

# DNA Genealogy, Mutation Rates, and Some Historical Evidence Written in the Y-Chromosome, Part II: Walking the Map

Anatole Klyosov

## Abstract

Employing the methodology developed in Part I of this two-part series of articles, haplotypes of the many different populations were analyzed for their time spans to the (most recent) common ancestor (TSCA). These populations included Haplogroup R1b1b2 populations from: Ireland (2), Britain, Sweden, Flanders, Central Europe, and Iberia; Haplogroup I1 populations from Scandinavia, Isles (Britain-Ireland-Scotland), Central Europe, Eastern Europe, and Middle East, Haplogroup R1a1 populations from England, Ireland, Scotland, Germany, Norway, Sweden, Poland, Czech and Slovak Republics, Russia, Armenia, Anatolia, Balkans, India (2), Arabia, and 17 remaining countries lumped; Haplogroup J1 populations Cohanim and Arab; Haplogroup J2 populations Cohanim and non-Cohanim Jewish; Haplogroup H1 populations Bulgarian Roma and Croatian Roma; Haplogroup C2 population from Polynesia; and Haplogroup Q population Native Americans.

## Introduction

This second part of a two-part series of articles applies the methods developed in Part 1 to many different world populations. While some of the results presented here are supported by other studies, many of the results are quite different from those found in previous studies, or, in cases where only theories of origin have been proposed, from popular expectations. We show here that our methods of calculating the TMRCA for a group of haplotypes, leads to a need for a reevaluation of current theories of the populating of Europe and other areas.

The data presented below will show several major conclusions:

(a) The male Basques living today have rather recent roots of less than four thousand years ago, contrary to the legend that proposes they lived some 30 thousand years ago.

(b) There is no justification in the results of a "Ukrainian refuge" for the R1a1 ancient population allegedly 15,000 years ago; instead, evidence has been obtained that the oldest R1a1 lived circa 20,000 years before the present (ybp) in South Siberia. There are two sets of data and these provide ages of  $21,000 \pm 3,000$  ybp and  $19,625 \pm 2,800$  ybp, calculated by two different methods, and  $11,650 \pm 1,550$  years ago appeared in the Balkans (Serbia, Kosovo, Bosnia, Macedonia).

(c) Except the South Siberian and Balkans populations, present-day bearers of R1a1 across Western and Eastern Europe have common ancestors who lived between 3550 and 4750 years ago (the "youngest" in Scotland, Ireland and Sweden, the "oldest" in Russia ( $4750 \pm 500$  ybp) and Germany ( $4,700 \pm 520$  ybp),

(d) There are two different groups of Indian R1a1 haplotypes; one shows a good match with the Russian Slavic R1a1 group, having a common ancestor several hundred years "younger" than the Russian R1a common ancestor ( $4,050 \pm 500$  vs.  $4,750 \pm 500$  ybp). This supports the idea that a proto-Slavic migration to India as Aryans occurred (mentioned in classic ancient Indian literature) around 3600 ybp. The other Indian R1a population is significantly older, with a common ancestor living  $7,125 \pm 950$  ybp; they could have migrated from South Siberia to South India.

---

Address for correspondence: Anatole Klyosov, [aklyosov@comcast.net](mailto:aklyosov@comcast.net)

Received: January 8, 2009; accepted: August 1, 2009

(e) South India Chenchu R1a1 match the current Russian Slavic R1a1 haplotypes, and the Chenchu R1a common ancestor appeared some  $3200 \pm 1900$  ybp, apparently after the R1a1 migration from the North to India. Another Chenchu R1a1 lineage originated about  $350 \pm 350$  ybp, around the 17th century CE.

(f) The so-called Cohen Modal Haplotype in Haplogroup J1 originated  $9,000 \pm 1,400$  years ago, if all related J1 haplotypes are considered. About  $4,000 \pm 520$  ybp it appeared in the proto-Jewish population, and  $1,050 \pm 190$  years ago (if to consider only CMH) or  $1,400 \pm 260$  years ago (if to consider only Jewish J1 population) split a "recent CMH" lineage.

(g) another so-called CMH, of Haplogroup J2, appeared in the Jewish population  $1,375 \pm 300$  years ago,

(h) The South African Lemba population of Haplogroup J has nothing to do with ancient Jewish patriarchs, since the haplogroup appears to have penetrated the Lemba population some  $625 \pm 200$  ybp, around the 14th century CE.

(i) Native American Haplogroup Q1a3a contains at least six lineages, the oldest of which originated  $16,000 \pm 3,300$  years ago, in accord with archaeological data.

## Methods

The methods used in the present study are described in the companion article.

## Results and Discussion: R1b Populations

### *The Basques*

The set of Basque R1b haplotypes was already discussed as an example in the companion article with the following results: The common ancestor of the Basques lived  $3,625 \pm 370$  years ago, and had the following haplotype

13-24-14-11-11-14-12-12-12-13-13-29-17-9-10-11-11-25-14-18-29-15-15-17-17

### *Iberia*

Since the 750 19-marker Iberian R1b1 haplotypes contain 16 identical, base haplotypes, the time span to their common ancestor can also be calculated as  $\ln(750/16)/0.0285 = 135$  generations (without a correction for back mutations), or 156 generations with the correction (see Table 2 in the companion paper) to the common ancestor, that is 3900 ybp (by the "logarithmic" method). It is only 7.6% higher than the above value, obtained by mutations count (by the "linear" method), and fits well into the standard error of the calculation.

It is remarkable that the Basque ancestral 25-marker haplotype is practically identical to the 25-marker ancestral haplotypes of R1b-U152 and R1b1b2-M269 subclades

13-24-14-11-11-14-12-12-12-13-13-29-17-9-10-11-11-25-15-19-29-15-15-17-17

with the respective common ancestors having lived  $4375 \pm 450$  and  $4,450 \pm 460$  years before present (Klyosov, 2008a) with estimated error being shown as the 95% confidence interval.

As was considered in detail in the preceding paper (Part I), these relatively tight confidence intervals result from the large number of alleles involved in each of the considered haplotype series, such as 750 of the 19-marker Iberian R1b1 haplotype series (14,250 alleles), 184 of the 25-marker U152 haplotypes (4,600 alleles), and 197 of the 25-marker M269 haplotypes (4,925 alleles). The standard errors of the measurements—for the average number of mutations per marker in said haplotype series—resulted in 95% confidence intervals of  $\pm 2.00\%$ ,  $2.84\%$ , and  $2.73\%$ , respectively, while the corresponding values for the mutation rates for the employed 12-, 19- and 25-marker haplotypes were equal to 10%, as it was discussed in Part I. Naturally, for smaller haplotype series the standard errors and standard deviations are significantly higher, as shown below.

There are only two differences in alleles (in bold) of the Basque (Iberian) and the M269-U152 base haplotypes (in DYS 437 and 448), which are 14-18 in the Basque base haplotype and 15-19 in the latter. These mutations are quite insignificant because the corresponding mean values in the Basque haplotypes are 14.53 and 18.35, respectively.

The data show that all three populations, including the Basques, are likely to be descendants from the same common ancestor of the R1b1b2-M269 haplogroup. The principal conclusion is that the male Basques living today have rather recent roots of less than four thousand years, contrary to legend that proposes they lived some 30,000 years ago. Despite the ancient language, it is very likely that the present day Basques represent a rather recent Iberian population, in terms of DNA genealogy. It is very unlikely that their ancestors had encountered Neanderthals in Europe or had been associated with the Aurignacian culture (34,000-23,000 ybp), nor did they make sophisticated cave paintings in South of France, Spain, and Portugal. Arguably, the Basque ancient and unique language was brought to Iberia around 3600 ybp by the M269 bearers from their place of preceding location(s) and/or their origin, presumably in Asia. The origin of R1b1b2, however, is beyond the scope of this study, and will be discussed in more detail elsewhere.

*Ireland R1b1 and I2 haplotypes*

A list of 243 of 19-marker Irish haplotypes encompassing 35 surnames with origins in the province of Munster was published in (McEvoy et al, 2008). The listed haplotypes were not identified by the authors in terms of haplogroups. However, when a haplotype tree was composed, as shown in Figure 1, it became obvious that it included quite a distinct branch of 25 haplotypes of a different origin, which clearly descended from a different common ancestor from apparently a different haplogroup. Indeed, those 25 haplotypes had the following base haplotype (in the format DYS 19-388-389-389-390-391-392-393-434-435-436-437-438-439-460-461-462-385a-385b, employed by Adams et al, 2008):

15-13-13-16-23-10-11-13-11-11-12-15-10-11-10-11-12-12-15

It was identified as a member of Haplogroup I2. For example, the Iberian base I2 haplotype, deduced from an extended haplotype list in (Adams et al, 2008), is as follows:

15-13-13-16-23-10-11-13-11-11-12-15-10-11-10-12-12-13-15

It differs by only one mutation on two markers (shown in bold), and its assumed common ancestor lived apparently  $12,800 \pm 2,600$  ybp (Klyosov, 2009a). The 25 Irish I2 haplotypes contain 199 mutations, which bring their common ancestor to  $9,600 \pm 1,200$  ybp. This is evidently a combination of subclades I2a1 ( $5600 \pm 620$  ybp), I2a2 ( $6250 \pm 800$  and  $2275 \pm 380$  ybp for two separate branches), I2b1 ( $5700 \pm 590$  ybp) and I2b2 ( $5000 \pm 630$  ybp), with timespans to the respective common ancestors shown in parentheses (Klyosov, to be published). Nev-

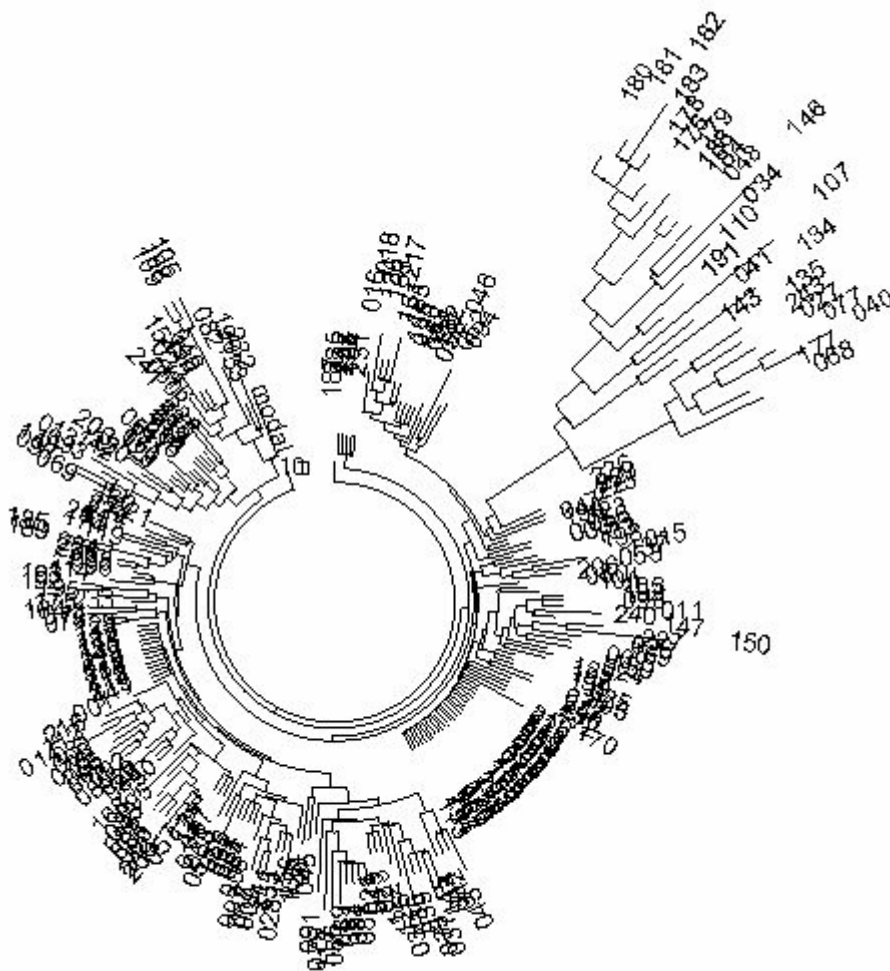


Figure 1. A tree of 243 Irish 19-marker haplotypes (McEvoy et al, 2008). The distinct branch of the right belongs to Haplogroup I2.

ertheless, this apparent dating shows an upper limit of the respective timespan, and depends on the fraction of the subclades that are represented in the dataset.

Since a difference in two mutations in two 19-marker haplotypes approximately corresponds to a time difference between them of about  $1900 \pm 1300$  years, it brings the two ancestral I2 haplotypes (in the Iberia and on the Isles) into a rather close proximity in time. Besides, the ancient I2 haplotypes in Ireland and Iberia resemble the ancient I2 base haplotype in Eastern Europe (Poland, Ukraine, Belarus, Russia) which in said 19-marker format is as follows:

15-13-13-16-23-10-12-13-X-X-X-14-X-11-X-X-X-15-15

and its common ancestor lived  $10,800 \pm 1,170$  ybp (Klyosov, 2009a). The same comment regarding a combination of I2a and I2b subclades (see above) is applicable here as well.

When the 25 ancient I2 haplotypes were removed from the tree of the Irish haplotypes, the presumably R1b1 218-haplotype tree became as shown in Figure 2. Its base haplotype

14-12-13-16-24-11-13-13-11-11-12-15-12-12-11-12-11-11-14

turned out to be identical with the Iberian R1b1 ancestral haplotype (see the preceding paper, Part I), and with the Atlantic Modal Haplotype, shown in the same format:

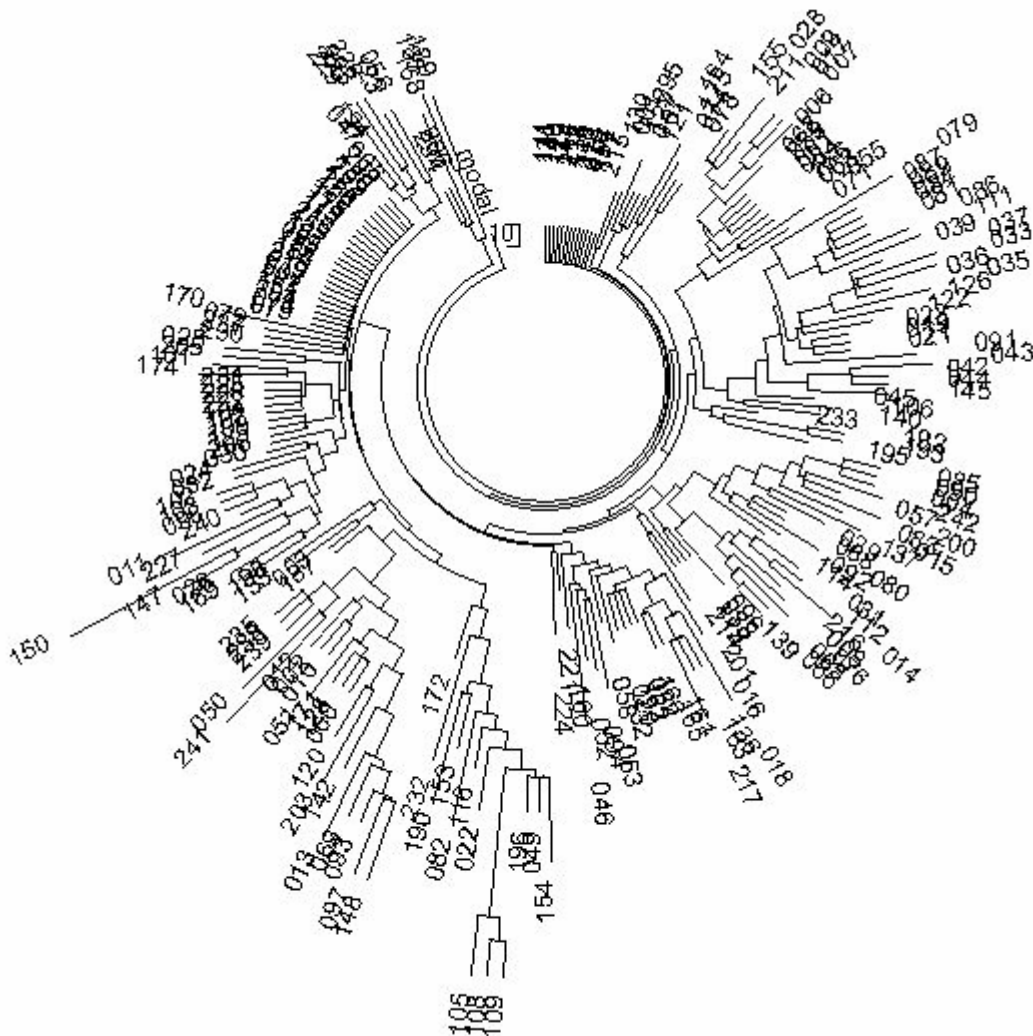


Figure 2. A tree of 218 Irish 19-marker haplotypes (McEvoy et al, 2008), probably all of Haplogroup R1b1.

14-12-13-16-24-11-13-13- X- X- Y- 15-12-12-11- X- X- 11-14

Here X replaces the alleles which are not part of the 67-marker FTDNA format, and Y stands for DYS436 which is not specified for the AMH. The same haplotype is the base one for the subclade U152 (R1b1c10), with a common ancestor of  $4,125 \pm 450$  ybp, P312 ( $3,950 \pm 400$  ybp), L21 ( $3,600 \pm 370$  ybp) and for R1b1b2 (M269) haplogroup (with subclades) with a common ancestor of  $4,375 \pm 450$  ybp (Klyosov, 2008a, and to be published), and only one mutation away (DYS390=23) for R1b-U106 ( $4,175 \pm 430$  ybp). Hence, they all are likely to have a rather recent origin, where the word "recent" is used in comparison to some widespread expectations in the literature of some 30,000 years to a common ancestor.

All the 218 Irish R1b1 haplotypes have 731 mutations in their 4,142 alleles, which brings their common ancestor to  $3350 \pm 360$  ybp. The average number of mutations per marker is slightly lower in the Irish R1b1 haplotypes ( $0.176 \pm 0.007$ ) compared to the Iberian ones ( $0.196 \pm 0.004$ ); the mutation rate was the same in the calculations of the 19-marker haplotypes. It appears that the M269 bearers had come to both Iberia and Ireland fairly late compared with the time of their inhabiting the continent or elsewhere, where the common ancestor of R1b1b2 subclades lived between  $3600 \pm 370$  and  $4175 \pm 430$  ybp (see above).

Since this dataset (McEvoy et al, 2008) was published without assigning of haplotypes to haplogroups, one might think that once the I2 haplotypes were separated from the series, the set would contain some other extraneous haplotypes from other haplogroups besides R1b1, and the dating would be distorted. However, it is very unlikely. First, the tree (Figure 2) does not contain obviously "foreign" branches, which would otherwise produce a very distinct "protuberances." Second, the value on the marker DYS392 in the remaining haplotypes was completely characteristic of Haplogroup R1b (predominantly a value of 13, with a few 14s and one 12). Third, the young "age" of  $3350 \pm 360$  ybp for this Irish dataset, which is even younger than that of a number of R1b1b2 subclades ( $3,600 \pm 370$  to  $4175 \pm 430$  ybp) shows that there were no any noticeable admixtures of haplogroups other than R1b, which would sharply increase the "age" of the dataset.

It might appear that the common ancestor(s) of this group of Irish individuals had arrived to the Isles somewhat late compared to his relatives in Iberia. However, we know that the asymmetry of mutations can affect the observed number of mutations per marker, and if the Iberian R1b1 haplotype series is more asymmetrical in terms of mutations, compared to the Irish one, it might explain the difference.

However, it is not so. The Iberian haplotypes are practically symmetrical (the degree of mutations is 0.56), and from  $0.196 \pm 0.004$  (the observed number of mutations per marker, without the corrections) the corrections brought the result to  $0.218 \pm 0.004$  (corrected for back mutations) and further to  $0.217 \pm 0.004$  (corrected for asymmetry of mutations). The degree of asymmetry of the Irish 188 haplotypes was 0.54, so the respective results will be  $0.176 \pm 0.007$ ,  $0.192 \pm 0.007$  and  $0.190 \pm 0.007$ . Hence, the Irish R1b1 haplotypes might still appear to be a little "younger" compared to the Iberian R1b1 haplotypes.

The pattern of the Irish R1b1 haplotypes, however, is a bit more complicated, since the above series of 218 haplotypes also contains two different base haplotypes: one, in which DYS439=11 (shown in bold), and there are 13 of such base haplotypes in the whole haplotype series

14-12-13-16-24-11-13-13-11-11-12-15-12-11-11-12-11-11-14

and another, in which DYS391=10 and DYS385b=15 (shown in bold), and there are 24 of such base haplotypes among the total of 218

14-12-13-16-24-10-13-13-11-11-12-15-12-12-11-12-11-11-15

The last one is obviously a "young" lineage, since the 24 base haplotypes are sitting on the 98-haplotype branch (on the left-hand side in Figure 2), which gives an estimate of the TSCA for this branch of  $\ln(98/24)/0.0285 = 49$  generations (without the correction) and 52 generations with the correction for back mutations, that is around 1300 ybp. The "older" 13 base haplotypes in the total amount of 218 haplotypes would give  $2750 \pm 290$  years to the common ancestor.

This dating actually matches fairly well the TSCA for 1242 haplotypes from the British Isles, for which 262 haplotypes are identical to each other, hence, the base haplotypes in the FTDNA format (X and Y stand for not typed DYS385a,b):

13-24-14-X-Y-12-12-12-13-13-29

The fraction of the base haplotype gives  $\ln(1242/262)/0.0179 = 87$  generations (without correction) or 96 generations with the correction, that is  $2400 \pm 250$  years to the common ancestor. This population of R1b1 is certainly "younger" compared with the Iberian R1b1 series of haplotypes.

Which of the base (ancestral) R1b1 haplotypes were "younger" in terms of the TSCA, the Irish or the Iberian, was further examined using a larger series of 983 Irish

R1b1 haplotypes, published in (McEvoy and Bradley, 2006). Their base haplotypes was as follows:

14-12-13-16-24-11-13-13-9-11-12-15-12-12-11-10-11-11-14

In all 983 haplotypes 966 had the allele of "9" (98% of total) on DYS434, while in the preceding series of 218 Irish haplotypes 216 of them had "11" (99% of total) on the same very locus. It seems that the authors simply changed their notation of haplotypes. We have the same situation with DYS461, where 189 out of 218 had the 12 allele (87%), while in the larger series it was 10 in 833 of 943 haplotypes (88%). It appears that these haplotypes are in fact identical to each other.

All the 983 R1b1 haplotypes have 3706 mutations from the base haplotype, that is  $0.198 \pm 0.003$  for the average number of mutations per marker. This is practically identical with  $0.196 \pm 0.004$  for the Iberian R1b1 haplotypes. The degree of asymmetry for the larger series is

exactly the same (0.54) as in the 218-haplotype series, and cannot shift the number of mutations per marker, hence, the "age" of the common ancestor stays at  $3,800 \pm 380$  ybp. Therefore, the Irish and the Iberian R1b1 haplotypes ( $3,625 \pm 370$  ybp) have practically the same common ancestor, from the viewpoint of DNA genealogy.

It is of interest to compare them to a Central European series, such as the Flemish R1b 12-marker 64-haplotype series (Mertens, 2007). In that case a non-FTDNA format of 12 markers was employed, in which DYS426 and DYS388 were replaced with DYS437 and DYS438. The average mutation rate for the format was calculated and shown in Table 1 in the preceding paper. All 64 haplotypes have the following base haplotype (the first 10 markers in the format of the FTDNA, plus DYS437 and DYS438):

13-24-14-11-11-14-X-Y-12-13-13-29-15-12

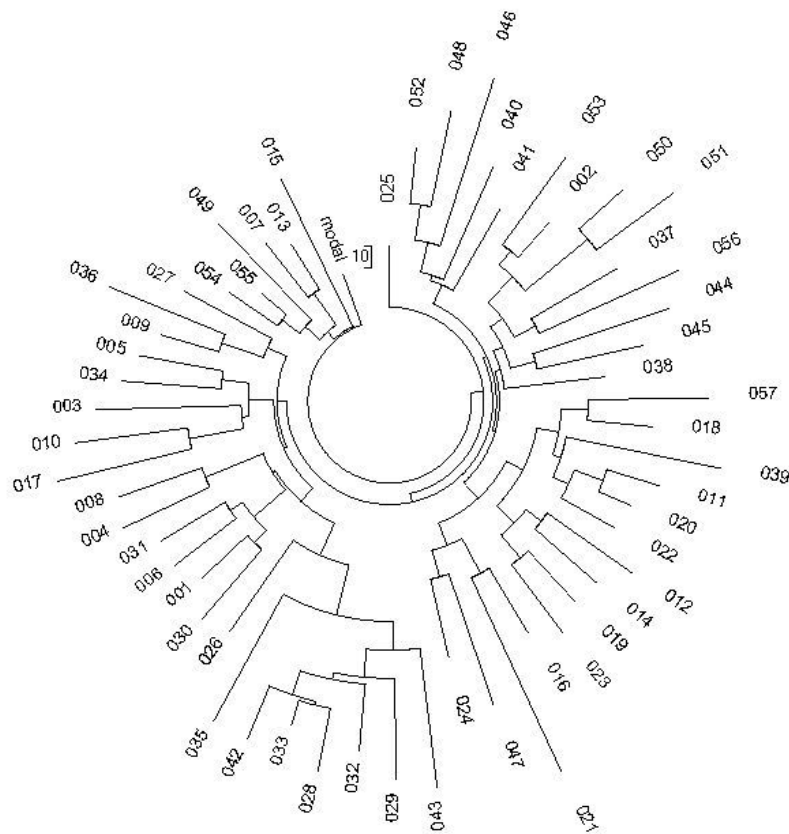


Figure 3. The 25-marker haplotype tree for England, Haplogroup R1a1. The 57-haplotype tree was composed from data of YSearch database. A seven-haplotype branch at the bottom (between 035 and 043) plus haplotypes 001, 006 and 030 is a family of haplotypes with DYS388=10 (all other mostly have DYS388=12, in one case DYS388=14, haplotype 031).

All 64 haplotypes contained 215 mutations, which results in  $4150 \pm 500$  years to the common ancestor.

All those principal ancestral (base) haplotypes, as well as the Swedish series (Karlsson et al, 2006) of 76 of the 9-marker R1b1b2 haplotypes with the base

13-24-14-11-11-14-X-Y-Z-13-13-29

fit to the Atlantic Modal Haplotype. All the 76 haplotypes included seven base haplotypes and 187 mutations from it. It gives  $\ln(76/7)/0.017 = 140$  generations (without the correction) or 163 (with the correction), when the logarithmic method is employed, and  $187/76/9/0.00189 = 145$  generations (without the correction) or 169 (with the correction for back mutations), which gives  $4225 \pm 520$  years to the common ancestor.

The TSCA values for the various R1b populations and the data from which they were derived are shown in Table 1. The "ages" of the Irish ( $3350 \pm 360$  ybp and  $3800 \pm 380$  ybp), Iberian ( $3625 \pm 370$  ybp), Flemish

( $4150 \pm 500$  ybp) and Swedish ( $4225 \pm 520$  ybp) populations differ insignificantly from each other in terms of their standard deviations (all within the 95% confidence interval). However, it still can provide food for thought about history of the European R1b1b2 population.

## Results and Discussion: I1 Populations

### *The Isles (England, Ireland and Scotland)*

These haplotypes were briefly considered in the preceding paper (Part I) as an example for calculating the TSCA for 1527 of 25-marker haplotypes, taking into account the effect of back mutations and the degree of asymmetry of mutations.

These 1527 haplotypes included 857 English haplotypes, 366 Irish haplotypes and 304 Scottish haplotypes. All of them turned out to be strikingly similar, and very likely descended from the same common ancestor, who had the following haplotype:

**Table 1**  
Time-Span to the Common Ancestor for R1b Populations

Population	N	Haplotype Size (mkrs)	TSCA (years)	Source of Haplotypes
Basque	17	25	$3625 \pm 370$	Basque Project, see <b>Web Resources</b>
Basque	44	12	$3500 \pm 735$	Basque Project, see <b>Web Resources</b>
Iberian	750	19	$3625 \pm 370$	Adams et al. (2008)
Ireland--First Series (from Munster)	218	19	$3800 \pm 380$	McEvoy (2008)
Ireland--Second Series	983	19	$3800 \pm 380$	McEvoy (2006)
Ireland--"Younger" (from Munster)	98	19	$2750 \pm 290$	McEvoy (2008)
British	1242	9	$2400 \pm 250$	Campbell (2007)
Flemish	64	12	$4150 \pm 500$	Mertens (2007)
Swedish	76	9	$4225 \pm 520$	Karlsson (2006)
R-U106	284	25	$4175 \pm 430$	Weston (2009), see <b>Web Resources</b>
R-U152	184	25	$4125 \pm 450$	Kerchner (2008), see <b>Web Resources</b>
R-P312	464	25, 67	$3950 \pm 400$	Stevens (2009a), see <b>Web Resources</b>
R-L21	509	25, 67	$3600 \pm 370$	Sreves (2009b), see <b>Web Resources</b>
R-L23*	22	25, 67	$5475 \pm 680$	Ysearch, see <b>Web Resources</b>
R-M222	269	25, 67	$1450 \pm 150$	Wilson (2009), see <b>Web Resources</b>
R-M269 (with all subclades)	197	25, 67	$4375 \pm 450$	

Note: Where the number of markers is shown as "25, 67," it means that the calculations were done with 25 markers, but a diagram was constructed with 67 markers to look for possible problematic branches,

13-22-14-10-13-14-11-14-11-12-11-28-15-8-9-8-11-  
23-16-20-28-12-14-15-16

All 1527 haplotypes contained 8785 mutations from the above base haplotype, which gives the “observed” value of  $0.230 \pm 0.002$  mutations per marker in the 95% confidence interval. Since the degree of asymmetry of the haplotype is 0.65, the corrected (for back mutations and the asymmetry of mutations) value is equal to  $0.255 \pm 0.003$  mutations/marker, which results in  $139 \pm 14$  generations, that is  $3475 \pm 350$  years to the common ancestor at the 95% confidence level.

The TSCA values for the English, Irish and Scottish 25-marker haplotypes, calculated separately (the degree of asymmetry for each series were equal to 0.66, 0.64 and 0.64, respectively), were  $136 \pm 14$ ,  $151 \pm 16$  and  $131 \pm 15$  generations, that is  $3400 \pm 350$ ,  $3775 \pm 400$ , and  $3275 \pm 375$  ybp. Indeed, their averaged value equals to  $139 \pm 10$  generations, and the obtained dispersion shows that the calculated standard deviations (based on  $\pm 10\%$  as the 95% confidence interval) are reasonable. Hence, the time span to the common ancestor of  $3475 \pm 350$  years is a reliable estimate for more than 1500 English, Irish and Scottish individuals.

The modal haplotype for the sub-clade of Haplogroup I1 defined by P109 (presently named I1d1 according to ISOGG-2009), is different from that of Isles I1 on two markers out of 25 (those two shown in bold):

13-23-14-10-14-14-11-14-11-12-11-28-15-8-9-8-11-  
23-16-20-28-12-14-15-16

A set of I1-P109 haplotypes has a TSCA of  $2275 \pm 330$  years (Klyosov, to be published).

#### *Other I1 Populations*

The value for TSCA seen in the populations discussed above are quite typical of other European I1 populations. For the Northwest European/Scandinavian (Denmark, Sweden, Norway, Finland) combined series of haplotypes the TSCA equals to  $3375 \pm 345$  ybp. For the Central and South Europe (Austria, Belgium, Netherlands, France, Italy, Switzerland, Spain) it equals to  $3425 \pm 350$  ybp, for Germany  $3225 \pm 330$  ybp, for the East European countries (Poland, Ukraine, Belarus, Russia, Lithuania) it is equal to  $3225 \pm 360$  ybp (to be published). It is of interest that even Middle Eastern I1 haplotypes (Jordan, Lebanon, and presumably Jewish ones) descend from a common ancestor who lived at about the same time,  $3475 \pm 480$  ybp (Klyosov, to be published).

The ASD methods gave a noticeably higher values for the TSCAs for the English, Irish, Scottish, and the pooled haplotype series. For the first three series of 25-marker

haplotypes, calculated separately, the TSCAs were 158, 175 and 155 generations, respectively, with their averaged value equal to  $162 \pm 11$  generations, which can be compared to  $139 \pm 10$  generations, calculated by the “linear” method (corrected for back-mutations, see companion article), corrected for back mutations. As was pointed out, the ASD method typically overestimates the TSCA by 15-20% due to double and triple mutations, and some almost unavoidable extraneous haplotypes, to which the ASD method is rather sensitive. In this case of the Isle I1 haplotypes the overestimation of the ASD-derived TSCA was a typical 17% compared with the “linear” method (corrected for back-mutations).

#### **Results and Discussion: Eurasian R1a1 Populations**

The “mapping” of the enormous territory from the Atlantic through Russia and India to the Pacific, and from Scandinavia to the Arabian Peninsula, reveals that Haplogroup R1a men are marked with practically the same ancestral haplotype, which is about 4,500 - 4,700 years “old.” through much of its geographic range. Exceptions in Europe are found only in the Balkans (Serbia, Kosovo, Macedonia, Bosnia), where the common ancestor is significantly more ancient, about  $11,650 \pm 1,550$  ybp, and in the Irish, Scottish and Swedish R1a1 populations, which have a significantly “younger” common ancestor, some thousands years “younger” compared with the Russian, the German, and the Poland R1a1 populations. These geographic patterns will be explored below in this section. The haplogroup name R1a1 is used here to mean Haplogroup R-M17, because that is the meaning in nearly all of the referenced articles.

The entire map of base (ancestral) haplotypes and their mutations, as well as “ages” of common ancestors of R1a1 haplotypes in Europe, Asia, and the Middle East show that approximately six thousand years ago bearers of R1a1 haplogroup started to migrate from the Balkans in all directions, spreading their haplotypes. A recent excavation of 4,600 year-old R1a1 haplotypes (Haak et al., 2008) revealed their almost exact match to present-day R1a1 haplotypes, as it is shown below.

Significantly, the oldest R1a population appears to be in Southern Siberia.

#### *England and Ireland R1a1 haplotypes*

The 57 R1a 25-marker haplotype series of English origin (YSearch database) contains ten haplotypes that belong to a DYS388=10 series and was analyzed separately. The remaining 47 haplotypes contain 304 mutations compared to the base haplotype shown below, which corresponds to  $4,125 \pm 475$  years to a common ancestor in the 95% confidence interval. The respective haplotype tree is shown in Figure 3.



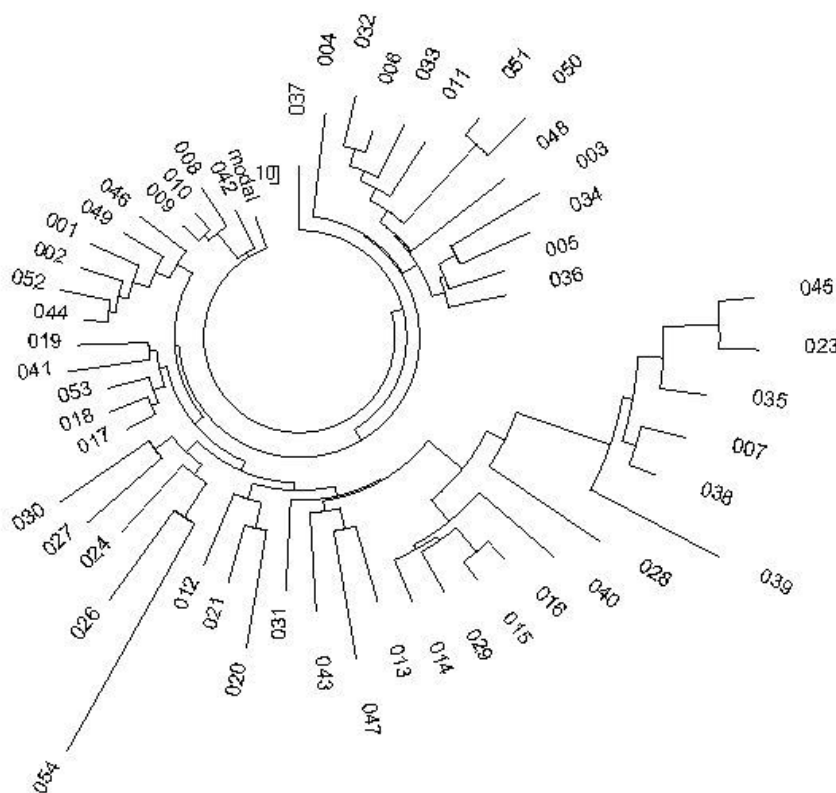


Figure 4. The 25-marker haplotype tree of Haplogroup R1a1 for Ireland. The 52-haplotype tree was composed from the YSearch database. A twelve-haplotype branch at the bottom left (between 014 and 045) is a family of haplotypes with DYS388=10 (all others primarily have DYS388=12, in two cases DYS388=14, haplotypes 003 and 034).

The 52 haplotype series of Ireland origin, all with 25 markers (YSearch database) contains 12 haplotypes which belong to a DYS388=10 distinct series as shown in Figure 4, and was analyzed separately. The remaining 40 haplotypes contain 244 mutations compared to the base haplotype, shown below, which corresponds to  $3,850 \pm 460$  years to a common ancestor.

Thus, R1a1 haplotypes sampled on the British Isles point to English and Irish common ancestors who lived  $4,125 \pm 475$  and  $3,850 \pm 460$  years ago. The English base (ancestral) haplotype is 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

and the Irish one:

13-25-15-11-11-14-12-12-10-13-11-30-15-9-10-11-11-  
23-14-20-32-12-15-15-16

An apparent difference in two alleles between the British and Irish ancestral haplotypes is in fact fairly insignificant, since the respective average alleles are equal to 10.51 and 10.73, and 23.98 and 23.55, respectively. Hence, their ancestral haplotypes are practically the same, within approximately one mutational difference.

#### *A DYS388=10 Subfamily of North-Western European R1a1 Haplotypes*

About 20% of both English and Irish R1a haplotypes have a mutated allele in eighth position in the FTDNA format (DYS388=12  $\rightarrow$  10), with a common ancestor of that population who lived  $3,575 \pm 450$  years ago (172 mutations in the 30 25-marker haplotypes with DYS388=10). 61 of these haplotypes were pooled from a number of European populations (Figure 5), and the tree splits into a relatively younger branch on the left, and the "older" branch on the lower right-hand side.

This DYS388=10 mutation is observed in northern and western Europe, mainly in England, Ireland, Norway, and to a much lesser degree in Sweden, Denmark, Netherlands and Germany. In areas further east and south that mutation is practically absent.

31 haplotypes on the left-hand side and on the top of the tree (Figure 5) contain collectively 86 mutations from the base haplotype

13-25-15-10-11-14-12-10-10-13-11-30-15-9-10-11-11-25-14-19-32-12-14-14-17

which corresponds to  $1625 \pm 240$  years to the common ancestor. It is a rather recent common ancestor, who lived around the 4<sup>th</sup> century CE. The closeness of the branch to the trunk of the tree in Figure 5 also points to the rather recent origin of the lineage.

The older 30-haplotype branch in the tree provides with the following base DYS388=10 haplotype:

13-25-16-10-11-14-12-10-10-13-11-30-15-9-10-11-11-24-14-19-32-12-14-15-16

These 30 haplotypes contain 172 mutations, for which the linear method gives  $3575 \pm 450$  years to the common ancestor.

These two DYS388=10 base haplotypes differ from each other by less than four mutations, which brings their common ancestor to about 3500 ybp. It is very likely that it is the same common ancestor as that of the right-hand branch in Figure 5.

The upper, “older” base haplotype differs by six mutations on average from the DYS388=12 base haplotypes from the same area (see the English and Irish R1a1 base haplotypes, above). This brings *their* common ancestor to about  $5,700 \pm 600$  ybp.

This common ancestor of both the DYS388=12 and DYS388=10 populations lived presumably in the Balkans (see below) since the Balkan population is the only European R1a1 population that is that old, for almost two thousand years before bearers of that mutation arrived to northern and western Europe some 4000 ybp (DYS388=12) and about 3600 ybp (DYS388=10). This mutation has continued to be passed down through the generations to the present time.

#### Scotland R1a1 Haplotypes

A set of 29 R1a1 25-marker haplotypes from Scotland have the same ancestral haplotype as those in England and Ireland:

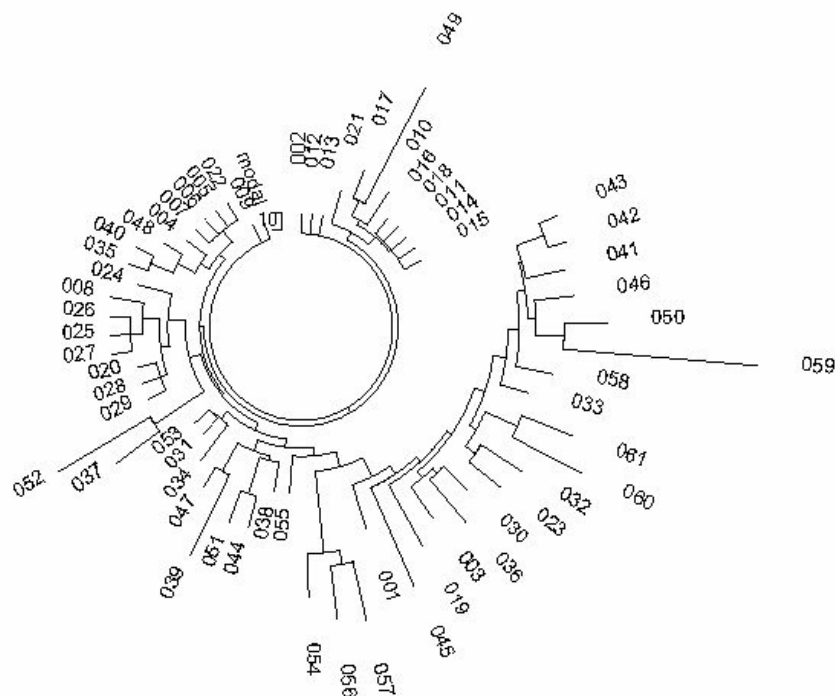


Figure 5. The 25-marker haplotype tree for 61 North West-European R1a1 haplotypes with DYS388 = 10. The haplotypes were collected from YSearch database.

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

This set of haplotypes contained 164 mutations, which gives  $3550 \pm 450$  years to the common ancestor.

#### *Germany R1a1 Haplotypes*

A 67-haplotype series with 25 markers from Germany revealed the following ancestral haplotype:

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

There is an apparent mutation in the third allele from the left (in bold) compared with the Isles ancestral R1a1 haplotypes, however, the average value equals to 15.48 for England, 15.33 for Ireland, 15.26 for Scotland, and 15.84 for Germany, so there is only a small difference between these populations. The 67 haplotypes contain 488 mutations, or  $0.291 \pm 0.013$  mutations per marker on average. This value corresponds to  $4,700 \pm 520$  years from a common ancestor in the German territory.

These results are supported by the very recent data found during excavation of remains from several males near Eulau, Germany. Tissue samples from one of them was derived for the SNP, SRY10831.2 (Haak et al., 2008), so it was from Haplogroup R1a1 and probably R1a1a as well (the SNPs defining R1a1a were not tested). The skeletons were dated to 4,600 ybp using strontium isotope analysis. The 4600 year old remains yielded some DNA and a few STRs were detected, yielding the following partial haplotype:

13(14)-25-16-11-11-14-X-Y-10-13-Z-30-15

These haplotypes very closely resemble the above R1a1 ancestral haplotype in Germany both in the structure and in the dating ( $4,700 \pm 520$  and  $4,600$  ybp).

#### *Norway and Sweden R1a1 Haplotypes*

The ancestral R1a1 haplotypes for the Norwegian and Swedish groups are almost the same, and both are similar to the German 25-marker ancestral haplotype. Their 16- and 19-haplotype sets (after DYS388=10 haplotypes were removed, five and one, respectively) contain on average  $0.218 \pm 0.023$  and  $0.242 \pm 0.023$  mutations per marker, respectively, which gives a result of  $3,375 \pm 490$  and  $3,825 \pm 520$  years back to their common ancestors. This is likely the same time span within the error margin.

#### *Polish, Czech, and Slovak R1a1 Haplotypes*

The ancestral haplotypes for the Polish, Czech, and Slovak groups are very similar to each other, having only one insignificant difference in DYS439 (shown in bold). The average value is 10.43 on DYS439 in the Polish

group, 10.63 in the Czech and Slovak combined groups. The base haplotype for the 44 Polish haplotypes is:

13-25-16-10-11-14-12-12-10-13-11-30-16-9-10-11-11-23-14-20-32-12-15-15-16

and for the 27 Czech and Slovak haplotypes it is:

13-25-16-10-11-14-12-12-11-13-11-30-16-9-10-11-11-23-14-20-32-12-15-15-16

The difference in these haplotypes is really only 0.2 mutations--the alleles are just rounded in opposite directions.

The 44 Polish haplotypes have 310 mutations among them, and there are 175 mutations in the 27 Czech-Slovak haplotypes, resulting in  $0.282 \pm 0.016$  and  $0.259 \pm 0.020$  mutations per marker on average, respectively. These values correspond to  $4,550 \pm 520$  and  $4,125 \pm 430$  years back to their common ancestors, respectively, which is very similar to the German TSCA.

#### *The Remaining European R1a1 Haplotypes*

Many European countries are represented by just a few 25-marker haplotypes in the databases. 36 R1a1 haplotypes were collected from the YSearch database where the ancestral origin of the paternal line was from Denmark, Netherlands, Switzerland, Iceland, Belgium, France, Italy, Lithuania, Romania, Albania, Montenegro, Slovenia, Croatia, Spain, Greece, Bulgaria and Moldavia. The respective haplotype tree is shown in Figure 6.

The tree does not show any noticeable anomalies and points at just one common ancestor for all 36 individuals, who had the following base haplotype:

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

The ancestral haplotype match is exactly the same as those in Germany, Russia (see below), and has quite insignificant deviations from all other ancestral haplotypes considered above, within fractions of mutational differences. All the 36 individuals have 248 mutations in their 25-marker haplotypes, which corresponds to  $4,425 \pm 520$  years to the common ancestor. This is quite a common value for European R1a1 population.

#### *Russia and Ukraine R1a1 haplotypes*

The haplotype tree containing 110 of 25-marker haplotypes collected over 10 time zones from the Western Ukraine to the Pacific Ocean and from the northern tundra to Central Asia (Tadzhikistan and Kyrgyzstan) is shown in Figure 7.

The ancestral (base) haplotype for the haplotype tree is

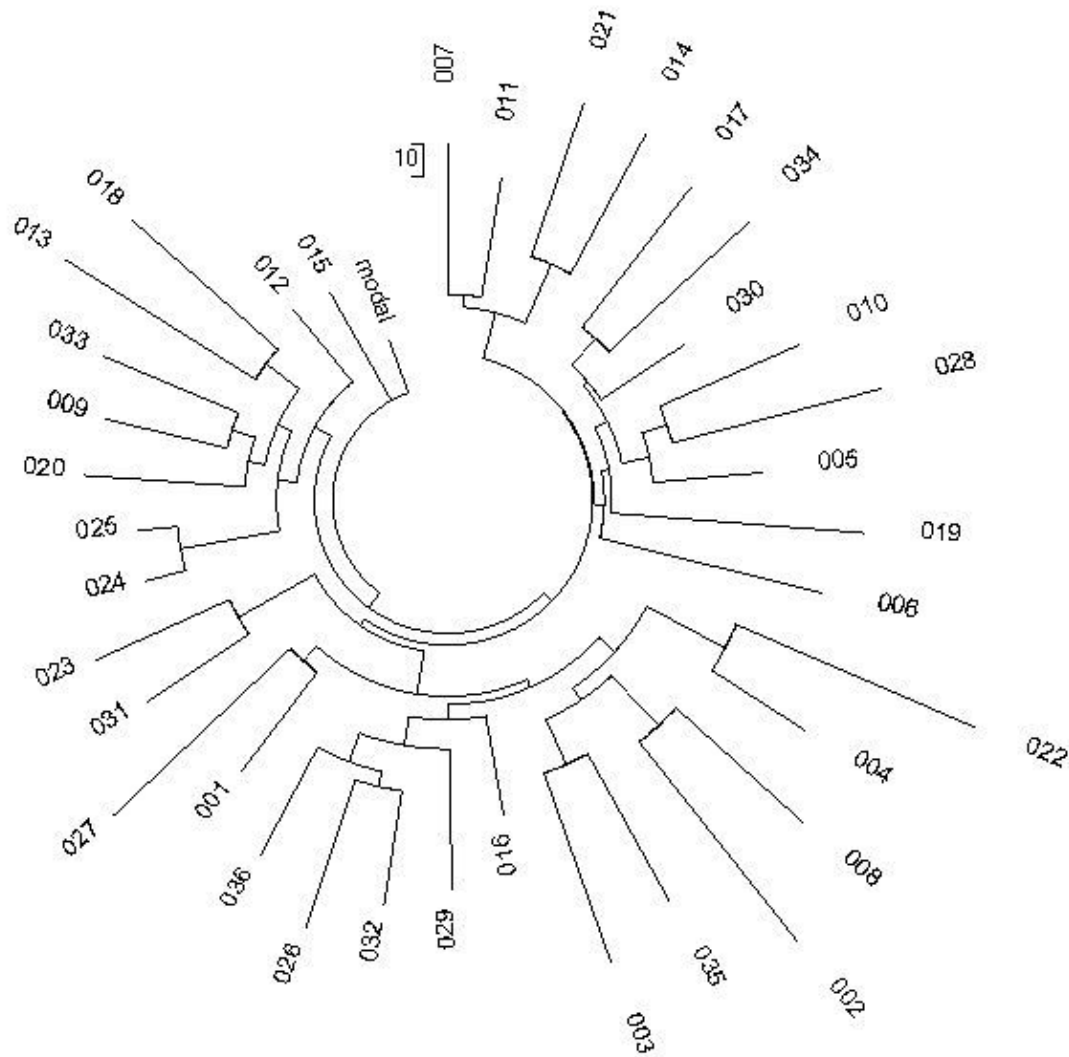


Figure 6. The 25-marker haplotype tree for various European countries (small series of haplotypes from Denmark, Netherlands, Switzerland, Iceland, Belgium, France, Italy, Lithuania, Romania, Albania, Montenegro, Slovenia, Croatia, Spain, Greece, Bulgaria and Moldavia), Haplogroup R1a1. The 36-haplotype tree was created using haplotypes from the YSearch database.

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

It is almost exactly the same ancestral haplotype as that in Germany. The average value on DYS391 is 10.53 in the Russian haplotypes, and was rounded to 11, and the base haplotype has only one insignificant deviation from the ancestral haplotype in England, which has the third allele (DYS19) 15.48, which in Russia/Ukraine it is 15.83 (calculated from one DYS19=14, 32 DYS19=15, 62 DYS19=16, and 15 DYS19=17). There are similar insignificant deviations with the Polish and Czechoslovakian base haplotypes, at DYS458 and DYS447, re-

spectively, within 0.2-0.5 mutations. This indicates a small mutational difference between their common ancestors.

The 110 Russian/Ukrainian haplotypes contain 804 mutations from the base haplotype, or  $0.292 \pm 0.010$  mutations per marker, resulting in  $4,750 \pm 500$  years from a common ancestor. The degree of asymmetry of this series of haplotypes is exactly 0.50, and does not affect the calculations.

R1a1 haplotypes of individuals who considered themselves of Ukrainian and Russian origin, present a practi-

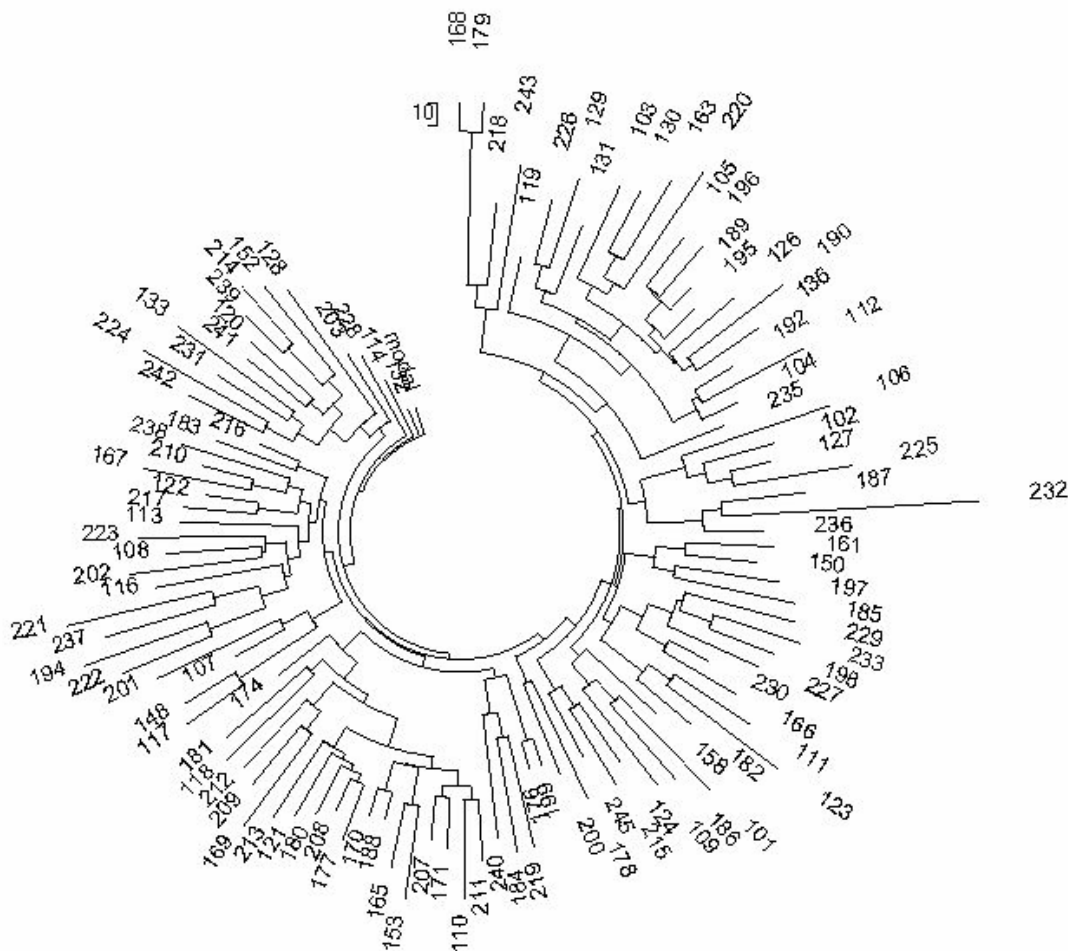


Figure 7. The 25-marker haplotype tree for Russia and Ukraine, Haplogroup R1a1. The 110-haplotype tree was composed from data of YSearch database and local Russian sources.

cally random geographic sample. For example, the haplotype tree contains two local Central Asian haplotypes (a Tadzhik and a Kyrgyz, haplotypes 133 and 127, respectively), as well as a local of a Caucasian Mountains Karachaev tribe (haplotype 166), though the male ancestry of the last one is unknown beyond his present tribal affiliation. They did not show any unusual deviations from other R1a1 haplotypes. Apparently, they are derived from the same common ancestor as are all other individuals of the R1a1 set.

The literature frequently refers to a statement that R1a1-M17 originated from a “refuge” in the present Ukraine about 15,000 years ago, following the Last Glacial Maximum. This statement was never substantiated by any actual data related to haplotypes and haplogroups.

It is just repeated over and over through a relay of references to references. The oldest reference is apparently that of Semino et al. (2000) which states that “this scenario is supported by the finding that the maximum variation for microsatellites linked to Eu19 [R1a1] is found in Ukraine” (Santachiara-Benerecetti, unpublished data). Now we know that this statement is incorrect. No calculations were provided in (Semino et al, 2000) or elsewhere which would explain the dating of 15,000 years.

Then, a paper by Wells et al. (2001) states “M17, a descendant of M173, is apparently much younger, with an inferred age of ~ 15,000 years.” No actual data or calculations are provided. The subsequent sentence in the paper says, “It must be noted that these age estimates

are dependent on many, possibly invalid, assumptions about mutational processes and population structure." This sentence has turned out to be valid in the sense that the estimate was inaccurate and overestimated by about 300% from the results obtained here. However, see the results below for the Chinese and southern Siberian haplotypes, below.

A more detailed consideration of R1a1 haplotypes in the Russian Plain and across Europe and Eurasia in general has shown that R1a1 haplogroup appeared in Europe between 12 and 10 thousand years before present, right after the Last Glacial Maximum, and after about 6,000 ybp had populated Europe, though, probably, with low

density. After 4,500 ybp R1a1 practically disappeared from Europe, incidentally, along with I1. Maybe more incidentally, it corresponded with the time period of populating of Europe with R1b1b2. Only those R1a1 who migrated to the Russian Plain from Europe around 6,000-5,000 years bp, stayed. They had expanded to the East, established on their way a number of archaeological cultures, including the Andronovo culture, which has embraced Northern Kazakhstan, Central Asia and South Ural and Western Siberia, and about 3600 ybp they migrated to India and Iran as the Aryans. Those who left behind, on the Russian Plain, re-populated Europe between 3200 and 2500 years bp, and stayed mainly in the Eastern Europe (present-day Poland, Germany,

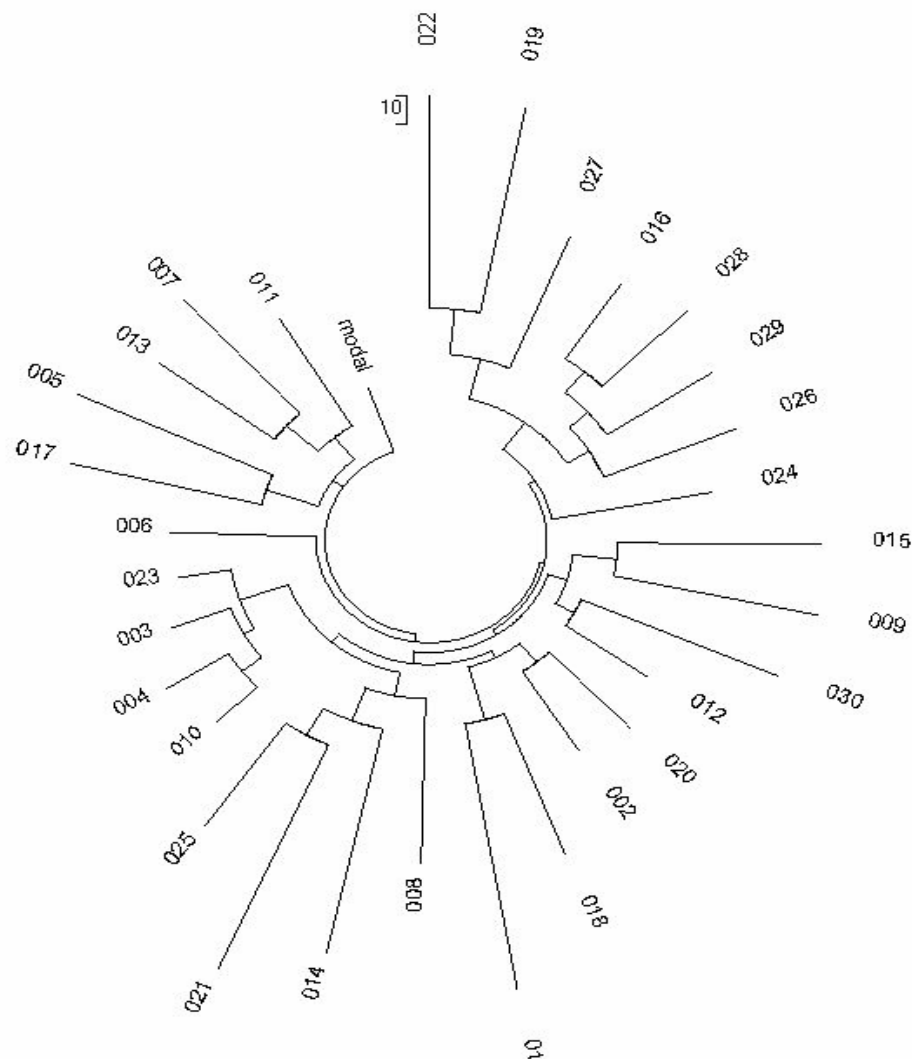


Figure 8. The 25-marker haplotype tree for India, Pakistan and Sri-Lanka, Haplogroup R1a1. The 30-haplotype tree was composed from data of YSearch and the India project (<http://www.familytreedna.com/public/India/default.aspx>) databases.

Czech, Slovak, etc. regions). Among them were carriers of the newly discovered R1a1-M458 subclade (Underhill et al, 2009).

### India R1a1 haplotypes

The YSearch database contains 22 of the 25-marker R1a1 haplotypes from India, including a few haplotypes from Pakistan and Sri Lanka. Their ancestral haplotype follows:

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

The only one apparent deviation in DYS458 (shown in bold) is related to an average alleles equal to 16.05 in Indian haplotypes, and 15.28 in Russian ones.

The India (Regional) Y-project at FTDNA (Rutledge, 2009) contains 15 haplotypes with 25 markers, eight of which are not listed in the YSearch database. Combined with others from Ysearch, a set of 30 haplotypes is available. Their ancestral haplotype is exactly the same as shown above. Figure 8 shows the respective haplotypes tree

All 30 Indian R1a1 haplotypes contain 191 mutations, that is  $0.255 \pm 0.018$  mutations per marker. It is statistically lower (with 95% confidence interval) than  $0.292 \pm 0.014$  mutations per marker in the Russian haplotypes and corresponds to  $4,050 \pm 500$  years from a common ancestor of the Indian haplotypes, compared to  $4,750 \pm 500$  years for the ancient "Russian" TSCA.

Archaeological studies have been conducted since the 1990's in the South Ural's Arkaim settlement and have revealed that the settlement was abandoned 3,600 years ago. The population apparently moved to northern India. That population belonged to Andronovo archaeological culture. Excavations of some sites of Andronovo culture, the oldest dating between 3,800 and 3,400 ybp, showed that nine inhabitants out of ten shared the R1a1 haplogroup and haplotypes (Bouakaze et al., 2007; Keyser et al, 2009). The base haplotype is as follows,:

13-25(24)-16(17)-11-11-14-X-Y-10-13(14)-11-31(32)

In this example, alleles that have not been assessed are replaced with letters. One can see that the ancient R1a1 haplotype closely resembles the Russian (as well as the other R1a1) ancestral haplotypes.

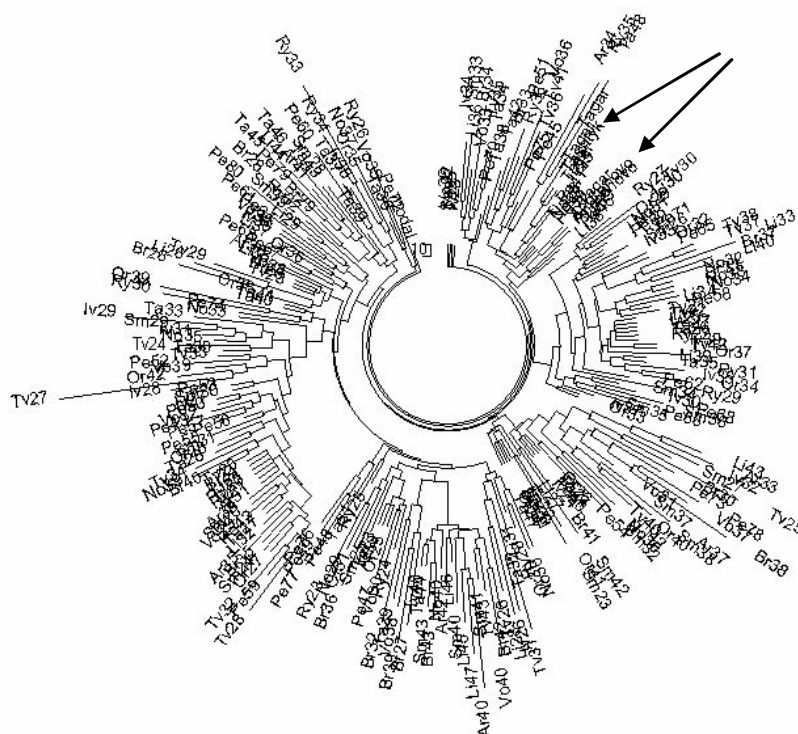
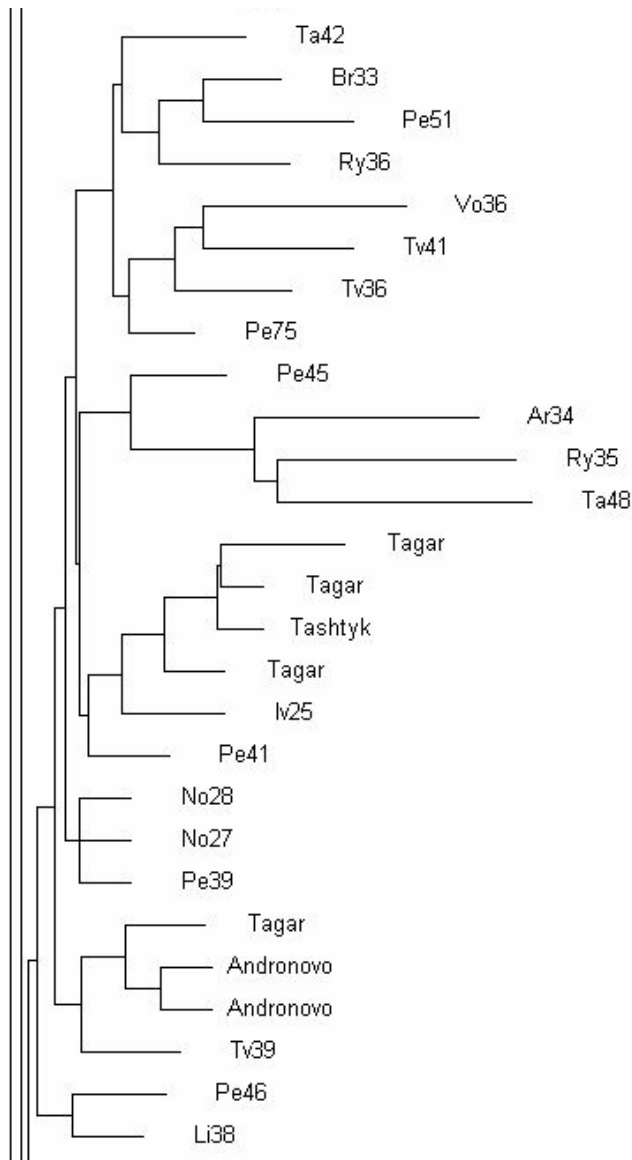


Figure 9. The 17-marker haplotype tree for R1a1 haplotypes of 262 ethnic Russians, most from Roewer et al (2008), to which seven ancient (excavated) Andronovo, Tagar, and Tashtyk haplotypes (Keyser et al, 2009) were added (their positions are indicated with the arrows; two more haplotypes were identical with two from the selection. See also Figure 10.



**Figure 10.** A fragment of the 17-marker haplotype tree for R1a1 haplotypes of ethnic Russians, to which seven ancient (excavated) Andronovo, Tagar, and Tashtyk haplotypes were added (Figure 9). Two-letter designations for the Russian regions are as follows: Ar–Archangelsk, Br–Bryansk, Iv–Ivanovo, Li–Lipezk, No–Novgorod, Pe–Penza, Ry–Ryazan, Ta–Tambov, Tv–Tver, Vo–Vologda.

In a recent paper (Keyser et al, 2009) the authors have extended the earlier analysis and determined 17-marker haplotypes for 10 Siberian individuals assigned to Andronovo, Tagar, and Tashtyk archaeological cultures. Nine of them shared the R1a1 haplogroup (one was of Haplogroup CxC3). The authors have reported that “none of the Y-STR haplotypes perfectly matched those

included in the databases,” and for some of those haplotypes “even the search based on the 9-loci minimal haplotype was fruitless.” However, as Figure 9 shows, all of the haplotypes nicely fit to the 17-marker haplotype tree of 252 Russian R1a1 haplotypes (with a common ancestor of  $4,750 \pm 490$  ybp, calculated using these 17-marker haplotypes [Klyosov, 2009b]), a list of which was recently published (Roewer et al, 2008). All seven Andronovo, Tagar and Tashtyk haplotypes (the other two were incomplete) are located in the upper right-hand side corner in Figure 9, and the respective branch on the tree is shown in Figure 10.

As one can see, the ancient R1a1 haplotypes excavated in Siberia, are comfortably located on the tree next to the haplotypes from the Russian cities and regions named in the legend to Figure 10.

The above data provide rather strong evidence that the R1a1 tribe migrated from Europe to the East between 5,000 and 3,600 ybp. The pattern of this migration is exhibited as follows: 1) the descendants who live today share a common ancestor of  $4,725 \pm 520$  ybp, 2) the Andronovo (and the others) archaeological complex of cultures in North Kazakhstan and South and Western Siberia dates 4,300 to 3,500 ybp, and it revealed several R1a1 excavated haplogroups, 3) they reached the South Ural region some 4,000 ybp, which is where they built Arkaim, Sintashta (contemporary names), and the so-called “a country of towns” in the South Ural region around 3,800 ybp, 4) by 3,600 ybp they abandoned the area and moved to India under the name of Aryans. The Indian R1a1 common ancestor of  $4,050 \pm 500$  ybp chronologically corresponds to these events. Currently, some 16% of the Indian population, that is about 100 millions males, and the majority of the upper castes, are members of Haplogroup R1a1 (Sengupta et al, 2006; Sharma et al, 2009).

#### *The Origin of Haplogroup R1a1*

Since we have mentioned Haplogroup R1a1 in India and in the archaeological cultures north of it, it is worthwhile to consider the question of where and when R1a1 haplotypes appeared in India. On the one hand, it is rather obvious from the above, that R1a1 haplotypes were brought to India around 3500 ybp from what is now Russia. The R1a1 bearers, known later as the Aryans, brought to India not only their haplotypes and the haplogroup, but also their language, thereby closing the loop, or building the linguistic and cultural bridge between India (and Iran) and Europe, and possibly creating the Indo-European family of languages. On the other hand, there is evidence that some Indian R1a1 haplotypes show a high variance, exceeding that in Europe (Kivisild et al, 2003; Sengupta et al, 2006; Sharma et al, 2009; Thanseem et al, 2009; Fornarino et al, 2009), thereby antedating the 4000-year-old migra-



**Table 3**  
**31 Five-Marker R1a1 Haplotypes from China**  
**(DYS 19, 388, 389-1, 389-2, 393)**

17 12 13 30 13	14 12 13 30 12	14 12 12 31 13
14 13 13 32 13	14 12 13 30 13	17 12 13 32 13
14 12 13 32 13	17 12 13 30 13	17 12 13 31 13
17 12 12 28 13	14 12 13 30 13	14 12 12 29 13
17 12 14 31 13	14 12 12 28 13	14 13 14 30 13
14 12 12 28 12	15 12 13 31 13	15 12 13 31 13
14 12 14 29 10	14 12 13 31 10	14 12 13 31 10
14 14 14 30 13	17 14 13 29 10	14 12 13 29 10
17 13 13 31 13	14 12 12 29 13	14 12 14 28 10
17 12 13 30 13	14 13 13 29 13	16 12 13 29 13
14 12 13 31 13		

tion from the north. However, there is a way to resolve the conundrum.

It seems that there are two quite distinct sources of Haplogroup R1a1 in India. One, indeed, was probably brought from the north by the Aryans. However, the most ancient source of R1a1 haplotypes appears to be provided by people who now live in China.

In an article by Bittles et al (2007) entitled "Physical anthropology and ethnicity in Asia: the transition from anthropometry to genome-based studies" a list of frequencies of Haplogroup R1a1 is given for a number of Chinese populations, however, haplotypes were not provided. The corresponding author, Professor Alan H. Bittles, kindly sent me the following list of 31 five-marker haplotypes (presented here in the format DYS19, 388, 389-1, 389-2, 393), in Table 3. The haplotype tree is shown in Figure 11.



**Figure 11.** The 5-marker haplotype tree for R1a1 haplotypes from China. The 31-haplotype tree was composed from data provided by Dr. A.H.Bittles and collected in ethnic communities Hui, Bolan, Dongxiang, and Sala (Bittles et al, 2007) (no haplotypes were provided in the referenced article).

These 31 5-marker haplotypes contain 99 mutations from the base haplotype (in the FTDNA format)

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

which gives  $0.639 \pm 0.085$  mutations per marker, or  $21,000 \pm 3,000$  years to a common ancestor. Such a large timespan to a common ancestor results from a low mutation rate constant, which was calculated from the Chandler's data for individual markers, as 0.00677 mutation/haplotype/generation, or 0.00135 mutation/marker/generation (see Table 1 in Part I of the companion article).

Since haplotypes descended from such a ancient common ancestor have many mutations, which makes their base (ancestral) haplotypes rather uncertain, the ASD permutational method was employed for the Chinese set of haplotypes (Part I). The ASD permutational method does not need a base haplotype and it does not require a correction for back mutations. For the given series of 31 five-marker haplotypes 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. It gives an average number of mutations per marker of 1.060. After division of this value by the mutation rate for the five-marker haplotypes (0.00135 mut/marker/generation for 25 years per generation), we obtain  $19,625 \pm 2,800$  years to a common ancestor. It is within the margin of error with that calculated by the linear method, as shown above.

It is likely that Haplogroup R1a1 had appeared in South Siberia around 20 thousand years ago, and its bearers split. One migration group headed West, and had arrived to the Balkans around 12 thousand years ago (see below). Another group had appeared in China some 21 thousand years ago. Apparently, bearers of R1a1 haplotypes made their way from China to South India between five and seven thousand years ago, and those haplotypes were quite different compared to the Aryan ones, or the "Indo-European" haplotypes. This is seen from the following data.

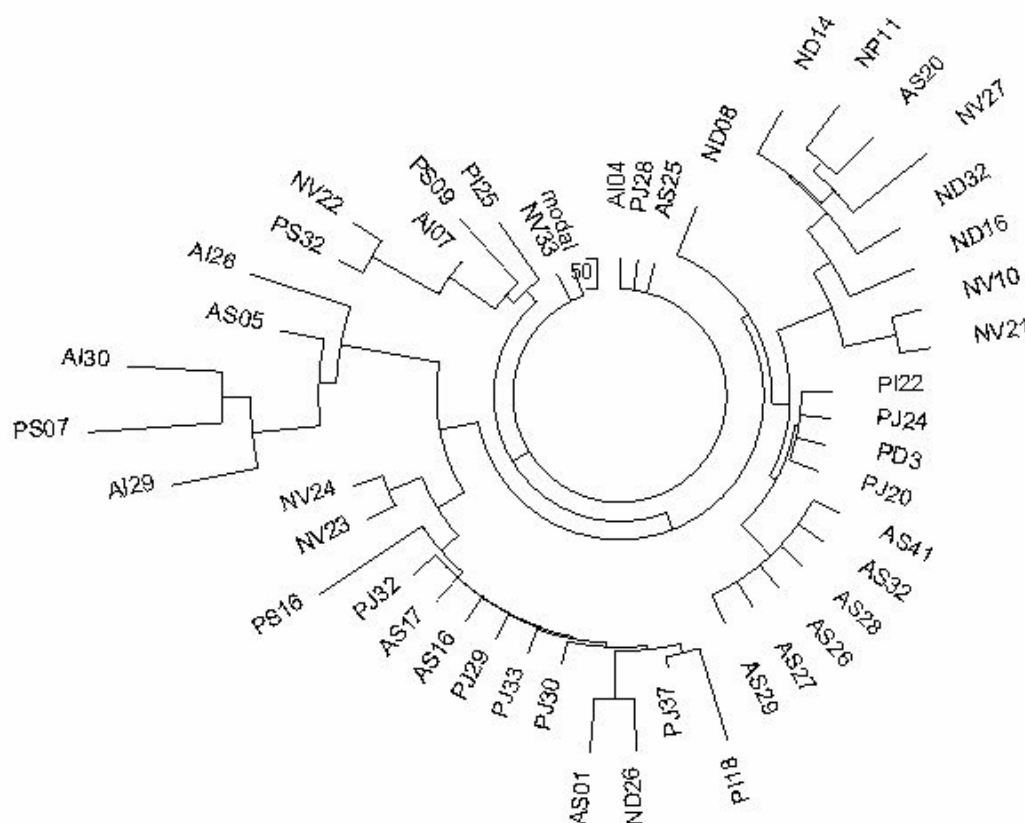


Figure 12. The 6-marker haplotype tree for R1a1 haplotypes in Andhra Pradesh (tribes Naikpod, Andh, and Pardhan), South India. The 46-haplotype tree was composed from data listed in (Thanseem et al, 2006). The designations of haplotypes are those used in the article.

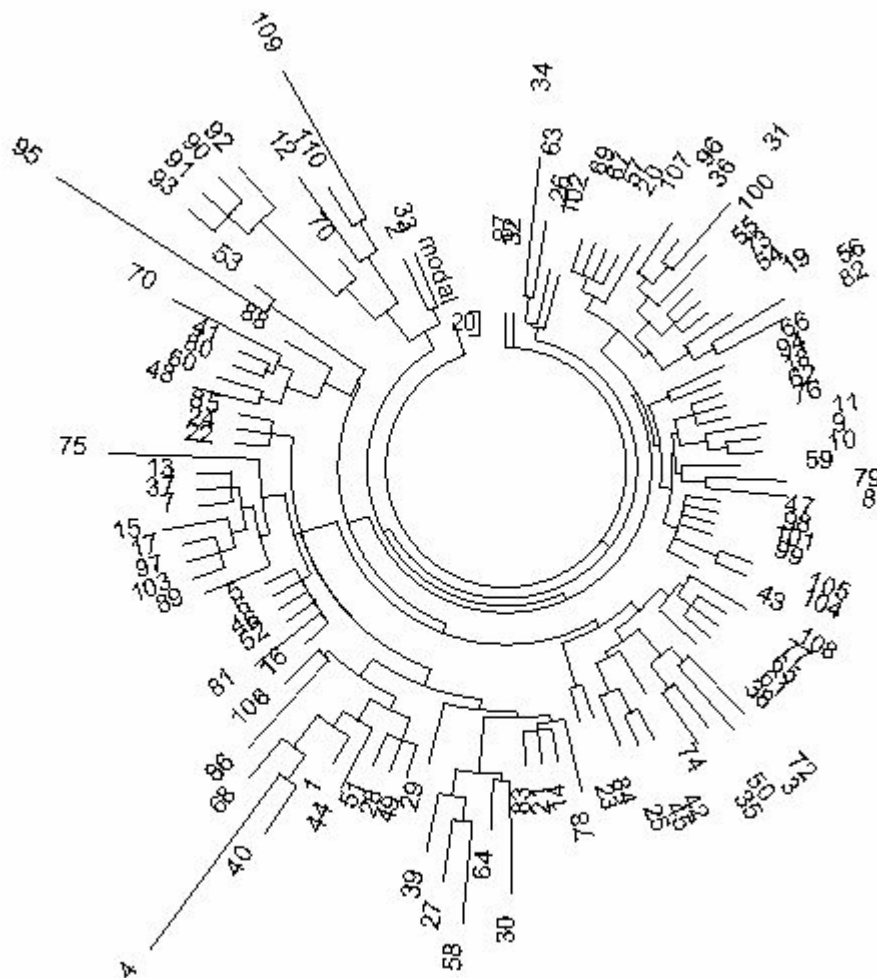


Figure 13. The 10-marker haplotype tree for R1a1 haplotypes in India (mixed population, including tribes and castes). The 110-haplotype tree was composed from data listed in (Sengupta et al, 2006). The article contains 114 Indian R1a1 haplotypes, however, four of them were incomplete.

Thanseem et al (2006) found 46 R1a1 haplotypes of six markers each in three different tribal population of Andra Pradesh, South India (tribes Naikpod, Andh, and Pardhan). The haplotypes are shown in a haplotype tree in Figure 12. The 46 haplotypes contain 126 mutations, so there were  $0.457 \pm 0.050$  mutations per marker. This gives  $7,125 \pm 950$  years to a common ancestor. The base (ancestral) haplotype of those populations in the FTD-NA format is as follows:

13-25-17-9-X-X-X-X-14-X-32

It differs from the “Indo-European” Indian haplotype,

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

by four mutations on six markers, which corresponds to 11,850 years between their common ancestors, and places their common ancestor at approximately  $(11850 + 4050 + 7125)/2 = 11,500$  ybp.

110 of 10-marker R1a1 haplotypes of various Indian populations, both tribal and Dravidian and Indo-European castes, listed in (Sengupta et al, 2006) shown in a haplotype tree in Figure 13, contain 344 mutations, that is  $0.313 \pm 0.019$  mutations per marker. It gives  $5,275 \pm 600$  years to a common

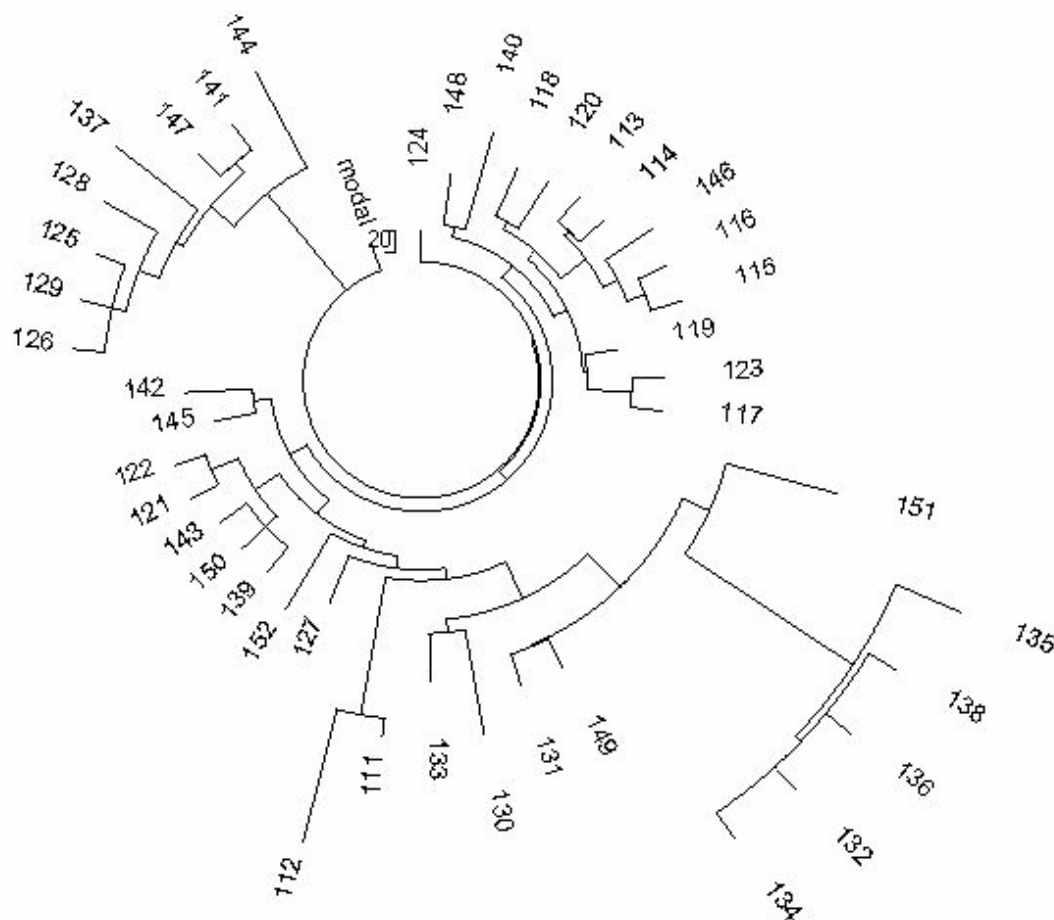


Figure 14. The 10-marker haplotype tree for R1a1 haplotypes in Pakistan. The 42-haplotype tree was composed from data listed in (Sengupta et al, 2006).

ancestor. The base (ancestral) haplotype of those populations in the FTDNA format is as follows:

13-25-15-10-X-X-X-12-10-13-11-30

It differs from the “Indo-European” Indian haplotype

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

by just 0.5 mutations on nine markers (DYS19 has, in fact, a mean value of 15.5 in the Indian haplotypes), which makes them practically identical. However, in this haplotype series two different populations, the “Indo-European” one and the “South-Indian” one were mixed, therefore an “intermediate”, apparently phantom “common ancestor” was artificially created with an intermediate TSCA, between  $4,050 \pm 500$  and  $7,125 \pm 950$  (see above).

For a comparison, let us consider Pakistani R1a1 haplotypes listed in the article by Sengupta (2003) and shown in Figure 14. The 42 haplotypes contain 166 mutations, which gives  $0.395 \pm 0.037$  mutations per marker, and  $7,025 \pm 890$  years to a common ancestor. This value fits within the margin of error to the “South-Indian” TSCA of  $7,125 \pm 950$  ybp. The base haplotype was as follows:

13-25-17-11-X-X-X-12-10-13-11-30

It differs from the “Indo-European” Indian haplotype by two mutations in the nine markers.

Finally, we will consider Central Asian R1a1 haplotypes listed in the same paper by Sengupta (2006). Ten haplotypes contain only 25 mutations, which gives  $0.250 \pm 0.056$  mutations per haplotype, and  $4050 \pm 900$

years to a common ancestor. It is the same value that we have found for the “Indo-European” Indian haplotypes.

The above suggests that there are two different subsets of Indian R1a1 haplotypes. One was brought by European bearers known as the Aryans, seemingly on their way through Central Asia, in the 2<sup>nd</sup> millennium BCE, another, much more ancient, made its way apparently through China, and arrived in India earlier than seven thousand years ago.

#### *R1a1 Haplotypes of the Arabian Peninsula*

Sixteen R1a1 10-marker haplotypes from Qatar and United Arab Emirates have been recently published (Cadenas et al., 2008). They split into two branches, with 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

which are only slightly different, and on DYS19 the mean values are almost the same, just rounded up or down. The first haplotype is the base one for seven haplotypes with 13 mutations in them, on average  $0.186 \pm 0.052$  mutations per marker, which gives  $2,300 \pm 680$  years to a common ancestor. The second haplotype is the base one for nine haplotypes with 26 mutations, an average  $0.289 \pm 0.057$  mutations per marker or  $3,750 \pm 825$  years to a common ancestor. Since a common ancestor of R1a1 haplotypes in Armenia and Anatolia lived  $4,500 \pm 1,040$  and  $3,700 \pm 550$  ybp, respectively (Klyosov, 2008b), it does not conflict with  $3,750 \pm 825$  ybp in the Arabian peninsula.

#### *The Balkan Ancient Branch: the Oldest Trace of Haplogroup R1a in Europe?*

A series of 67 haplotypes of Haplogroup R1a1 from the Balkans was published (Barac et al., 2003a, 2003b; Pericic et al., 2005). They were presented in a 9-marker format only. The respective haplotype tree is shown in Figure 15.

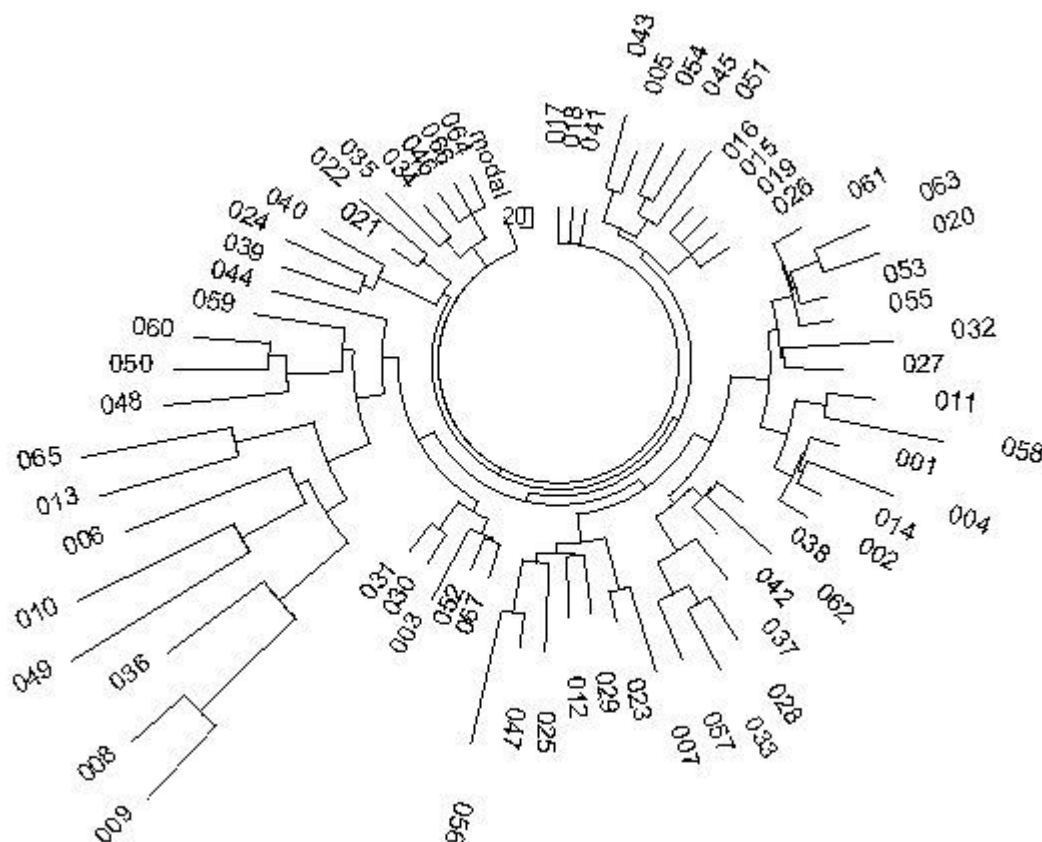


Figure 15. The 9-marker haplotype tree for the Balkans, Haplogroup R1a1. The 67-haplotype tree was composed from data published (Barac et al., 2003a, 2003b; Pericic et al., 2005).

One can see a remarkable branch on the left-hand side of the tree which stands out as an “extended and fluffy” one. These are typically features of a very old branch compared with others on the same tree. Also, a common feature of ancient haplotype trees is that they are typically “heterogeneous” ones and consist of a number of branches.

The tree in Figure 15 includes a rather small branch of twelve haplotypes on top of the tree, which contains only 14 mutations. This results in  $0.130 \pm 0.035$  mutations per marker, or  $1,850 \pm 530$  years to a common ancestor. Its base haplotype

13-25-16-10-11-14-X-Y-Z-13-11-30

is practically the same as that in Russia and Germany.

The wide 27-haplotype branch on the right contains  $0.280 \pm 0.034$  mutations per marker, which is rather typical for R1a1 haplotypes in Europe. It gives typical in kind  $4,350 \pm 680$  years to a common ancestor of the branch. Its base haplotype

13-25-16-11-11-14-X-Y-Z-13-11-30

is again typical for Eastern European R1a1 base haplotypes, in which the fourth marker (DYS391) often fluctuates between 10 and 11. Of 110 Russian-Ukrainian haplotypes diagramed in Figure 7, 51 haplotypes have “10”, and 57 have “11” in that locus (one has “9” and one has “12”). In 67 German haplotypes, discussed above, 43 haplotypes have “10”, 23 have “11” and one has “12.” Hence, the Balkan haplotypes from this branch are more close to the Russian than to the German haplotypes.

The “extended and fluffy” 13-haplotype branch on the left contains the following haplotypes:

13 24 16 12 14 15 13 11 31  
 12 24 16 10 12 15 13 13 29  
 12 24 15 11 12 15 13 13 29  
 14 24 16 11 11 15 15 11 32  
 13 23 14 10 13 17 13 11 31  
 13 24 14 11 11 11 13 13 29  
 13 25 15 9 11 14 13 11 31  
 13 25 15 11 11 15 12 11 29  
 12 22 15 10 15 17 14 11 30  
 14 25 15 10 11 15 13 11 29  
 13 25 15 10 12 14 13 11 29  
 13 26 15 10 11 15 13 11 29  
 13 23 15 10 13 14 12 11 28

The set does not contain a haplotype which can be defined as a base. This is because common ancestor lived too long ago, and all haplotypes of his descendants living today are extensively mutated. In order to deter-

mine when that common ancestor lived, we have employed three different approaches, described in the preceding paper (Part 1), namely the “linear” method with the correction for reverse mutations, the ASD method based on a deduced base (ancestral) haplotype, and the permutational ASD method (no base haplotype considered). The linear method gave the following deduced base haplotype, an alleged one for a common ancestor of those 13 individuals from Serbia, Kosovo, Bosnia and Macedonia:

13-24-15-10-12-15-X-Y-Z-13-11-29

The bold notations identify deviations from typical ancestral (base) East-European haplotypes. The third allele (DYS19) is identical to the Atlantic and Scandinavian R1a1 base haplotypes. All 13 haplotypes contain 70 mutations from this base haplotype, which gives  $0.598 \pm 0.071$  mutations on average per marker, and results in  $11,425 \pm 1,780$  years from a common ancestor.

The “quadratic method” (ASD) gives the following “base haplotype” (the unknown alleles are eliminated here, and the last allele is presented as the DYS389-2 notation):

12.92 – 24.15 – 15.08 – 10.38 – 12.08 – 14.77 – 13.08  
 – 11.46 – 16.62

A sum of square deviations from the above haplotype results in 103 mutations total, including reverse mutations “hidden” in the linear method. Seventy “observed” mutations in the linear method amount to only 68% of the “actual” mutations including reverse mutations. Since all 13 haplotypes contain 117 markers, the average number of mutations per marker is  $0.880 \pm 0.081$ , which corresponds to  $0.880/0.00189 = 466 \pm 62$  generations or  $11,650 \pm 1,550$  years to a common ancestor. 0.00189 is the mutation rate (in mutations per marker per generation) for the given 9-marker haplotypes (companion paper, Part I).

A calculation of  $11,650 \pm 1,550$  years to a common ancestor is practically the same as  $11,425 \pm 1,780$  years, obtained with linear method and corrected for reverse mutations.

The all-permutation “quadratic” method (Adamov & Klyosov, 2008) gives 2,680 as a sum of all square differences in all permutations between alleles. When divided by N (N = number of haplotypes, that is 13), by 9 (number of markers in haplotype), and by 2 (since deviation were both “up” and “down”), we obtain an average number of mutations per marker equal to 0.881. It is near exactly equal to 0.880 obtained by the quadratic method above. Naturally, it gives again  $0.881/0.00189 = 466 \pm 62$  generations or  $11,650 \pm 1,550$  years to a common ancestor of the R1a1 group in the Balkans.

These results suggest that the first bearers of the R1a1 haplogroup in Europe lived in the Balkans (Serbia, Kosovo, Bosnia, Macedonia) between 10 and 13 thousand ybp. It was shown below that Haplogroup R1a1 has appeared in Asia, apparently in China or rather South Siberia, around 20,000 ybp. It appears that some of its bearers migrated to the Balkans in the following 7-10 thousand years. It was found (Klyosov, 2008a) that Haplogroup R1b appeared about 16,000 ybp, apparently in Asia, and it also migrated to Europe in the following 11-13 thousand years. It is of a certain interest that a common ancestor of the ethnic Russians of R1b is dated  $6775 \pm 830$  ybp (Klyosov, to be published), which is about two to three thousand years earlier than most of the European R1b common ancestors. It is plausible that the Russian R1b have descended from the Kurgan archaeological culture.

The data shown above suggests that only about 6,000-5,000 ybp bearers of R1a1, presumably in the Balkans, began to mobilize and migrate to the west toward the Atlantic, to the north toward the Baltic Sea and Scandinavia, to the east to the Russian plains and steppes, to the south to Asia Minor, the Middle East, and far south to the Arabian Sea. All of those local R1a1 haplotypes point to their common ancestors who lived around 4,800 to 4,500 ybp. On their way through the Russian plains and steppes the R1a1 tribe presumably formed the Andronovo archaeological culture, apparently domesticated the horse, advanced to Central Asia and formed the "Aryan population" which dated to about 4,500 ybp. They then moved to the Ural mountains about 4,000 ybp and migrated to India as the Aryans circa 3,600-3,500 ybp. Presently, 16% of the male Indian population, or approximately 100 million people, bear the R1a1 haplogroup's SNP mutation (SRY10831.2), with their common ancestor of  $4,050 \pm 500$  ybp, of times back to the Andronovo archaeological culture and the Aryans in the Russian plains and steppes. The current "Indo-European" Indian R1a1 haplotypes are practically indistinguishable from Russian, Ukrainian, and Central Asian R1a1 haplotypes, as well as from many West and Central European R1a1 haplotypes. These populations speak languages of the Indo-European language family.

The next section focuses on some trails of R1a1 in India, an "Aryan trail."

#### *The Chenchu R1a1 Haplotypes*

Kivisild et al. (2003) reported that eleven out of 41 individuals tested in the Chenchu, an australoid tribal group from southern India, are in Haplogroup R1a1 (or 27% of the total). It is tempting to associate this with the Aryan influx into India, which occurred some 3,600-3,500 ybp. However, questionable calculations of time spans to a common ancestor of R1a1 in India,

and particularly in the Chenchu (Kivisild et al., 2003; Sengupta et al., 2006; Thanseem et al, 2006; Sahoo et al, 2006; Sharma et al, 2009; Fornarino et al, 2009) using methods of population genetics rather than those of DNA genealogy have precluded an objective and balanced discussion of the events and their consequences.

The eleven R1a1 haplotypes of the Chenchu (Kivisild et al., 2003) do not provide good statistics; however, they can allow a reasonable estimate of a time span to a common ancestor for these 11 individuals. Logically, if these haplotypes are more or less identical, with just a few mutations in them, a common ancestor would likely have lived within a thousand or two thousands of ybp. Conversely, if these haplotypes are all mutated, and there is no base (ancestral) haplotype among them, a common ancestor lived thousands ybp. Even two base (identical) haplotypes among 11 would tentatively give  $\ln(11/2)/0.0088 = 194$  generations, which, corrected to back mutations, would result in 240 generations, or 6,000 years to a common ancestor, with a large margin of error. If eleven of the six-marker haplotypes are all mutated, it would mean that a common ancestor lived apparently earlier than 6 thousand ybp. Hence, even with such a small set of haplotypes one can obtain useful and meaningful information.

The eleven Chenchu haplotypes have seven identical (base) six-marker haplotypes (in the format of DYS 19-388-290-391-392-393, commonly employed in earlier scientific publications):

16-12-24-11-11-13

They are practically the same as those common East European ancestral haplotypes considered above, if presented in the same six-marker format:

16-12-25-11(10)-11-13

Actually, the author of this study, himself an R1a1 Slav (R1a1) has the "Chenchu" base six-marker haplotype.

These identical haplotypes are represented by a "comb" in Figure 16. If all seven identical haplotypes are derived from the same common ancestor as the other four mutated haplotypes, the common ancestor would have lived on average of only 51 generations bp, or less than 1300 years ago [ $\ln(11/7)/0.0088 = 51$ ], with a certain margin of error (see estimates below). In fact, the Chenchu R1a1 haplotypes represent two lineages, one  $3,200 \pm 1,900$  years old and the other only  $350 \pm 350$  years old, starting from around the 17<sup>th</sup> century CE. The tree in Figure 16 shows these two lineages.

A quantitative description of these two lineages is as follows. Despite the 11-haplotype series containing seven identical haplotypes, which in the case of one

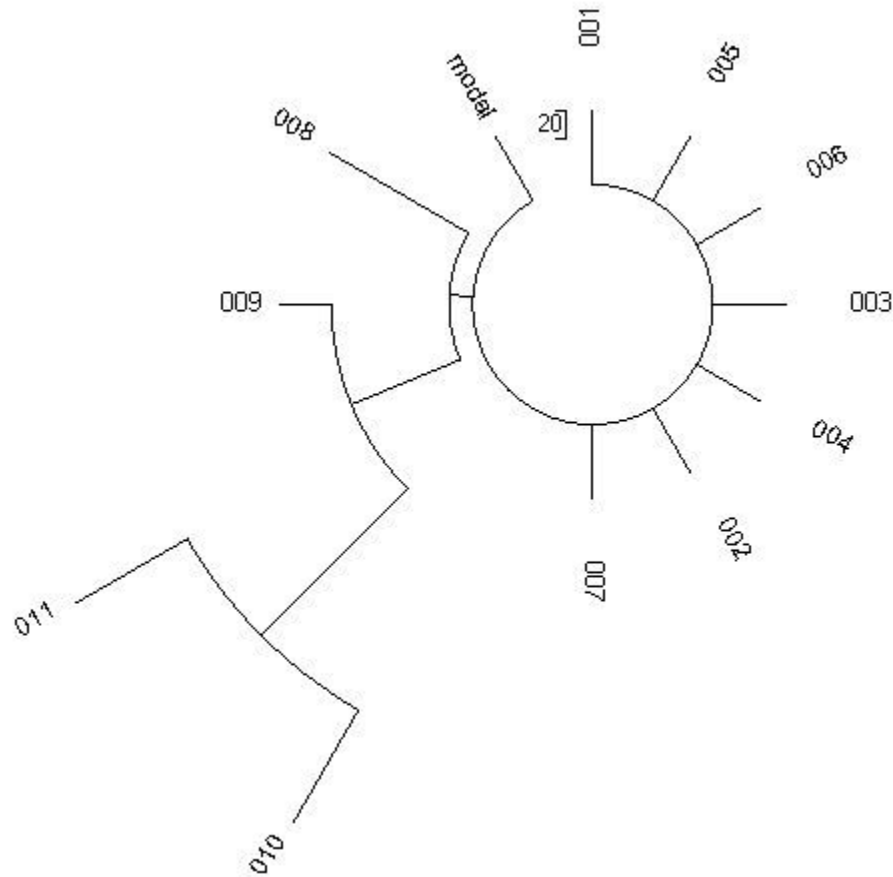


Figure 16. The six-marker haplotype tree for the South Indian tribe Chenchu, Haplogroup R1a1. The 11-haplotype tree was created from data of Kivisild et al. (2003).

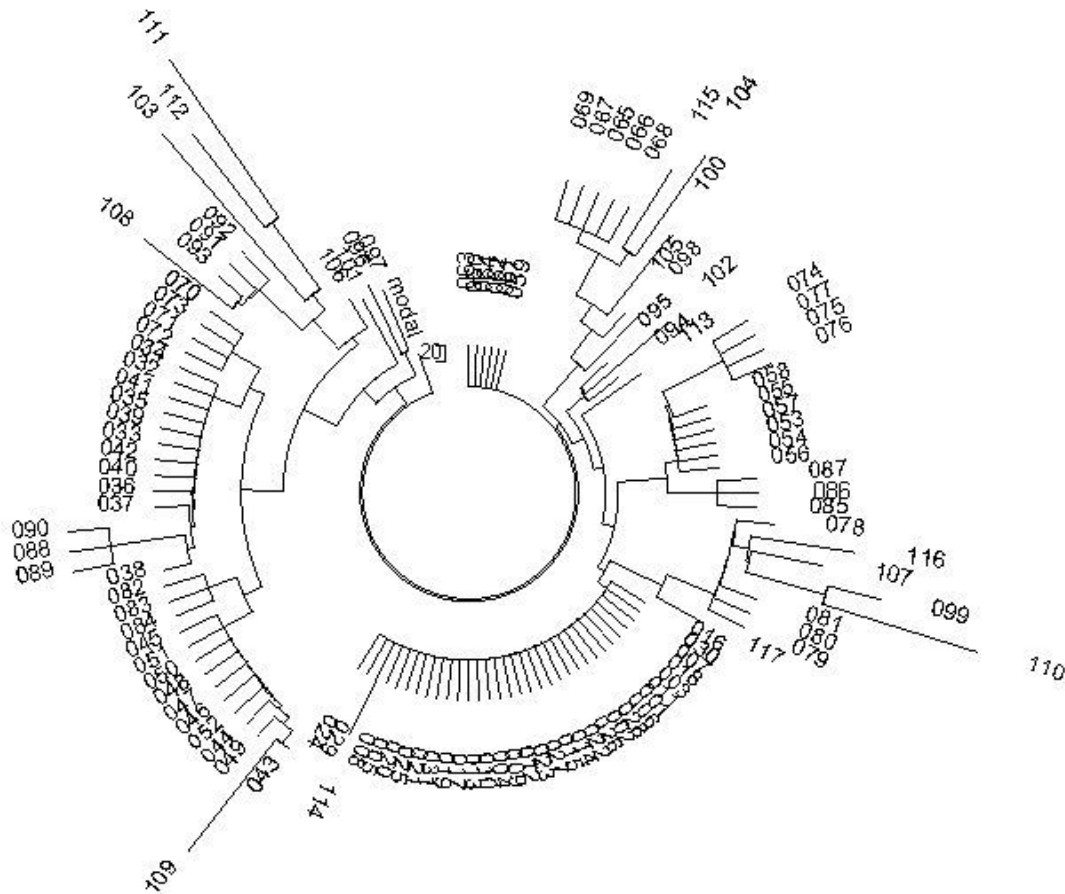
common ancestor for the series, would have indicated 51 generations (with a proper margin of error) from a common ancestor, the same 11 haplotypes contain 9 mutations from the above base haplotype. The linear method gives  $9/11/0.0088 = 93$  generations to a common ancestor (both values without a correction for back mutations). Because there is a significant mismatch between the 51 and 93 generations, one can conclude that the 11 haplotypes descended from more than one common ancestor. Clearly, the Chenchu R1a1 haplotype set points to a minimum of two common ancestors, which is confirmed by the haplotype tree (Figure 16). A recent branch includes eight haplotypes, seven being base haplotypes, and one with only one mutation. The older branch, contains three haplotypes containing three mutations from their base haplotype:

15-12-25-10-11-13

The recent branch results in  $\ln(8/7)/0.0088 = 15$  generations (by the logarithmic method), and  $1/8/0.0088 = 14$  generations (the linear method) from the residual seven base haplotypes and a number of mutations (just one), respectively. It shows a good fit between the two estimates. This confirms that a single common ancestor for eight individuals of the eleven lived only about  $350 \pm 350$  ybp, around the 17<sup>th</sup> century. The old branch of haplotypes points at a common ancestor who lived  $3/3/0.0088 = 114 \pm 67$  generations BP, or  $3,200 \pm 1,900$  ybp with a correction for back mutations.

Considering that the Aryan (R1a1) migration to northern India took place about 3,600-3,500 ybp, it is quite





plausible to attribute the appearance of R1a1 in the Chenchu by  $3,200 \pm 1,900$  ybp to the Aryans.

Native American Haplotypes of Haplogroup Q1a3a

The tree contains 31 identical (base) haplotypes and 273 mutations from that “base” haplotype. It is obvious

An analysis of the haplotype tree in Figure 17 shows that it includes at least six lineages, each with its own common ancestor. Four of them turned out to be quite

recent common ancestors, who lived within the last thousand years. They had the following base haplotypes:

13-12-24-10-14-13  
13-12-23-10-14-13  
13-12-24-10-15-12  
13-12-24-10-13-14

The oldest branch contains 11 haplotypes with the following base (ancestral) haplotype:

13-13-24-10-14-14

This branch contains 32 mutations, which gives  $0.485 \pm 0.086$  mutations per marker on average for six-marker haplotypes, that is  $12,125 \pm 2,460$  years to a common ancestor for those 11 individuals.

However, this was the most ancient ancestor of just one branch of haplotypes. From several base haplotypes shown above one can see that a mutational difference between this base haplotype and more recent base haplotypes reach 4 mutations per six-marker haplotype. This corresponds to about 19,700 years of mutational difference between them and points out that *their* common ancestor lived  $16,300 \pm 3,300$  ybp.

Hence, a common ancestor of several groups of individuals among Native Americans of Haplogroup Q1a3a and having largely varied haplotypes, lived between 13,000 and 19,600 ybp with the 95% confidence interval. This dating is in line with many independent dates from archaeological, climatological, and genetic studies of Native American origins. Some researchers refer the peopling of the Americas to the end of the last glacial maximum, approximately 23,000 to 19,000 years ago and suggest a strong population expansion started approximately 18,000 and finished 15,000 ybp (Fagundes et al, 2008). Others refer to archaeology data of Paleo-Indian people between 11,000 to approximately 18-22,000 ybp (Haynes, 2002, p. 52; Lepper, 1999, pp. 362-394; Bradley and Stanford, 2004; Seielstad et al., 2003). In any event, the time span of 16,000 years ago is quite compatible with those estimates.

## The Cohen Modal Haplotype of Haplogroups J1 and J2 (the Jewish and Arabic Haplotypes)

### *CMH, Haplogroup J1*

The “Cohen Modal Haplotype” (CMH) was introduced (Thomas et al., 1998) ten years ago to designate the following six-marker haplotype (in DYS 19-388-390-391-392-393 format):

14-16-23-10-11-12

Further research showed that this haplotype presents in both J1 and J2 haplogroups.

In Haplogroup J1 if 12-marker haplotypes are collected and analyzed, the 12-marker haplotype corresponding to the CMH may be shown to split into two principal lineages (Klyosov, 2008c), with the following base (ancestral) haplotypes:

12-23-14-10-13-15-11-16-12-13-11-30

and

12-23-14-10-13-17-11-16-11-13-11-31

which differ from each other by four mutations (shown in bold).

In Haplogroup J2 the same six-marker “CMH” also exists, and if 12-marker haplotypes containing the CMH are collected and analyzed, the following 12-marker “CMH” modal or base haplotype can be determined (Klyosov, 2008c):

12-23-14-10-13-17-11-16-11-13-11-30

It differs from both J1 CMHs by three and one mutations, respectively.

In fact, actual Cohanim of Haplogroup J2, recognized by the Cohanim Association (private communication from the President of the Cohanim Association of Latin America, Mr. Mashuah HaCohen-Pereira; also [www.cohen.org.br](http://www.cohen.org.br)), have the following base 12-marker haplotypes

12-23-15-10-14-17-11-15-12-13-11-29

which differs by five mutations from the above CMH-J2 base haplotype (to be published). Another Cohanim base haplotype

12-23-15-10-13-18-11-15-12-13-11-29

is currently assigned to the Cohanim-Sephardim lineage. The above Cohanim haplotype belonged to the common ancestor who lived  $3,000 \pm 560$  ybp. The presumably Cohanim-Sephardim base haplotype belonged to the common ancestor who lived  $2,500 \pm 710$  ybp (to be published).

In this section we consider the J1 “Cohen Modal Haplotype” in its extended format of the 25-, 37- and 67-marker haplotypes.

A 25-marker haplotype tree of 49 presumed Jewish J1 haplotypes is shown in Figure 18. In this tree the two CMH branches are located on either side at the top of

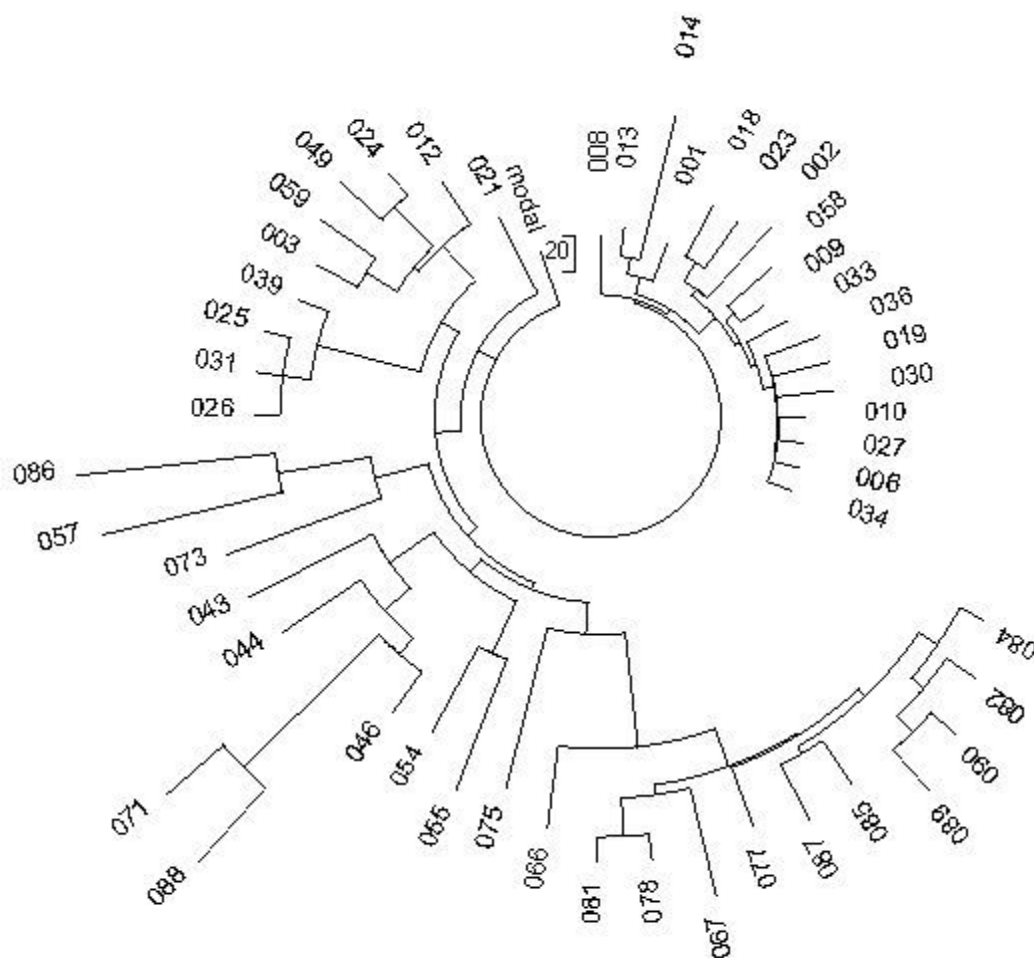


Figure 18. The 25-marker Haplogroup J1 haplotype tree for 49 presumably Jewish haplotypes. The haplotypes were collected in the YSearch database (Klyosov, 2008c).

the tree, the “recent CMH” (rCMH) is the more compact 17-haplotype branch on the right (between haplotypes 008 and 034), and the 9-haplotype “older CMH” (oCMH) branch is on the left (between haplotypes 012 and 026). The base haplotype for the “recent CMH” is

12-23-14-10-13-15-11-16-12-13-11-30-17-8-9-11-11-  
26-14-21-27-12-14-16-17

and for the “older CMH”

12-23-14-10-13-17-11-16-11-13-11-31-18-8-9-11-11-  
25-14-20-25-12-14-16-17

There are 9 mutations between these two base haplotypes as shown above in bold. If we sum the distances between them and their common ancestor's haplotype

(using the mean values for the ancestors values), there are 7.2 mutations between them. This corresponds to about 4,650 years of a mutational difference between them (that is, a sum of the distances between them and THEIR common ancestor). We will use this result later.

The rCMH branch contains 41 mutations, which gives only  $0.0965 \pm 0.0015$  mutations per marker on average and corresponds to  $1,400 \pm 260$  years to a common ancestor. The oCMH branch contains 36 mutations, which gives  $0.160 \pm 0.027$  mutations per marker on average and corresponds to  $2,400 \pm 470$  years to a common ancestor.

From the data obtained we can calculate that the common ancestor of the two branches lived about  $4,225 \pm 520$  ybp. That is when a common ancestor of the

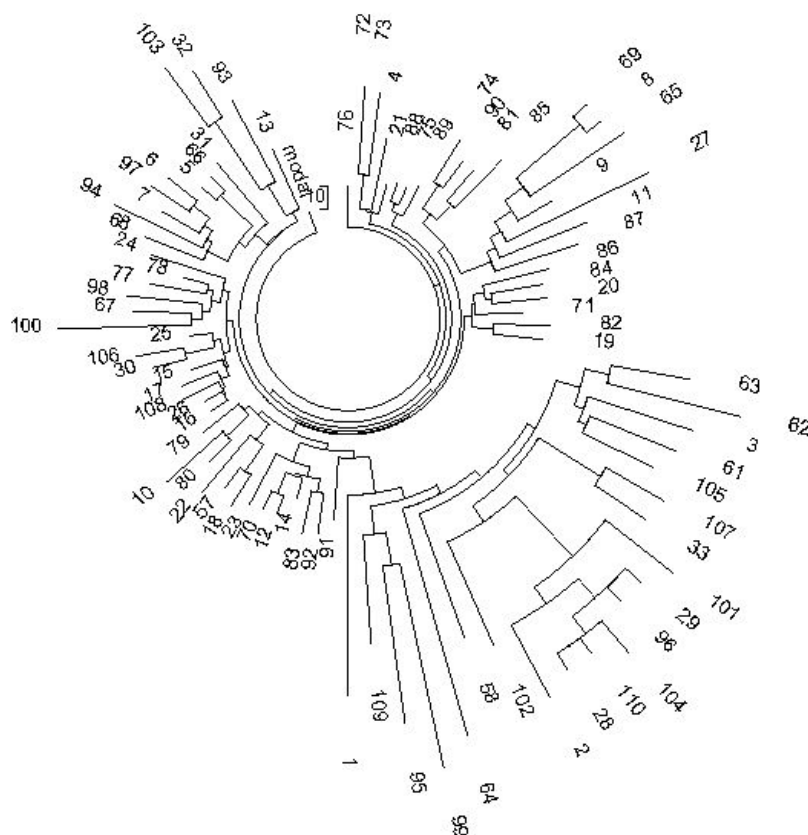
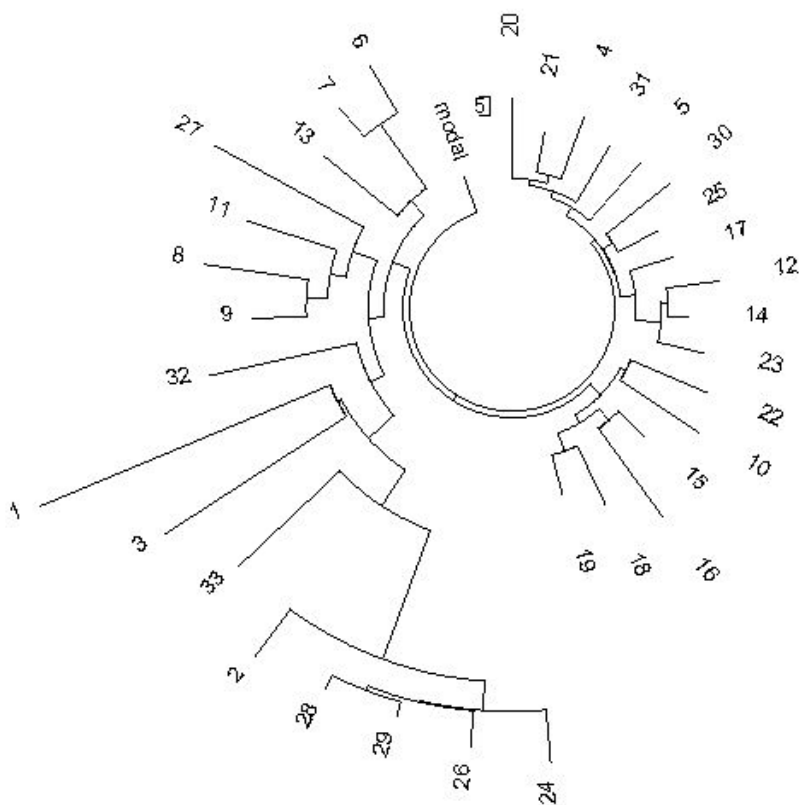


Figure 19. The 37-marker haplotype tree for the “Cohen Modal Haplotypes”, Haplogroup J1. The 85 haplotype tree was composed of haplotypes collected in YSearch database (Klyosov, 2008c) and private “Cohen Haplotype” projects, and provided by Dr. Alberto Aburto.

Figure 20. The 67-marker haplotype tree for the “Cohen Modal Haplotypes”, haplogroup J1. The 33 haplotype tree was composed of haplotypes collected from three sources: 1) YSearch database (Klyosov, 2008c), 2) private “Cohen Haplotype” projects, and 3) provided by Dr. Alberto Aburto.



“Cohen modal haplotype” lived among the future Jewish community of Haplogroup J1, according to information stored in their 25-marker haplotypes.

#### “Cohen Modal Haplotypes” of Jewish and non-Jewish Descent.

In order to study the question of the origin of the CMH further, we examined 37-marker haplotypes. 85 J1 haplotypes having 37 markers and also the embedded six-marker CMH were collected from both Jewish and non-Jewish descendants. 33 of these haplotypes also contained 67 markers and will be discussed below. Figure 19 shows a 37-marker CMH haplotype tree. The left-hand side has the “recent CMH” branch and the “older CMH” branch is located at the lower right.

The “older CMH” 22-haplotype branch contains 126 mutations in just the first 25 markers and 243 mutations using all 37 markers, which results in TSCAs of  $3,575 \pm 480$  and  $3,525 \pm 420$  years from a common ancestor, respectively, or on average  $3,525 \pm 450$  years.

Similarly, the “recent CMH” branch corresponds to  $975 \pm 135$  and  $1175 \pm 140$  years to a common ancestor, respectively, or on average  $1,075 \pm 190$  ybp.

The most resolution comes from the 67-marker haplotypes, but there are fewer of these. The 33 haplotypes are shown as a 67-marker CMH tree in Figure 20.

The tree again splits into two quite distinct branches: a recent one, the 17-haplotype branch on the right, and an older one, the 16-haplotype branch on the left. Again, these are same two principal “CMH” branches, each one with its own common ancestor, who lived about 3,000 years apart. If we examine this smaller number of haplotypes (33) on both 25 markers and 37 markers, we get results consistent with those obtained above with the larger 85-haplotype set, though naturally the confidence intervals are somewhat larger. A common ancestor of the “older CMH”, calculated from 25-, and 37-marker haplotypes, lived  $4,150 \pm 580$  and  $3,850 \pm 470$  ybp, on average  $4,000 \pm 520$  ybp. A common ancestor of the “recent CMH,” also calculated from 25- and 37-marker haplotypes, lived  $975 \pm 205$  and  $1,150 \pm 180$  years to a common ancestor, respectively, on average  $1,050 \pm 190$  ybp, around the 9<sup>th</sup> to the 11<sup>th</sup> century. This coincides with the Khazarian times; however, it would be a stretch to make a claim related to this. It could also have corresponded with the time of Jewish Ashkenazi appearance in Europe.

The base (ancestral) 67-marker haplotype of the “older CMH” 16-haplotype branch (Figure 20) is as follows:

12 23 14 10 13 17 11 16 11 13 11 30 17 8 9 11 11 25  
14 20 26 12 14 16 17 11 10 22 22 15 14 18 18 32 35

12 10 11 8 15 16 8 11 10 8 11 9 12 21 22 18 10 12 12  
15 8 12 25 21 13 12 11 14 12 12 12 11

The 17-haplotype “recent CMH” branch has the following 67-marker base haplotype, with differences in bold:

12 23 14 10 13 15 11 16 12 13 11 30 17 8 9 11 11 26  
14 21 27 12 14 16 17 11 10 22 22 15 14 20 18 31 35  
13 10 11 8 15 16 8 11 10 8 11 9 12 21 22 17 10 12 12  
15 8 12 24 21 13 12 12 14 12 12 12 11

According to data provided in databases, two-thirds of the bearers of the “older CMH” (10 individuals of the 16) in the respective 67-marker branch did not claim Jewish heritage. They are descendants of people who lived in Italy, Cuba, Lebanon, Puerto-Rico, Spain, England, and France (Basque). That may explain why the “older CMH” haplotype differs in three alleles in the first 25 markers from the Jewish oCMH, shown earlier. On the other hand, 16 out of 17 haplotypes in the recent CMH branch claimed Jewish heritage, and several claimed themselves to be descendants of Cohens. Their base haplotype is identical with the Jewish rCMH shown earlier.

To study this issue further, three haplotypes of inhabitants of the Arabian Peninsula with typical Arabic names, and having the following 37-marker “CMH” haplotypes

12 23 14 10 14 17 11 16 12 12 11 29 17 8 9 11 11 25  
14 20 26 12 14 16 17 10 10 22 22 14 15 18 17 33 36  
12 10

12 23 14 10 12 16 11 16 11 13 11 29 17 8 9 10 11 25  
14 19 30 13 13 13 16 11 9 19 20 16 13 16 17 33 36  
12 10

12 23 14 10 12 16 11 16 11 13 11 29 21 8 9 11 11 26  
14 20 26 12 14 15 16 10 10 20 22 14 14 17 18 32 34  
13 9

were added to the set of haplotypes shown in Figure 19. All three Arabic haplotypes joined the lower, predominantly non-Jewish branch on the right of Figure 19. After the addition of the Arab haplotypes, all 25 of the haplotypes in the branch contained 162 mutations on 25 markers, which gives  $4,125 \pm 525$  years to a common ancestor, slightly older than age of the branch without those three haplotypes. This time period is close to that of the legendary Biblical split into the Jewish and the Arab lineages of the Abrahamic tribes, although the applicability of those events to the results of this study are uncertain, especially given that the uncertainty in the dating of Abraham is at least as large, if not larger, as the uncertainties in our TSCA values.

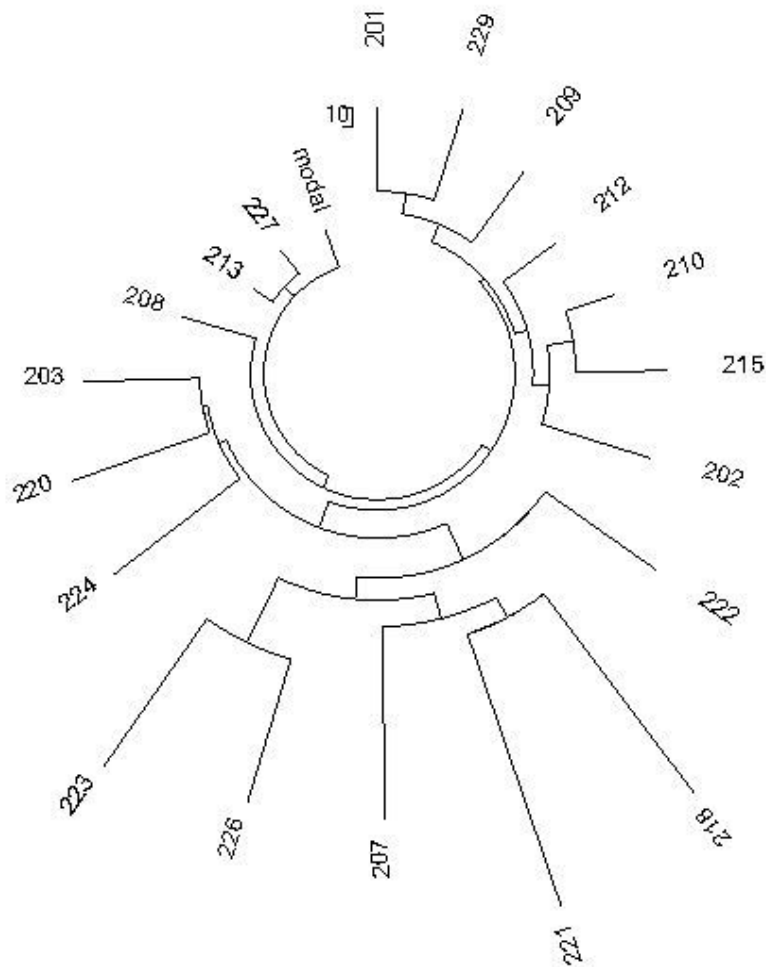


Figure 21. The 37-marker haplotype tree for Arabian haplotypes of Haplogroup J1. 19 haplotypes were listed in the Arabian Peninsula YDNA Project (2008).

The following section demonstrates that the “CMH” in fact appeared as long as 9,000 ybp or earlier on the Arabian Peninsula. The above time spans of about  $4,000 \pm 520$  or  $4,125 \pm 525$  (the “older CMH”), and  $1,050 \pm 190$  (the “recent CMH”) ybp was generated as a result of migrations of haplotype bearers from the Arabian Peninsula to the Middle East and further to the north. We can neither prove nor disprove as yet that the “recent CMH” appeared in the Khazar Khaganate between 9<sup>th</sup> and 11<sup>th</sup> centuries. At any rate, about a thousand years ago, the bearer of the base “recent CMH” became a common ancestor to perhaps millions of present-day bearers of this lineage.

#### *The Arabian “CMH” and the Arabian Peninsula*

Obviously, the name “Cohen Modal Haplotype” was a misleading one. Though, by the end of the 1990’s it had certainly attracted attention to DNA genealogy. Even as a “modal” haplotype it is not exclusively associated with a Jewish population. A haplotype tree of Haplogroup J1 Arabs from the Arabian Peninsula is shown in Figure 21. The tree is composed of 19 haplotypes, each containing 37 markers and belonging to Haplogroup J1, listed in the Arabian Peninsula YDNA J1 Project (Al-Jasmi, 2008).

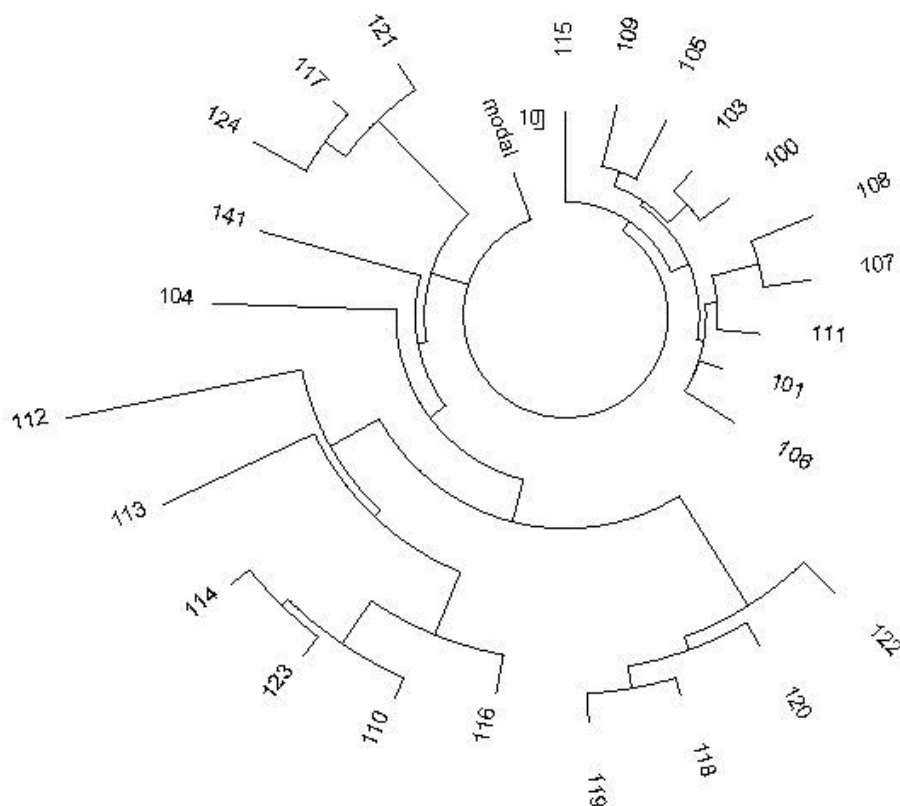


Figure 22. The 37-marker haplotype tree for J2 Jewish haplotypes. Twenty five haplotypes were collected in YSearch database (Klyosov, 2008c).

There are two branches of the tree, one of which has the “CMH” embedded and has the following ancestral (base) haplotype:

12-23-14-10-12-18-11-16-11-13-11-30-18-8-9-11-11-  
25-14-20-26-12-13-15-16-10-10-19-22-15-14-18-18-  
33-35-12-10

There are 73 mutations in 25 markers on the six -haplotype branch, or  $0.487 \pm 0.057$  mutations per marker on average, or  $9,000 \pm 1,400$  years to a common ancestor.

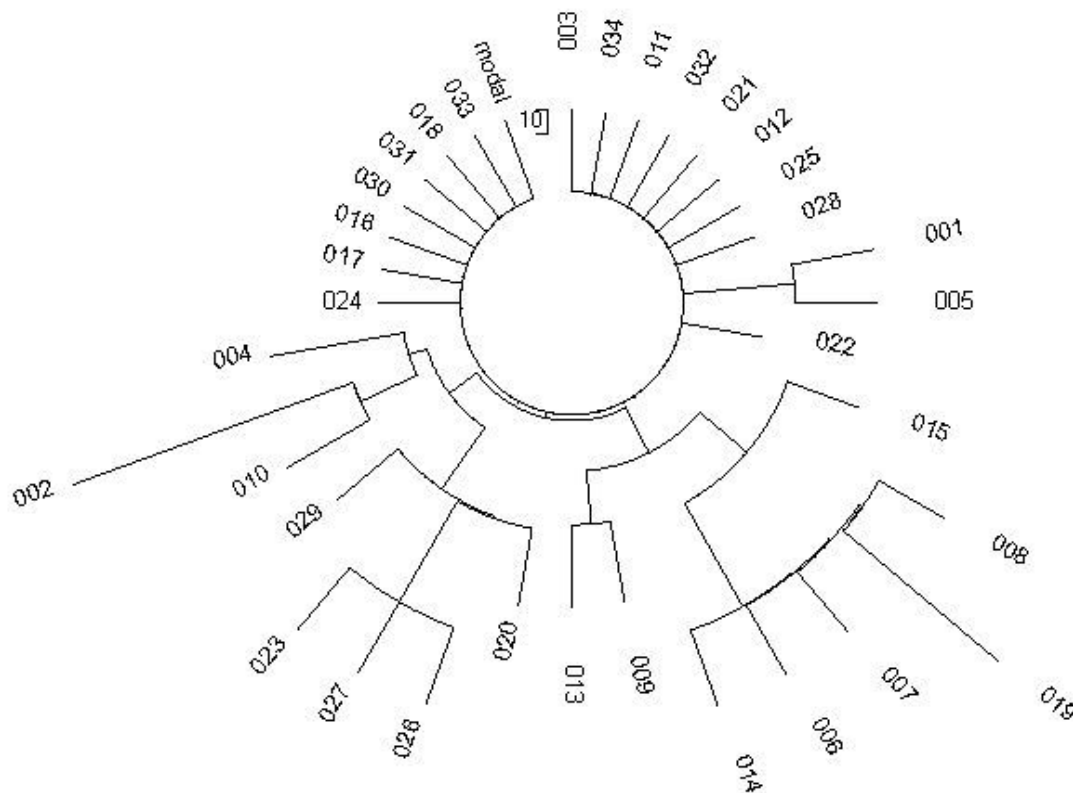
The second branch, a seven-haplotype branch (Figure 21), has a significantly “younger” common ancestor, since it is much less extended from the “trunk” of the tree. It has the following base haplotype, with markers differing from the older base haplotype shown in bold:

12-23-14-11-13-19-11-17-11-13-11-30-19-8-9-11-11-  
26-14-20-25-12-14-16-17-10-10-22-22-14-14-18-18-  
32-36-11-10

The seven haplotypes contain 27 mutations in their first 25 markers or  $0.154 \pm 0.030$  mutations per marker and  $2,300 \pm 500$  years to a common ancestor for this branch.

These two ancestral haplotypes differ by ten mutations on the first 25 markers, which corresponds approximately to 7,000 years between them. The “younger” common ancestor who lived  $2,300 \pm 500$  ybp is very likely a direct descendant of the “older” one who lived  $9,000 \pm 1,400$  ybp and had the “CMH” haplotype.

It seems that the “Cohen Modal Haplotype” likely became an ancestral haplotype for a significant fraction of the inhabitants of the Arabian Peninsula between 7600 and 10400 ybp (the 95% confidence interval) and was a common feature of the background population from which both the Arabs and Jews arose. When the ancient Israelites became a distinct group, the CMH was quite naturally in this population, though in a slightly drifted form, which gave rise to the Jewish “older CMH.” Still later, when the Israelite priesthood was



established, the CMH was again available for inclusion, though whether it actually was included from the beginning or not is still uncertain. No one knows exactly when or how the priesthood was formed—whether it indeed started with one man according to tradition, or how it expanded, whether only by patrilineal descent from the founder or, as seems more likely given the present-day variety of haplogroups within the Cohanim, by a variety of inclusion mechanisms. By around the 7<sup>th</sup> century CE, the “recent CMH” split from the “older CMH” and became the ancestral haplotype for a separate albeit recent Jewish lineage within Haplogroup J1. If we consider only “rCMH” haplotypes within this Cohanim population, a common ancestor can be identified who lived  $1,050 \pm 190$  ybp.

The Jewish “Cohen Modal Haplotype” of Haplogroup J2, that is, J2 haplotypes with the six-marker CMH embedded, represents a rather compact group of haplotypes with a recent ancestor who lived in about the 7<sup>th</sup> century CE (see below). As it was indicated above, this

This compact group of ten J2 haplotypes, located rather close to the trunk of the tree (indicating young age) as shown in Figure 22, represents the J2-CMH. Their 37-marker base haplotype is as follows:

The ten haplotypes contain 25 and 44 mutations in 25 and 37 markers, respectively. This gives  $1,450 \pm 320$  and  $1,300 \pm 230$  years to a common ancestor or  $1,375 \pm 300$  ybp when averaged. So, the common ancestor lived around the 6<sup>th</sup> to the 8<sup>th</sup> centuries CE.

The J2-CMH base haplotype differs from the J1 “recent” and “older” base (ancestral) haplotypes by 29



and 25 mutations, respectively, in their 37-marker haplotypes. This corresponds to about 11,800 and 9,600 years of mutational difference, respectively. Clearly, J1- and J2-CMH represent quite distant lineages. After all, they belong to two different haplogroups.

## The Roma of Haplogroup H1

### *Bulgarian Roma*

According to old records, the Roma arrived in Bulgaria during the Middle Ages. Haplotypes of Bulgarian Roma have been compiled from the testing of 179 males from 12 local tribes (Zhivotovski et al., 2004). All of the haplotypes were similar and apparently originated from the same rather recent common ancestor. It seems that a very narrow circle of Roma, perhaps a single tribe, came to Bulgaria some 500-700 years ago. Descendants of other unrelated tribes if included, apparently did not survive. It cannot be excluded from consideration that a few close relatives, rather than a single founder, were the patriarchs of the tribe that survived.

Let us consider both the 6- and 8-marker haplotypes, in order to understand how an elongation of haplotypes can affect data.

The most numerous tribe, the Rudari, had the following six-marker base haplotype which was represented by 62 identical haplotypes out of the total of 67 haplotypes from the tribe:

15-12-22-10-11-12

The same base haplotype was represented in 12 of 13 members tested from the Kalderash tribe, in 9 of 26 members of the Lom tribe, in 4 of 4 members of the Torgovzi ("Traders") tribe, in 20 of 29 from the Kalaidjii tribe, and in 12 of 19 from the Musicians tribe. Other haplotypes also contained very few mutations. It is obvious that the ancestral haplotype was rather "young", no older than several hundred ybp.

Overall, all 179 haplotypes of Bulgarian Roma contained 146 identical (base) six-marker haplotypes and 34 mutations compared to the base haplotype.

Considering remaining base haplotypes, this gives  $\ln(179/146)/0.0088 = 23 \pm 3$  generations or  $575 \pm 75$  years to the common ancestor for all 179 members of all the 12 tribes. Considering mutations, this gives  $34/179/0.0088 = 22 \pm 4$  generations or  $550 \pm 100$  years.

This represents a typical fit between the TSCAs obtained from the two quite different methods of calculations where there is a well defined cluster descended from one common ancestor, and again shows that there is a fundamental basis for such fit. This basis was explained in

detail in the companion paper (Part I). "Formal" calculations of the standard deviation will give about  $\pm 20\%$  for the 95% confidence interval for the second result above, that is  $550 \pm 110$  years to the common ancestor, and about the same margin of error for the first result.

Thus, the Roma show a straightforward, uncomplicated DNA genealogy family tree. Incidentally, these haplotypes belong to Haplogroup H1, which is very common in India. Beyond India it is found mainly among the Roma. The base six-marker haplotype for H1 in India is exactly the same as that for the Roma shown above. However, its common ancestor in India lived several thousand years ago. For example, the Indians of the Koya tribe have a base haplotype 15-12-22-10-11-12, with a common ancestor of  $2,575 \pm 825$  ybp, and the Indians of the Koraja tribe have a base haplotype X-Y-22-10-Z-12, with a common ancestor of  $2,450 \pm 810$  ybp. In general, an "age" of Haplogroup H1 in India determined in several datasets is  $3975 \pm 450$  ybp (Klyosov, to be published).

In the 8-marker format, the base haplotype for the Bulgarian Roma is as follows

15-12-22-10-11-12 -- 14-16

in which the last two are markers DYS389-1 and 389-2. Considering eight markers adds 25 mutations to the 179 haplotypes, bringing the total to 59 mutations. This gives  $59/179/8/0.00163 = 25 \pm 4$  generations to a common ancestor, that is  $625 \pm 100$  ybp. It is essentially the same as that for the six-marker haplotypes,  $(34/179/0.0088 = 22 \pm 4$  generations to a common ancestor, that is  $550 \pm 100$  ybp.

The 8-marker dataset has 20 fewer base haplotypes compared to those of the six-marker series, that is 126. This results in  $\ln(179/126)/0.013 = 27 \pm 4$  generations to a common ancestor, or  $675 \pm 100$  ybp. This is within the margin of error with the result for the six-marker haplotypes. As one can see,  $675 \pm 100$ ,  $575 \pm 75$  and  $550 \pm 100$  ybp fit each other quite satisfactorily.

### *Croatian Roma*

Another example of a population that apparently represents the Roma is presented by a series of 34 haplotypes from Croatia (Barac et al., 2003a, 2003b; Pericic et al., 2005). The haplotype tree composed of those haplotypes is shown in Figure 23.

The referenced articles did not specify the origin or the ethnic features of the tested individuals, however, a group of H1 bearers in Croatia will most likely be the Roma. This guess was further supported by the TSCA estimate, as follows.

Sixteen haplotypes, representing nearly half of the haplotypes, were identical, base, ancestral haplotypes:

12-22-15-10-15-17-X-Y-Z-14-11-29

Using the logarithmic method, we obtain  $\ln(34/16)/0.017 = 44$  generations to a common ancestor (without a correction for back mutations). The linear method gives  $24/34/0.017 = 42$  generations (without a correction). Clearly, all the 34 individuals have a single common ancestor who lived  $45 \pm 10$  generations ago (with a correction for back mutations), or  $1125 \pm 250$  years ago, between the 7<sup>th</sup> and 12<sup>th</sup> century CE.

### The Polynesian Haplotypes, Haplogroup C2

The Polynesians, such as the Maoris, Cook Islanders, Samoans, often have Haplogroup C2. In a published study (Zhivotovski et al., 2004), 37 ten-marker haplotypes were determined in these three populations, and the base haplotype for all of them follows (in the FTD-NA format plus DYS461):

14-20-16-10-X-X-X-15-13-12-12-30 -- 9

It is not clear from the tree, whether it is derived from one or more of common ancestors. Let us employ again the “fit” criterion of the logarithmic and the linear methods.

There were eight base haplotypes among the 37 haplotypes total and 49 mutations in the whole set with respect to the base haplotype. The average mutation rate for this haplotype equals to 0.0018 per marker per generation of 25 years (Klyosov, 2009, the companion article). Therefore, a common ancestor of all the 37 individuals lived  $49/37/10/0.0018 = 74$  generations (without a correction for back mutations), or 80 generations with the correction, which results in  $2000 \pm 350$  ybp. Eight base haplotypes (on top of the tree in Figure 24) give  $\ln(37/8)/0.018 = 85$  generations (without the correction) or 93 generations (with the correction), which equates to 2325 ybp. This is within the margin or error. The 16% difference, albeit not a dramatic one, might indicate that there was a slight admixture in the dataset by one or more extraneous haplotypes.

Incidentally, the theorized Polynesian expansion time cited by Zhivotovski (2004) was between 650 and 1,200

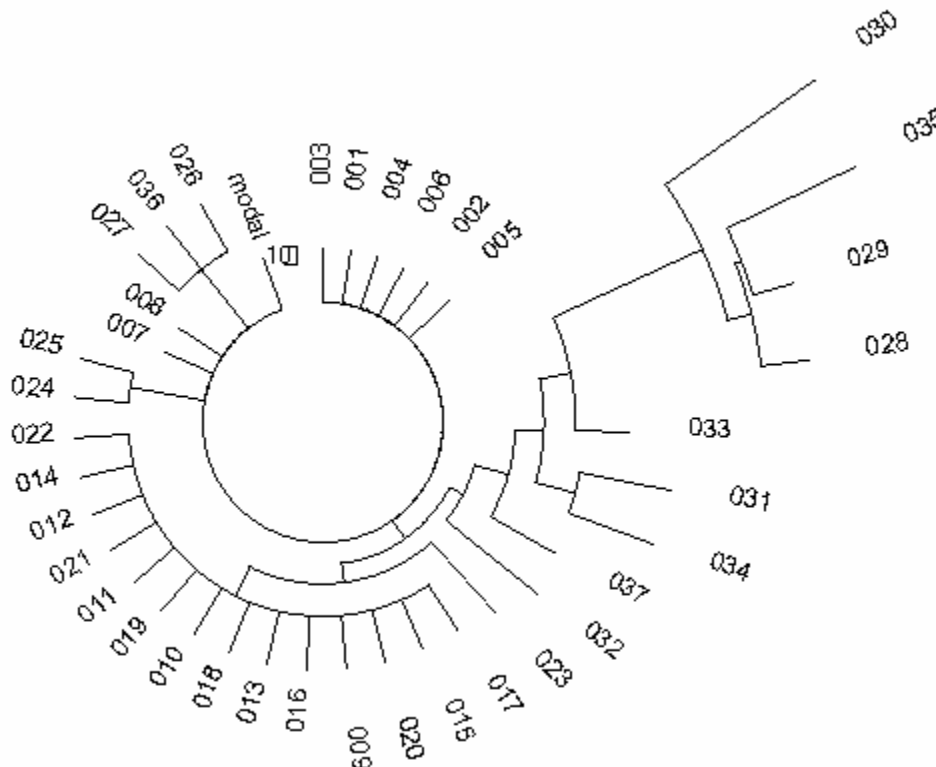


Figure 24. The ten-marker haplotype tree for 37 Polynesians of Haplogroup C2. The haplotype tree was created from data from Zhivotovsky, et al. (2004).

years ago. Either that timespan is underestimated, or the DNA-genealogy points at arrival of not one but a number of travelers to the Polynesian islands. They might have “brought” an earlier common ancestor in the DNA.

### The South African Lemba Haplotypes

Lemba is a South African tribe whose people live in Limpopo Province in Zimbabwe, in Malawi, and in Mozambique. A list of 136 Lemba haplotypes was published by Thomas et al. (2000), and the authors noted that some Lemba belong to the CMH Jewish lineage. We will demonstrate that it is very unlikely. Thomas did not publish the haplogroups for any of the individuals, only their six-marker STR haplotypes.

Of the 136 Lemba individuals, 41 had typical “Bantu” haplotypes belonging to Haplogroup E3a (by the author’s definition) with a base haplotype:

15-12-21-10-11-13

The 41 haplotypes contain 91 mutations from the above haplotype, that is  $0.370 \pm 0.039$  mutations per marker or  $8,300 \pm 1,200$  years from a common ancestor.

Another 23 Lemba who were tested had the following base haplotype :

14-12-23-10-15-14

All 23 haplotypes had only 16 mutations, which gives  $2,150 \pm 580$  years to a common ancestor for these individuals.

There were a few scattered Lemba haplotypes, apparently from different unidentified haplogroups, and finally there were 57 haplotypes of apparently Haplogroup J, which in turn split into three different branches as shown in Figure 25. Three base haplotypes, one for each branch, are shown below:

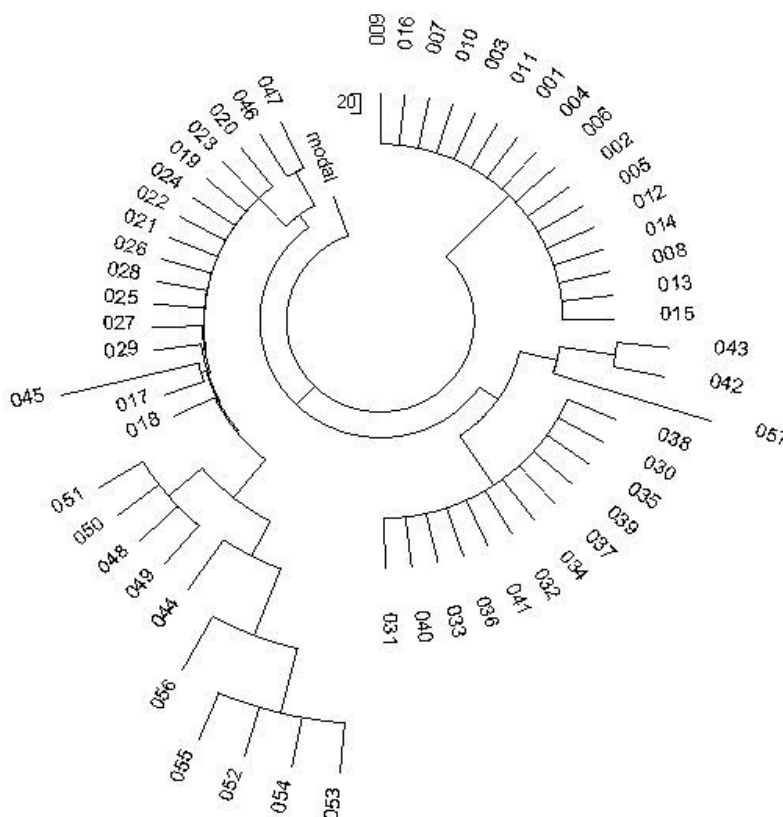


Figure 25. The Lemba six-marker haplotype tree, apparently of Haplogroup J. The 57-haplotype tree is composed of data published in Thomas et al. (2000).

14-16-24-10-13-12  
 14-15-24-10-11-12  
 14-16-23-10-11-12

The first one represents 16 identical haplotypes (the upper right area in Figure 24), which obviously came from a very recent common ancestor. As one can see from the haplotype tree, none of these haplotypes is mutated. Its common ancestor should have lived no more than a few centuries ago.

The second one, being a base haplotype for a 26-haplotype branch on the left-hand side in Figure 25, is a rather common haplotype in the Arabic world, and belongs likely to Haplogroups J2, but possibly to J1. The branch contains 21 mutations, which gives  $2,550 \pm 610$  years to a common ancestor, who most likely lived in the first millennium BCE. It is clearly not the "Cohen Modal Haplotype" and differs from the third haplotype by two mutations, which in the six-marker format corresponds to about 7300 years.

The third base haplotype, which is the CMH in its six-marker format, supports a branch of 15 haplotypes on the lower right-hand side of Figure 24. Twelve of those CMH haplotypes are identical to each other and form a flat branch. There are no mutations in them, and they must have come from a very recent ancestor of only a few centuries ago. From a fraction of the base haplotype, their common ancestor lived only  $\ln(15/12)/0.0088 = 25$  generations ago, or about  $625 \pm 200$  ybp, around the 14<sup>th</sup> century.

The three mutated haplotypes in this series are quite different from the CMH, and apparently do not belong to the same group of haplotypes. All of them have two or four mutations from the CMH:

14-23-14-10-11-12  
 14-23-14-10-11-12  
 16-24-14-10-11-12

Unfortunately, more extended haplotypes are not available. It is very likely that they are rather typical mutated Arabic haplotypes. Besides, it is not known with certainty to which haplogroup they belong, J1 or J2, though the CMH is the modal haplotype for J1.

Obviously, to call the Lemba haplotypes the "Cohen haplotype" is a huge stretch in regard to the Cohen reference. They could have been Jewish and originated just a few centuries ago, or they could have been Arabic. Since the CMH on six markers is the modal haplotype for all of Haplogroup J1, automatically attributing a group of CMH haplotypes to the Cohanim is misleading. Hence, the so-called "Cohen Modal Haplotype" in the "Black Jews of Southern Africa" has nothing to do

with an ancient history of either the Lemba or the Jewish people. It is a rather recent development.

## Conclusion

The present article (Part II) and the companion article (Part I) present a rather simple and self-consistent approach to dating both recent and distant common ancestors based upon appearance of haplotype trees. It provides an approach to a verification of haplotype sets in terms of a singular common ancestor for the set, or to determining that they represent a multiplicity of common ancestors (through use of the "logarithmic method"), as well as a way to calculate a time span to a common ancestor corrected for reverse mutations and asymmetry of mutations. Obviously, time spans to common ancestors refer to those ancestors whose descendants survived and present their haplotypes and haplogroups for analysis in the present. Naturally, their tribes or clans could have appeared earlier, since it is very likely that in many cases offspring and/or descendants did not survive.

Amazing as it is, our history is written in the DNA of each of us, the survivors. The "writing" represents just a "scribble on the cuff" of the DNA, though there is a deeper meaning to these scribbles. If we look at them for a single individual, without comparisons to others, they do not say much. They represent just a string of numbers. However, when compared with those in other people, these scribbles start to tell a story. These collective stories are about origins of mankind, appearances of tribes, their migrations, about our ancestors, and their contributions to current populations. This study is intended to contribute quantitative methods for studying the field of DNA genealogy.

## Acknowledgements

I am indebted to Dr. A. Aburto for providing a series of the "Cohen Modal Haplotypes", to Dr. A.H. Bittles for providing a series of haplotypes from China, to Theresa M. Wubben for valuable discussions.

## Web Resources

Arabian Peninsula J1 Project

[www.familytreedna.com/public/Arabian\\_YDNA\\_J1\\_Project/default.aspx](http://www.familytreedna.com/public/Arabian_YDNA_J1_Project/default.aspx)

Ashina Project

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

Basque DNA Project

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

Haplogroup Q Project

<http://m242.haplogroup.org>

India (regional) DNA Project

<http://www.familytreedna.com/public/India/default.aspx>

ISOGG 2009 Y-DNA Haplogroup Q and its Subclades

[http://www.isogg.org/tree/ISOGG\\_HapgrpQ09.html](http://www.isogg.org/tree/ISOGG_HapgrpQ09.html)

R-312 and Subclades Project

<http://www.familytreedna.com/public/atlantic-r1b1c/default.aspx>

R-L21 Project

<http://www.familytreedna.com/public/R-L21/default.aspx?publicwebsite.aspx>

R1b (U152+) Project

<http://www.familytreedna.com/public/R1b1c10/default.aspx>

R-M222 Project

<http://www.familytreedna.com/public/R1b1c7/default.aspx>

## References

- Adamov DS, Klyosov AA (2008b) Evaluation of an "age" of populations from Y chromosome using methods of average square distance (in Russian) *Proc Russian Academy of DNA Genealogy* (ISSN 1942-7484), 1:855-905.
- Al-Jasmi J (2008) The Arabian Y-DNA J1 Project. See **Web Resources**.
- Barać L, Perićić M, Klarić IM, Janićijević B, Parik J, Rootsi S, Rudan P (2003a) Y chromosome STRs in Croatians. *Foren Sci. Int.* 138:127-133.
- Barać L, Perićić M, Klarić IM, Rootsi S, Janićijević B, Kivisild T, Parik J, Rudan I, Villems R, Rudan P (2003) Y-chromosomal heritage of Croatian population and its island isolates. *Eur J Hum Genet.* 11:535-542.
- Bittles AH, Black ML, Wang W (2007) Physical anthropology and ethnicity in Asia: the transition from anthropology to genome-based studies. *J Physiol Anthropol.* 26:77-82.
- Bortolini M-C, Salzano FM, Thomas MG, Stuart S, Nasanen SPK, Bau CHD, Hutz MH, Layrisse Z, Petzl-Erler ML, Tsuneto LT, et al. (2003) Y-chromosome evidence for differing ancient demographic histories in the Americas. *Am J Hum Genet.* 73:524-539.
- Bouakaze C, Keyser C, Amory S, Crubezy E (2007) First successful assay of Y-SNP typing by SNaPshot minisequencing on ancient DNA. *Int J Legal Med.* 121:493-499.
- Bradley B, Stanford D (2004) The North Atlantic ice-edge corridor: a possible Paleolithic route to the New World. *World Archaeology*, 36:459-478.
- Cadenas AM, Zhivotovsky LA, Cavalli-Sforza LL, Underhill PA, Herrera RJ (2008) Y-chromosome diversity characterizes the Gulf of Oman. *Eur J Hum Genet.* 18:374-386.
- Campbell KD (2007) Geographic patterns of haplogroup R1b in the British Isles. *J Genet Geneal.* 3:1-13.
- Cervantes A (2008) Basque DNA Project. See **Web Resources**.
- Fagundes NJR, Kanitz R, Eckert R, Valls ACS, Bogo MR, Salzano FM, Smith DG, Silva WA, Zago MA, Ribeiro-dos-Santos, et al (2008) Mitochondrial population genomics supports a single pre-Clovis origin with a coastal route for the peopling of the Americas. *Am J Hum Genet.* 82:583-592.
- Fornarino, S, Pala, M, Battaglia, V, Maranta, R, Achilli, A, Modiano, G, Torroni, A, Semino, O, Santachiara-Benerecetti, S.A (2009) Mitochondrial and Y-chromosome diversity of the Tharus (Nepal): a reservoir of genetic variation. *BMC Evol Biol.* 2009:9:154.
- Haak W, Brandt G, de Jong HN, Meyer C, Ganslmeier R, Heyd V, Hawkesworth C, Pike AWG, Meller H, Alt KW (2008) Ancient DNA, strontium isotopes, and osteological analyses shed light on social and kinship organization of the later Stone Age. *Proc Natl Acad Sci. US* 105:18226-18231.
- Haynes G (2002) *The Early Settlement of North America: The Clovis Era*. Cambridge University Press, p. 52.
- Karlsson AO, Wallerström T, Götherström A, Holmlund G (2006) Y-chromosome diversity in Sweden – A long-time perspective. *Eur J Hum Genet.* 14:863-970.
- Kerchner C (2008) R1b (U152+) project. See **Web Resources**.
- Keyser C, Bouakaze C, Crubezy E, Nikolaev VG, Montagnon D, Reis T, Ludes B (2009) Ancient DNA provides new insights into the history of south Siberian Kurgan people. *Hum Genet*, published online 16 May 2009.
- Kivisild T, Rootsi S, Metspalu M, Mastana S, Kaldma K, Parik J, Metspalu E, Adojaan M, Tolk HV, Stepanov V, et al. (2003) The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. *Am J Hum Genet.* 72:313-332.

- Klyosov AA (2008a) The features of the "West European" R1b haplogroup (in Russian). *Proc Russian Acad DNA Geneal* (ISSN 1942-7484), 1:568-629.
- Klyosov, A.A (2008b) Where is the origin of the Slavs and Indo-Europeans? DNA genealogy gives an answer (in Russian). *Proc Russian Acad DNA Geneal* (ISSN 1942-7484), 1:400-477.
- Klyosov AA (2008b) Origin of the Jews via DNA genealogy. *Proc Russian Acad DNA Geneal* (ISSN 1942-7484), 1:54-232.
- Klyosov, A.A (2009a) Iberian haplotypes, and analysis of history of populations of Basques, Sephards and other groups in the Iberian peninsula. *Proc Russian Acad DNA Geneal* (ISSN 1942-7484), 2:390-421.
- Klyosov, A.A (2009b) Haplotypes of Eastern Slavs: nine tribes? *Proc Russian Acad DNA Geneal* (ISSN 1942-7484), 2:232-251.
- Lepper, B.T (1999) Pleistocene peoples of midcontinental North America, in: Bonnicksen R. and Turnmire K, eds, *Ice Age People of North America*. Oregon State University Press, pp. 362-394.
- McEvoy B, Simms K, Bradley DG (2008) Genetic investigation of the patrilineal kinship structure of early Medieval Ireland. *Am J Phys Anthropol*, 136:415-22.
- McEvoy B, Bradley DG (2006) Y-chromosomes and the extent of patrilineal ancestry in Irish surnames. *Hum Genet*, 119:212-219.
- Mertens G (2007) Y-Haplogroup frequencies in the Flemish population. *J Genet Geneal*, 3:19-25.
- Peričić M, Lauc LB, Klaric IM, Rootsi S, Janicijević B, Rudan I, Terzić R, Colak I, Kvesic A, Popovic D, Sijacki A, Behluli I, Thorthovic D, Efremovska L, Bajec ED, Stefanovic BD, Vilems R, and Rudan P (2005) High resolution phylogenetic analysis of southeastern Europe (SEE) traces major episodes of paternal gene flow among Slavic populations. *Mol Biol Evol*, 22:1964-1975.
- Roewer L, Willuweit S, Krüger C, Nagy M, Rychkov S, Morozowa I, Naumova O, Schneider Y, Zhukova O, Stoneking M, Nasidze I (2008) Analysis of Y chromosome STR haplotypes in the European part of Russia reveals high diversities but non-significant genetic distances between populations. *Int J Legal Med*, 122:219-23.
- Rutledge C (2009) India (regional) DNA Project (including Pakistan, Bangladesh, Nepal, Bhutan). See Web Resources.
- Sahoo S, Singh A, Himabindu G, Banerjee J, Sitalaximi T, Gaikwad S, Trivedi R, Endicott P, Kivisild T, Metspalu M, et al. (2006) A prehistory of Indian Y chromosomes: evaluating demic diffusion scenarios. *Proc Nat Acad Sci US*, 103:843-848.
- Seielstad M, Yuldasheva N, Singh N, Underhill P, Oefner P, Shen P, Wells RS (2003) A novel Y-chromosome variant puts an upper limit on the timing of first entry into the Americas. *Am J Hum Genet*, 73:700-705.
- Semino O, Passarino G, Oefner PJ, Lin AA, Arbuzova S, Beckman LE, De Benedictis G, Francalacci P, Kouvasi A, Limborska S, Marcikiae M, Mika A, Mika B, Primorac D, Santachiara-Benerecetti AS, Cavalli-Sforza LL, Underhill PA (2000) The genetic legacy of paleolithic *Homo sapiens sapiens* in extant Europeans: A Y chromosome perspective. *Science*, 90:1155-1159.
- Sengupta S, Zhivotovsky LA, King R, Mehdi SQ, Edmonds CA, Chow CT, Lin AA, Mitra M, Sil SK, Ramesh A, Rani MVU, Thakur CM, Cavalli-Sforza LL, Majumder PP, Underhill PA (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. *Am J Hum Gen*, 78:202-221.
- Sharma S, Rai E, Sharma P, Jena M, Singh S, Darvishi K, Bhat AK, Bhanwer AJS, Tiwari PK, Bamezai RNK (2009) The Indian origin of paternal haplogroup R1a1\* substantiates the autochthonous origin of Brahmins and the caste system. *J Hum Genet*, 54:47-55.
- Stevens R, Tilroe V (2009) R-P312 and subclades project (October 2009). See Web Resources.
- Stevens R, Secher B, Walsh M (2009) R-L21 project (October 2009). See Web Resources.
- Thanseem I, Thangaraj K, Chaubey G, Singh VK, Bhaskar LV, Reddy MB, Reddy AG, Singh L (2006) Genetic affinities among the lower castes and tribal groups of India: Inference from Y chromosome and mitochondrial DNA. *BMC Genet*, 2006 Aug 7;7(1):42.
- Thomas MG, Skorecki K, Ben-Ami H, Parfitt T, Bradman N, Goldstein DB (1998) Origins of Old Testament priests. *Nature*, 394:138-140.
- Thomas MG, Parfitt T, Weiss DA, Skorecki K, Wilson JF, le Roux M, Bradman N, Goldstein DB (2000) Y Chromosomes traveling South: the Cohen Modal Haplotype and the origin of the Lemba - the "Black Jews of Southern Africa." *Am J Hum Genet*, 66:674-686.
- Underhill, PA, Myres, NM, Rootsi, S. et al. (2009) Separating the post-Glacial coancestry of European and

Asian Y chromosomes within Haplogroup R1a. *Eur J Hum Genet* On-line publication, doi: 10.1038/ejhg.2009.194.

[Wells RS, Yuldasheva N, Ruzibakiev R, Underhill PA, Evseeva I, Blue-Smith J, Jin L, Su B, Pitchappan R, Shanmugalakshmi S, Balakrishnan K, Read M, Pearson NM, Zerjal T, Webster MT, Zholoshvili I, Jamarjashvili E, Gambarov S, Nikbin B, Dostiev A, Aknazarov O, Zalloua P, Tsoy I, Kitaev M, Mirrakhimov M, Chariev A, Bodmer WF \(2001\) The Eurasian heartland: a continental perspective on Y-chromosome diversity. \*Proc Natl Acad Sci \(USA\)\*, 98:10244–10249.](#)

Weston D, Corbett G, Maddi M (2009) YDNA Haplogroup R1b-U106/S21+ Research Group. See **Web Resources**.

Wilson D, McLaughlin J (2009) R-M222 haplogroup project. See **Web Resources**.

[Zhivotovsky LA, Underhill PA, Cinnoglu C, Kayser M, Morar B, Kivisild T, Scozzari R, Cruciani F, Destro-Bisol G, Spedini G, et al. \(2004\) The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. \*Am J Hum Genet\*, 74:50-61.](#)