Russian Plain around 4800 - 4600 ybp. From there R1a1 migrated (or moved as military expeditions) to the south, east, and south-east as the historic Aryans. Dates for these movements are strikingly similar, and they span 4200 and 3600 ybp. As a result, in Anatolia and Mitanni, South Ural, Iran, India, and beyond the Ural Mountains, in South Siberia, in all those areas today's linguists find the same languages: the Aryan, or the Indo-European language, or the Iranian family of languages. They all have the same Aryan roots. They founded common horse breeding terminology and shared essentially the same vocabulary for household items, gods and religious terms, although sometimes twisted due to “human factor” as found in India and Iran.

Currently, most of those with European R1a1 live in Eastern Europe, primarily in Russia (up to 62% of the population) and Poland, Ukraine, Belarus (up to 55% of the population in the last three countries). In depth reports on their haplotype branches and distinct SNP (characteristic mutations in the DNA) will be explored in forthcoming publications.

Acknowledgements

The authors are indebted to Laurie Sutherland 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*, 1959. doi:10.1186/1471-2156-10-59
- Adamov, D., & Klyosov, A. A. (2008). Theoretical and practical evaluations of back mutations in haplotypes of Y chromosome. *Proceedings of the Russian Academy of DNA Genealogy*, *1*, 631-645.
- Anthony, D. W. (2007). *The horse, the wheel, and language*. Princeton: Princeton University Press.
- Barac, L., Pericic, M., Klaric, I. M., Janicijevic, B., Parik, J., Rootsi, S., & Rudan, P. (2003a). Y chromosome STRs in Croatians. *Forensic Science International*, *138*, 127-133. doi:10.1016/j.forsciint.2003.09.004
- Barac, L., Pericic, M., Klaric, I. M., Rootsi, S., Janicijevic, B., Kivisild, T., Parik, J., Rudan, I., Villems, R., & Rudan, P. (2003b). Y chro-

- mosomal heritage of Croatian population and its island isolates. *European Journal of Human Genetics*, 11, 535-542. doi:10.1038/sj.ejhg.5200992
- Bittles, A. H., Black, M. L., & Wang, W. (2007). Physical anthropology and ethnicity in Asia: the transition from anthropology to genome-based studies. *Journal of Physical Anthropology*, 26, 77-82. doi:10.2114/jpa2.26.77
- Burgarella, C. & Navascues, M. (2011). Mutation rate estimates for 110 Y-chromosome STRs combining population and father-son paired data. *European Journal of Human Genetics*, 19, 70-75. doi:10.1038/ejhg.2010.154
- Cadenas, A. M., Zhivotovsky, L. A., CavalliSforza, L. L., Underhill, P. A., & Herrera, R. J. (2008). Y-chromosome diversity characterizes the Gulf of Oman. *European Journal of Human Genetics*, 16, 374-386. doi:10.1038/sj.ejhg.5201934
- Dixon, R. M. W. (1997). *The rise and fall of language*. Cambridge: Cambridge University Press.
- Fornarino, S., Pala, M., Battaglia, V., Maranta, R., Achilli, A., Modiano, G., Torroni, A., et al. (2009). Mitochondrial and Y-chromosome diversity of the Tharus (Nepal): A reservoir of genetic variation. *BMC Evolutionary Biology*, 9, 154-170. doi:10.1186/1471-2148-9-154
- Gamkrelidze, T.V., & Ivanov, V.V. (1995). *Trends in linguistics 80: Indo-European and the Indo-Europeans*. Berlin: Mouton de Gruyter.
- Gimbutas, M. (1973) The beginning of the bronze age in Europe and the Indo-Europeans 3500-2500 B.C. *Journal of Indo-Europeans Studies*, 1, 163-214.
- Gimbutas, M. (1994) The civilization of the goddess. In: J. Marler (Ed.), *The End of Old Europe: the Intrusion of Steppe Pastoralists from N. Pontic and the Transformation of Europe*. San-Francisco: Harper.
- 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
- Järve, M., Zhivotovsky, L. A., Rootsi, S., Help, H., Rogaev, E. I., Khusnutdinova, E. K. et al. (2009) Decreased rate of evolution in Y chromosome STR loci of increased size of the repeat unit. *PLOS One*, 4, e7276. doi:10.1371/journal.pone.0007276
- Kang, L., Lu, Y., Wang, C., Hu, K., Chen, F., Liu, K. et al. (2011). Y-chromosome O3 haplogroup diversity in Sino-Tibetan populations reveals two migration routes into the Eastern Himalayas. *Annals of Human Genetics*, 76, 92-99.
- Kivisild, T., Rootsi, S., Metspalu, M., Mastana, S., Kaldma, K., Parik, J., Metspalu, E. et al. (2003). The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. *American Journal of Human Genetics*, 72, 313-332. doi:10.1086/346068
- Klyosov, A. A. (2008). Where Slavs and Indo-Europeans came from? *Proceedings of the Russian Academy of DNA Genealogy*, 1, 400-477.
- 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. (2009c). A comment on the paper: Extended Y chromosome haplotypes resolve multiple and unique lineages of the Jewish priesthood. *Human Genetics*, 126, 719-724. doi:10.1007/s00439-009-0739-1
- Klyosov, A. A. (2010a). Haplogroup R1a1 and its subclades in Asia. *Proceedings of the Russian Academy of DNA Genealogy*, 3, 1866-1896 (in Russian).
- Klyosov, A. A. (2010b). Ancient (non-Indo European haplotypes of haplogroup R1a1 in North-Western China. *Proceedings of the Russian Academy of DNA Genealogy*, 3, 925-941 (in Russian).
- Klyosov, A. A. (2011a). Haplotypes of R1b1a2-P312 and related subclades: Origin and "ages" of most recent common ancestors. *Proceedings of the Russian Academy of DNA Genealogy*, 4, 1127-1195.
- Klyosov, A. A. (2011b). Biological chemistry as a foundation of DNA genealogy: The emergence of "Molecular history". *Biokhimiya*, 76, 517-533. doi:10.1134/S0006297911050026
- Klyosov, A. A., & Rozhanskii, I. L. (2011). Re-examining the "out of Africa" theory and the origin of Europeoids (Caucasoids) in light of DNA genealogy. *Advances in Anthropology*, 1, 1 (in press).
- Mallory, J. P. (1989). *In search of the Indo-Europeans: Language, archaeology and myth*. London: Thames and Hudson.
- Msaïdie, S., Ducourneau, A., Boetsch, G., Longepied, G., Papa, K., Allibert, C. et al. (2011). Genetic diversity on the Comoros Islands shows early seafaring as major determinant of human biocultural evolution in the Western Indian Ocean. *EEuropean Journal of Human Genetics*, 19, 89-94. doi:10.1038/ejhg.2010.128
- Myres, N. M., Rootsi, S., Lin, A. A., Jarve, M., King, R. J., Kutuev, I. et al. (2010). A Major Y-chromosome haplogroup R1b Holocene era founder effect in Central and Western Europe. *European Journal of Human Genetics*, 19, 95-101.
- Pericic, M., Lauc, L. B., Klaric, A. M. et al. (2005). High-resolution phylogenetic analysis of southeastern Europe traces major episodes of paternal gene flow among Slavic populations. *Molecular Biology and Evolution*, 22, 1964-1975. doi:10.1093/molbev/msi185
- Renfrew, C. (2000). Time depth in historical linguistics. In: C. Renfrew, A. McMahon, & L. Trask (Eds.) (pp. 413-439). Cambridge: The McDonald Institute for Archaeological Research.
- Rozhanskii, I. L., & Klyosov, A. A. (2009). Haplogroup R1a: Haplotypes, genealogical lineages, history, geography. *Proceedings of the Russian Academy of DNA Genealogy*, 2, 974-1099 (in Russian).
- 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
- Sahoo, S., Singh, A., Himabindu, G., Banerjee, J., Sitalaximi, T., Gaikwad, S., Trivedi, R., Endicott, P., Kivisild, T., Metspalu, M., Villemers, R., & Kashyep, V. K. (2006). A prehistory of Indian Y chromosomes: Evaluating demic diffusion scenarios. *PNAS*, 103, 843-848. doi:10.1073/pnas.0507714103
- 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., Bhanwer, A. J. S., Tiwari, P. K., & Bamezai, R. N. K. (2009). The Indian origin of paternal haplogroup R1a1¹ substantiates the autochthonous origin of Brahmins and the caste system. *Human Genetics*, 54, 47-55. doi:10.1038/jhg.2008.2
- Shou, W. H., Qiao, E. F., Wei, C. Y., Dong, Y. L., Tan, S. J., Shi, H. et al. (2010). Y-chromosome distributions among populations in Northwest China identify significant contribution from Central Asian pastoralists and lesser influence of western Eurasians. *Human Genetics*, 23, (Advance Online Publication).
- Thanseem, I., Thangaraj, K., Chaubey, G., Singh, V. K., Bhaskar, L. V., Reddy, M. B., Reddy, A. G., & Singh, L. (2006). Genetic affinities among the lower castes and tribal groups of India: Inference from Y chromosome and mitochondrial DNA. *BMC Genetics*, 7, 42-53. doi:10.1186/1471-2156-7-42
- Underhill, P. A., Myres, N. M., Rootsi, S., Metspalu, M., Zhivotovsky, M. A., King, R. J. et al. (2010). Separating the post-Glacial coancestry of European and Asian Y chromosomes within haplogroup R1a. *European Journal of Human Genetics*, 18, 479-484.
- Zhong, H., Shi, H., Qi, X.-B., Duan, Z.-Y., Tan, P.-P., Jin, L. et al. (2010). Extended Y-chromosome investigation suggests postGlacial migrations of modern humans into East Asia via the northern route. *Molecular Biology and Evolution*, 28, 717-727.

APPENDIX

The following DNA projects were selected as primary haplotype databases:

<http://www.familytreedna.com/public/sharifs/default.aspx?section=yresults>

<http://www.familytreedna.com/public/Arab%20Tribes/default.aspx?vgroup=Arab+Tribes§ion=yresults>

<http://www.familytreedna.com/public/R1aY-Haplogroup/default.aspx?vgroup=R1aY-Haplogroup§ion=yresults>

<http://www.familytreedna.com/public/R1a/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>

Reference data were selected according to SNP assignment from YSearch database:

(<http://www.ysearch.org>)

and public projects of FTDNA

(<http://www.familytreedna.com>)