DNA Genealogy, Mutation Rates, and Some Historical Evidences Written in Y-Chromosome. II. Walking the Map

Anatole A. Klyosov[1]

Abstract

Employing the methodology developed in Part I of this study, haplotypes of the following populations have been considered (TSCA, or time spans to [the most recent] common ancestors, shown as years before the present):

-- Basques/Iberian R1b1b2, 3625±370,

-- Ireland R1b1b2, 3800±380 and 3350±360, two extended haplotype series

-- Ireland R1b1b2 ("younger Irish"), 2750±290

-- British R1b, 2400±250

-- Flemish R1b, 4150±500

-- Swedish R1b1b2, 4225±520

-- Ireland I2 (9600±1200)

-- the Isles I1 (England, Ireland and Scotland), 3,475±350

-- NW Europe/Scandinavia I1 (Denmark, Sweden, Norway, Finland), 3375±345

-- Central Europe I1 (Austria, Belgium, Netherland, France, Italy, Spain, Switzerland), 3425±350

[1]Anatole Klyosov, 36 Walsh Road, Newton, Massachusetts 02459, USA

aklyosov@comcast.net; Phone (617)785-4548; Fax (617)964-4983

-- Eastern Europe (Poland, Ukraine, Belarus, Lithuania, Russia), 3225±360

-- Middle Eastern I1 (Jordan, Lebanon, Jewish populations), 3725±500

-- England R1a1, 4125±475

-- Ireland R1a1, 3850±460

-- Scotland R1a1, 3550±450

-- Germany R1a1, 4700±520

-- Norway R1a1, 3375±490

-- Sweden R1a1, 3825±520

-- Poland R1a1, 4550±520

-- Czechoslovakia R1a1, 4125±430

-- Seventeen other European countries R1a1, 4425±520

-- Russia R1a1, 4725±520

-- India R1a1, 4300±560

-- Armenia R1a1, 4500±1040

-- Anatolia R1a1, 3700±550

-- the Arabian Peninsula R1a1, 3,750±825

-- South India Chenchu R1a1, 3,200±1900 and 350±350, two lineages),

-- bearers of the J1 Cohen Modal Haplotype (CMH), 4000±520 and 1050±150, two lineages (”older J1-CMH” and “recent J1-CMH”)

-- bearers of the J2-CMH, 1375±300

-- Cohanim J2 (not the CMH), 3000±560

-- Arabian J1-CMH, 9,000±1400

-- Bulgarian Gypsies H1, 550±110

-- Croatian Gypsies H1, 1125±250

-- Polynesians C2, 800±260

-- South African Lemba, 625±200

-- the oldest Balkan R1a1 population, 11,650±1,550

-- Native Americans Q-M3, 16,000±3,300

Some of these findings are supported by independent estimates. The data show that (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 are no indications regarding “Ukrainian refuge” for R1a1 ancient population allegedly 15,000 years ago; instead, evidences have been obtained that the oldest R1a1 lived circa 11,600± years ago on the Balkans (Serbia, Kosovo, Bosnia, Macedonia), (c) except the Balkans, present-day bearers of R1a1 across Western and Eastern Europe have common ancestors who lived between 3550 and 4725 years ago (the “youngest” in Scotland, Ireland and Sweden, the “oldest” in Russia (4725±520 ybp) and Germany (4,700±520 ybp),

(d) the Indian R1a1 haplotypes show a good match with the Russian Slavic R1a1 ones, having a common ancestor several hundred years ”younger” compared to the Russian one (4300±560 vs. 4,725±520 years bp); this supports a concept of proto-Slavic migration to India as Aryans (according to classic ancient Indian literature) around 3600 years before present,

(e) South India Chenchu R1a1 match the current Russian Slavic R1a1 haplotypes, and their (Chenchu R1a1) common ancestor appeared some 3200±1900 ybp, apparently after

the R1a1 migration from the North to India; another Chenchu R1a1 lineage was originated 350±350 ybp, around the 17th century AD,

(f) the so-called Cohen Modal Haplotype J1 was originated 9,000±1,400 years ago; about 4,000±520 years ago it appeared in the emerging Jewish population, and 1,050±190 years ago (if to consider only CMH) or 1,400±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±300 years ago,

(h) South African Lemba population of haplogroup J has nothing to do with ancient Jewish patriarchs, since the haplogroup penetrated Lemba population some 625±200 years ago, around the 14th century AD,

(i) Native American Q1a3a contains at least six lineages, the oldest one has appeared 16,000±3,300 years ago, in accord with archaeological data.

Results and discussion

The Basques and Iberian R1b1 haplotypes

As it was shown in the preceding paper (Part I), a common ancestor of the Basques of R1b1b2 haplogroup lived 3,625±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

Since all the 750 of 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 preceding paper) to the common ancestor, that is 3900 ybp (by the "logarithmic" method). It is only 7.6% higher than the above figure, 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±450 and 4,450±460 years before present (Klyosov, 2008a) with the 95% confidence interval. As it was considered in detail in the preceding paper (Part I), these relatively low standard deviations are resulting from a large number of alleles in each of the considered haplotype series, such as 750 of 19-marker Iberian R1b1 haplotype series (14,250 alleles), 184 of 25-marker U152 haplotypes (4,600 alleles), and 197 of 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 – were equal to 2.00%, 2.84%, and 2.73%, respectively,

and the SD of the mutation rates for the employed 12-, 19- and 25-marker haplotypes were equal to 10% at the 95% confidence, as it was discussed in Part I. Naturally, for less numerous haplotype series the standard errors and standard deviations are significantly higher, as it is 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 figures in the Basque base haplotype are 14.53 and 18.35, respectively.

The data show that all the tree populations, including the Basques, are likely to be descendents 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 thousand 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 years bp), nor did they make sophisticated cave paintings in South of France, Spain, and Portugal. Those people were likely bearers of haplogroup I2. However, this theory remains to be proven. Arguably, the Basque ancient and unique language was brought to Iberia around 3600 ybp by the R1b1b2 bearers from their place of preceding location(s) and/or their origin, presumably in Asia, which they had left thousands of years ago. This question, 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, 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[1]-389[2]-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 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

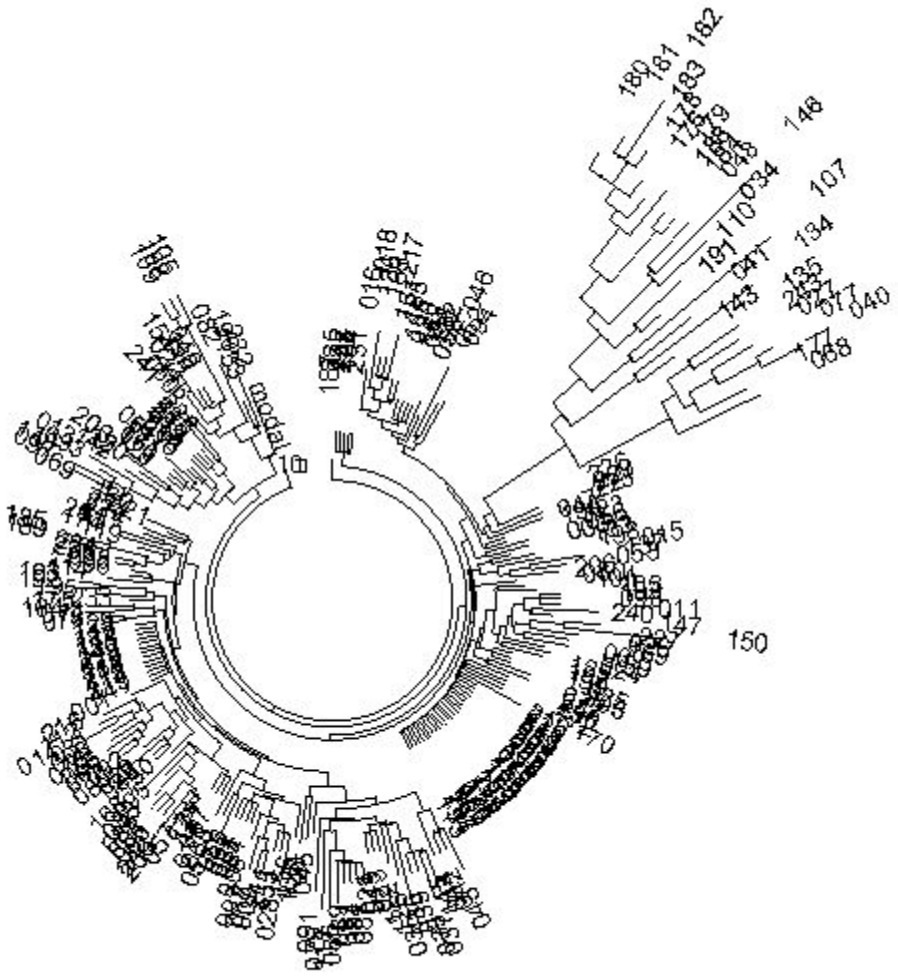

Fig. 1. A tree of 243 Irish 19-marker haplotypes (haplotypes were published in McEvoy et al, 2008). The distinct branch of the right belongs to I2 haplogroup.

It differs by only two mutations (shown in bold), and its assumed common ancestor lived 12,800±2,600 ybp (Klyosov, 2009). All 25 of the Irish I2 haplotypes contain 199 mutations, which bring their common ancestor to 9,600±1,200 ybp.

Since a difference in two mutations in two 19-marker haplotypes approximately corresponds to a mutational distance between them of about 1900±1300 years, it brings

the two ancestral I2 haplotypes (in the Iberia and on the Isles) into a rather close timewise proximity. 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±1,170 ybp (Klyosov, 2009)

R1b1 haplotypes

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 Fig. 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:

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 4375±450 ybp, and for R1b1b2 haplogroup with a common ancestor of 4,450±460 ybp (Klyosov, 2008a). Hence, they all are likely to have a rather recent origin, with "recent" is compared to some widespread expectations in the literature.

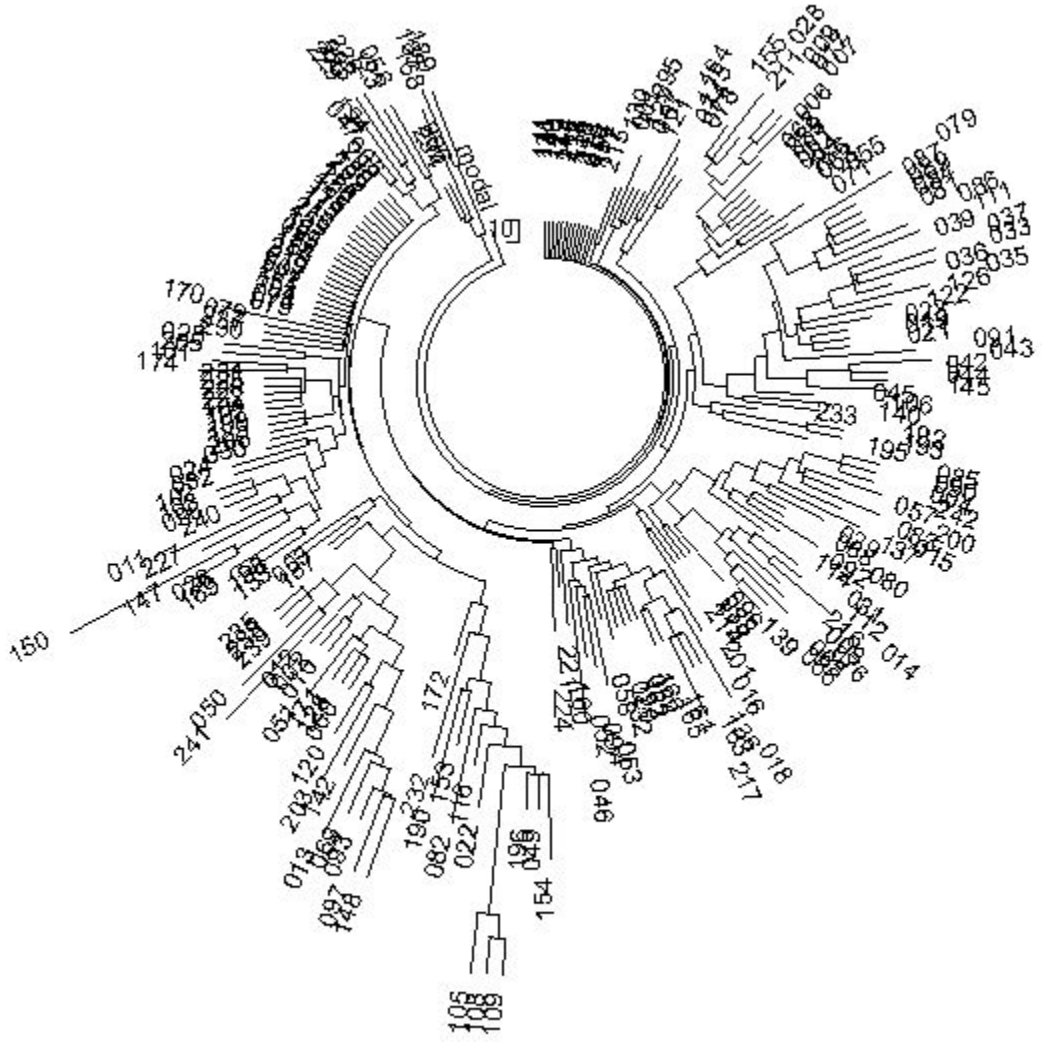

Fig. 2. A tree of 218 Irish 19-marker haplotypes (haplotypes were published in McEvoy et al, 2008), presumably of R1b1 haplogroup.

All the 218 Irish R1b1 haplotypes have 731 mutations in their 4,142 alleles, which brings their common ancestor to 3350±360 ybp. The average number of mutations per marker is slightly lower in the Ireland R1b1 haplotypes (0.176±0.007) compared to the Iberian ones (0.196±0.004); the mutation rate was the same in the calculations of the 19-marker haplotypes. It appears that the R1b1 bearers had come to the both areas fairly later compared with time of their inhabiting the continent or elsewhere, where the common ancestor of R1b1b2 lived 4,450±460 ybp. It might appear that they had arrived to the Isles somewhat later compared to that to 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±0.004 (the observed number of mutations per marker, without the corrections) the corrections brought the figure to 0.218±0.004 (corrected for back mutations) and further to 0.217±0.004 (corrected for asymmetry of mutations). The degree of asymmetry of the Irish 188 haplotypes was 0.54, that is the respective figures will be 0.176±0.007, 0.192±0.007 and 0.190±0.007. Hence, the Iberian 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 DYS385a=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 Fig. 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±290 years to the common ancestor.

This dating actually matches pretty 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 (withour correction) or 96 generations with the correction, that is 2400±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 of DYS434 had the allele of "9" (98% of total), while in the preceding series of 218 Irish haplotypes 216 of them had "11" (99% of total) in the same very locus.It seems that the authors simply changed their notation of haplotypes. The same situation was with DYS461, which had "12" in 189 of 218 alleles (87%), while in the larger series it was "10" in 833 events 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±0.003 for the average number of mutations per marker. This is practically identical with 0.196±0.004 for the Iberian R1b1 haplotypes. The degree of asymmetry for the larger series is exactly the same (0.54) as in the 118-haplotype series, and cannot shift the number of mutations per marker, hence, the "age" of the common ancestor stays at 3,800±380 ybp. Therefore, the Irish and the Iberian R1b1 haplotypes (3,625±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 not a standard format of markers was employed (in terms of the FTDNA), in which DYS426 and DYS388 were not types, and DYS437 and DYS438 were added to the series. 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

All the 64 haplotypes contained 215 mutations, which results in 4150±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 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±520 years to the common ancestor.

The “ages” of the Irish (3350±360 ybp and 3800±380 ybp), Iberian (3625±370 ybp), Flemish (4150±500 ybp) and Swedish (4225±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.

The Isles (England, Ireland and Scotland) I1 haplotypes

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 in their majority descended from the same common ancestor, who had the following haplotype:

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±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±0.003 mut/marker, which results in 139±14 generations, that is 3475±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±14, 151±16 and 131±15 generations, that is 3400±350, 3775±400, and 3275±375 ybp. Indeed, their averaged value equals to 139±10 generations, and the obtained dispersion shows that the calculated standard deviations (based on 10% standard deviation for the 95% confidence interval) are reasonable. Hence, the time span to the common ancestor of 3475±350 years is a reliable estimate for more than 1500 English, Irish and Scottish individuals.

This is a rather common TSCA for European I1 populations, and for the North-Western European/Scandinavian (Denmark, Sweden, Norway, Finland) combined series of haplotypes the TSCA equals to 3375±345 ybp, for the Central and South Europe (Austria, Belgium, Netherlands, France, Italy, Switzerland, Spain) it equals to 3425±350 ybp, for the East European countries (Poland, Ukraine, Belarus, Russia, Lithuania) it is equal to 3225±360 ybp (to be published). It is of interest that even Middle Eastern I1 haplotypes (Jordan, Lebanon, and presumably Jewish ones) descend from the common ancestor who lived at about the same time, 3725±500 ybp (to be published).

The ASD methods gave a noticeably higher figures fot the TSCAs for all four populations, that is 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±11 generations, which can be compared to 139±10 generations, calculated by the “linear” method (see above). As it was pointed out in the preceding paper (Part I) 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.

"Eurasian" haplogroup R1a1 from the Atlantics to Russia, India, and the Pacific Ocean, and from Scandinavia to the Arabian Peninsula

The "mapping" of the enormous territory outlined in this heading reveals that it is all marked with practically the same ancestral haplotype, which is about 4,500 - 4,700 years "old". An exception is presented only in the Balkans (Serbia, Kosovo, Macedonia, Bosnia), where the common ancestor is significantly more ancient, about 11,650±1,550 years bp. This will be explored below in this section. Another exception is presented by Irish, Scottish and Swedish R1a1 populations, which have a significantly "younger" common ancestor, some thousands years "younger" compared with those, e.g., of the Russian, the German, and the Poland R1a1 populations.

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 closeness to present-day R1a1 haplotypes, as it is shown below.

England and Ireland R1a1 haplotypes

The 57 of 25-marker haplotype series of England origin (YSearch database) contains ten haplotypes which 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±475 years to a common ancestor in the 95% confidence interval. The respective haplotype tree is shown in Fig. 3.

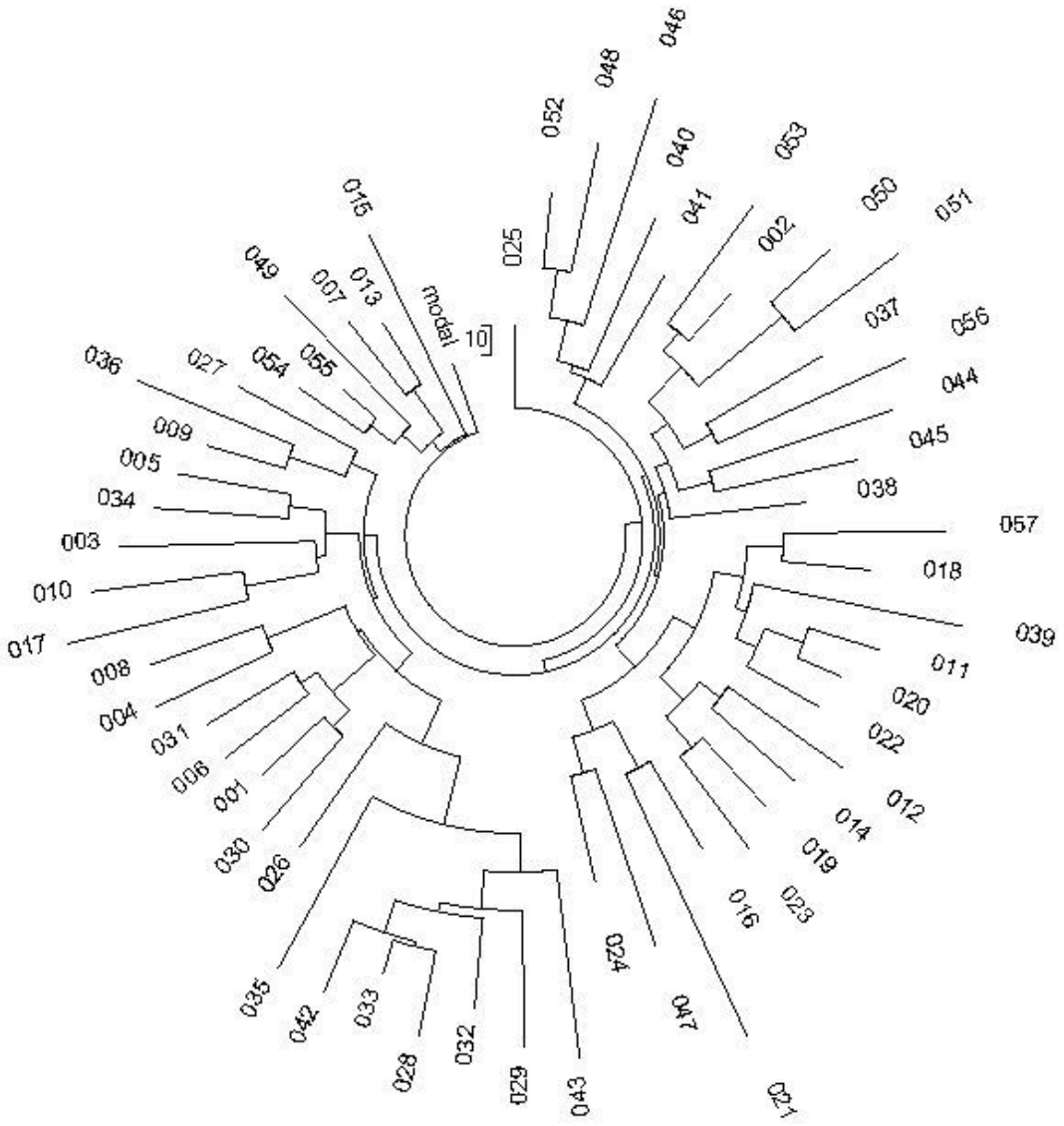

Figure 3. The 25-marker haplotype tree for England, haplogroup R1a1. The 57-haplotype tree was composed from data of YSearch database. A seven-haplotype branch at the bottom (between 035 and 043) plus haplotypes 001, 006 and 030 is a family of haplotypes with DYS388=10 (all other mostly have DYS388=12,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

The 52 of 25-marker haplotype series of Ireland origin (YSearch database) contains 12 haplotypes which belong to a DYS388=10 distinct series (Fig. 4), and was analyzed separately. The remaining 40 haplotypes contain 244 mutations compared to the base haplotype, shown below, which corresponds to 3,850±460 years to a common ancestor.

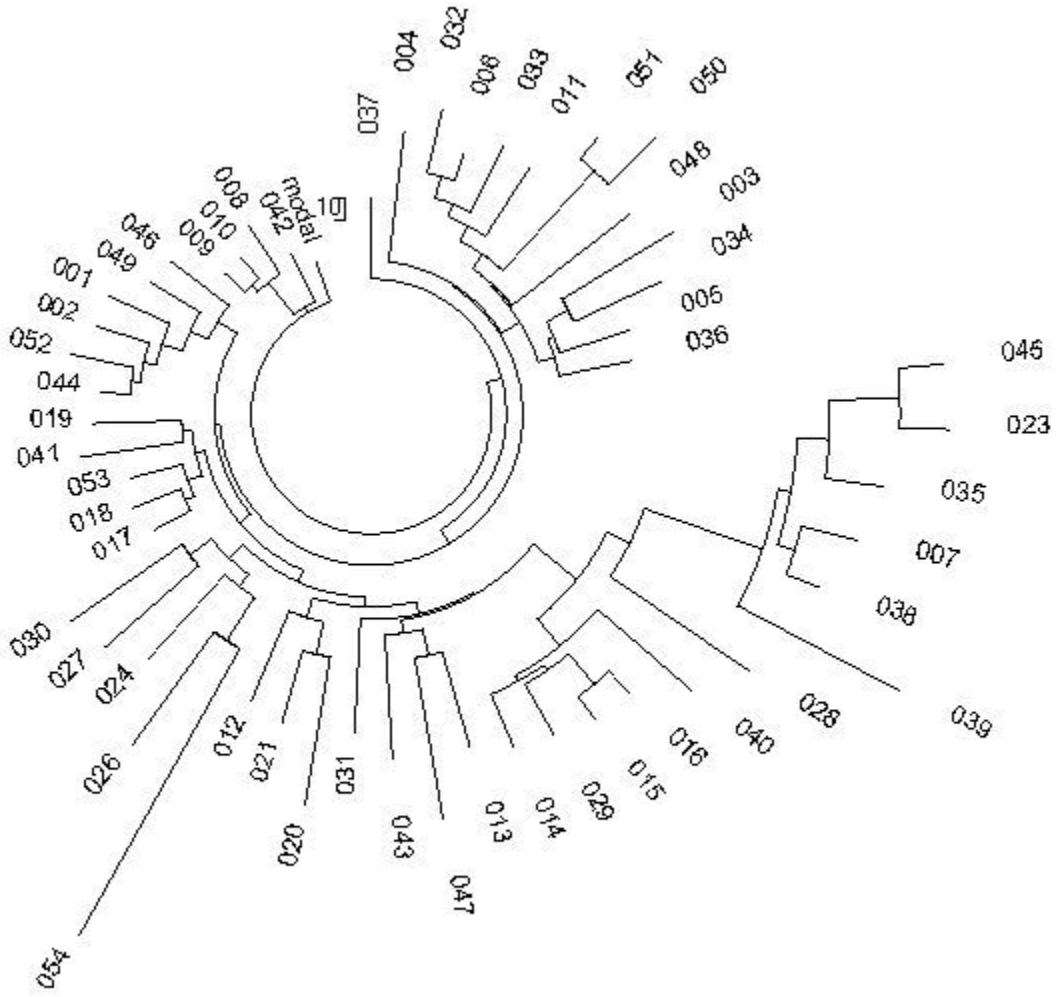

Figure 4. The 25-marker haplotype tree of haplogroup R1a1 for Ireland. The 52-haplotype tree was composed from the YSearch database. A twelve-haplotype branch at the bottom left (between 014 and 045) is a family of haplotype with DYS388=10 (all others primarily have DYS388=12, in two cases DYS388=14,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

Thus, R1a1 haplotypes sampled on the British Isles point at English and Irish common ancestors who lived 4,125±475 and 3,850±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 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 haplotypes have a mutated allele in eighth position in the FTDNA format (DYS388=12→10), with a common ancestor of that population who lived 3,575±450 years ago (172 mutations in 30 of 25-marker haplotypes with DYS388=10). 61 of these haplotypes were pooled from a number of European populations (Fig. 5), and the tree splits into a relatively younger branch on the left, and the "older" branch on the lower right-hand

side.

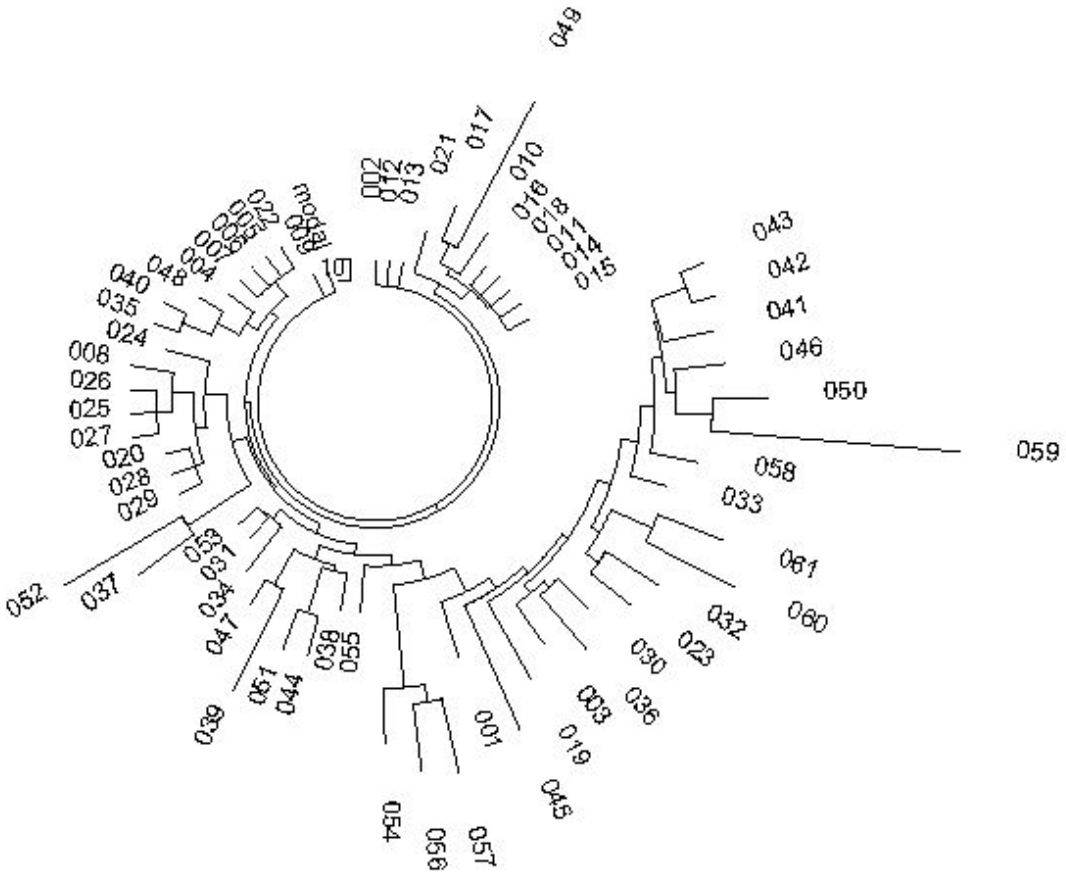

Figure 5. The 25-marker haplotype tree for 61 North West-European haplotypes with DYS388 = 10. The haplotypes were collected from YSearch database.

This DYS388=10 mutation is observed only 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 (Fig. 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±240 years to the common ancestor. It is a rather recent common ancestor, who lived in the middle of the first millennium AD. The closeness of the branch to the trunk of the tree (Fig. 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

All the 30 haplotypes contain 172 mutations, which gives 3575±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 Fig. 5.

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

This common ancestor of both DYS388=12 and DYS388=10 populations lived presumably in the Balkans (see below), 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 was continuing to pass over the generations up to the present time.

Scotland R1a1 haplotypes

R1a1 haplotypes in Scotland have the same ancestral haplotype as those in England and Ireland:

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

29 of 25-marker haplotypes contained 164 mutations, which gives 3550±450 years to the common ancestor.

Germany R1a1 haplotypes

A 67-haplotype series in 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, it equals to 15.48 for England, 15.33 for Ireland, 15.26 for Scotland, and 15.84 for Germany, that is pretty close to each other. All 67 of 25-marker haplotypes contain 488 mutations, that is 0.291±0.013 mutations per

marker on average. It corresponds to 4,700±520 years from a common ancestor in the German territory.

These results are supported by the very recent data on excavation of R1a (SNP SRY10831.2) and presumably R1a1 haplotypes near Eulau, Germany (Haak et al., 2008). The 4,600 old haplotypes follow with dating performed using strontium isotope analysis:

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

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

Norway and Sweden R1a1 haplotypes

Norwegian and Swedish ancestral R1a1 haplotypes are principally the same, and both look exactly as 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±0.023 and 0.242±0.023 mutations per marker, respectively, which give 3,375±490 and 3,825±520 years to their common ancestors. This is likely the same time span within the error margin.

Poland, Czech, and Slovak R1a1 haplotypes

These ancestral haplotypes are very similar to each other, having only one insignificant deviation in DYS439 (shown in bold), which is 10.43 in the Polish base haplotype of 44 haplotypes total.

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 10.63 in Czech and Slovakian combined base haplotype of 27 haplotypes total.

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 markers is 0.2 mutations only, the alleles are just rounded up in opposite directions.

These haplotypes have 310 mutations in the 44 Polish haplotypes, and 175 mutations in 27 Czechoslovak haplotypes, and results in 0.282±0.016 and 0.259±0.020 mutations per marker on average in their 25-marker haplotypes, which results in 4,550±520 and 4,125±430 years from their common ancestors, respectively.

The European R1a1 haplogroup (small haplotype sets across Europe)

Many countries are represented with just a few haplotypes in data bases. I have collected 36 of 25-marker R1a1 haplotypes provided into YSearch database by descendants referring to 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 Fig. 6.

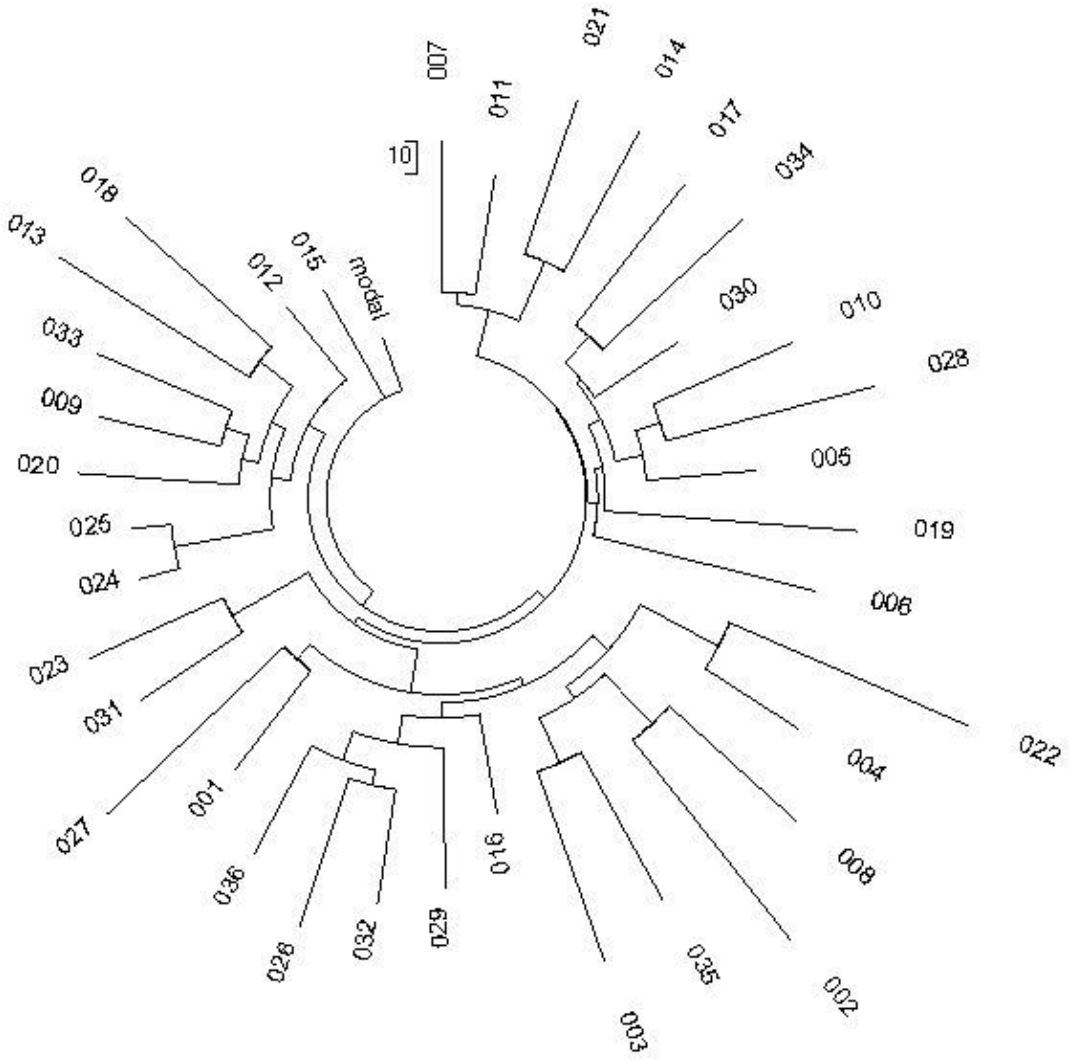

Figure 6. The 25-marker haplotype tree for various European countries (small series of haplotypes from Denmark, Netherlands, Switzerland, Iceland, Belgium, France, Italy, Lithuania, Romania, Albania, Montenegro, Slovenia, Croatia, Spain, Greece, Bulgaria and Moldavia), haplogroup R1a1. The 36-haplotype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

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±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 58 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 Kirgizstan) is shown in Fig. 7.

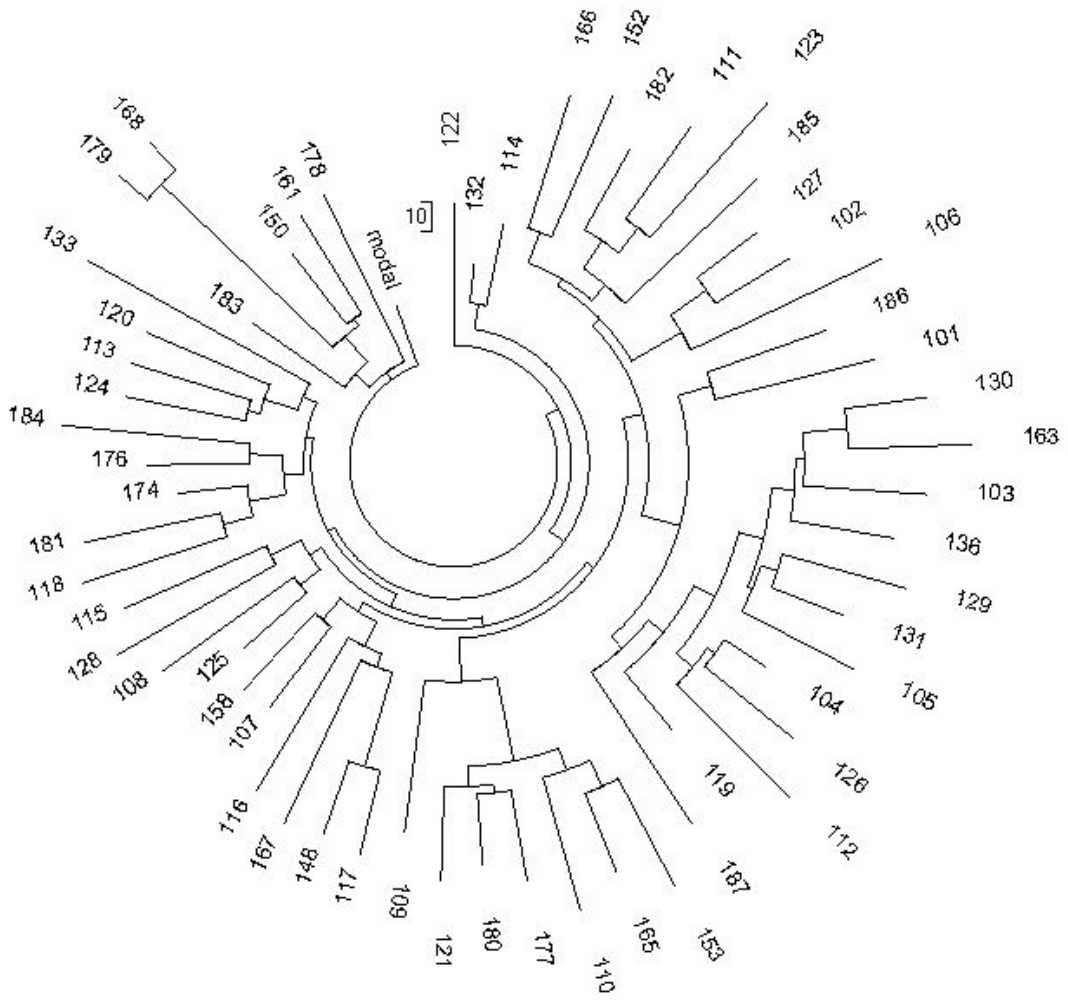

Figure 7. The 25-marker haplotype tree for Russia and Ukraine, haplogroup R1a1.

The 58-haplotype tree was composed from data of YSearch database.

The ancestral (base) haplotype for the haplotype tree is

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

It is exactly the same ancestral haplotype as that in Germany, and it 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 15.84. There are similar insignificant deviations with the Polish and Czechoslovakian base haplotypes, at DYS458 and DYS447, respectively, within 0.2-0.5 mutations. This points at a mutational difference between their common ancestors of only few generations.

All 58 haplotypes contain 423 mutations from the base haplotype, that is 0.292±0.014 mutations per marker or 4,725±520 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 practically random mix in the haplotype tree. 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. 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 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 carrying around through a relay of references to references. The oldest one is apparently 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" (ref. 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 of Wells et al. (2001) states "M17, a descendant of M173, is apparently much younger, with an inferred age of ~ 15,000 years". No 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 is turned out to be valid in a sense that the estimate was inaccurate and elevated by about 300%.

India R1a1 haplotypes

The YSearch database contains 22 of 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.

All 22 Indian R1a1 haplotypes contain 148 mutations, that is 0.269±0.022 mutations per marker. It is close to 0.292±0.014 mutations per marker in the Russian haplotypes and corresponds to 4,300±560 years from a common ancestor of the Indian haplotypes, compared to 4,725±520 years for the ancient "Russian" TSCA.

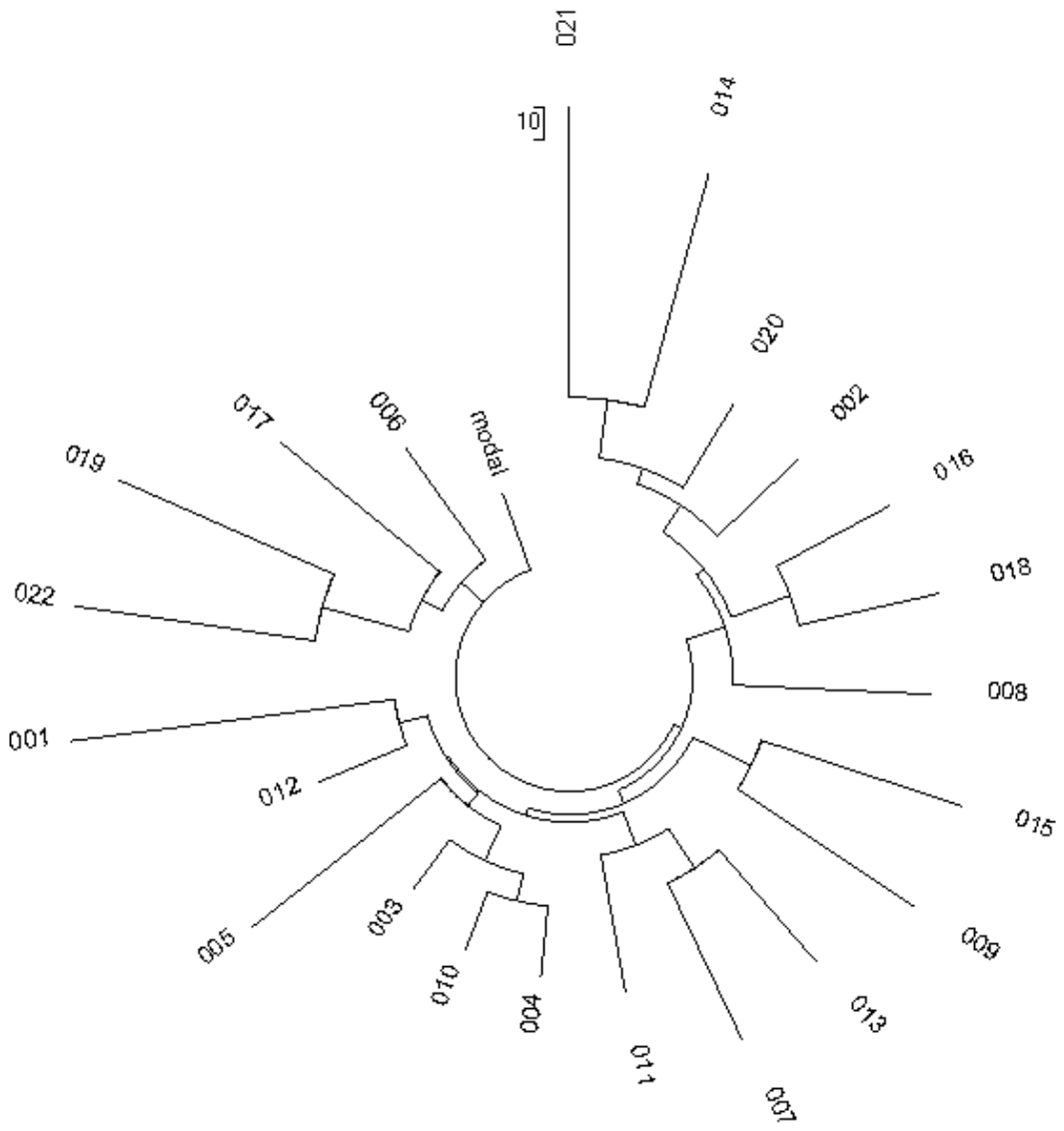

Figure 8. The 25-marker haplotype tree for India, Pakistan and Sri-Lanka, haplogroup R1a1. The 22-haplotype tree was composed from data of YSearch database.

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 showed that eight inhabitants out of nine shared R1a1 haplogroup and haplotypes (Bouakaze et al., 2007) as follows, dating between 5,500 and 1,800 years bp:

13-25-16-11-11-14-X-Y-Z-14-11-32

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

This provides rather strong evidence that the R1a1 tribe migrated from Europe to the East between 5,000 and 3,600 years bp. The pattern of this migration is exhibited as follows: 1) the descendants who live today share a common ancestor of 4,725±520 years prior, 2) the Andronovo archaeological complex of cultures in North Kazakhstan and South and Western Siberia dates 4,300 to 3,500 years bp, and it revealed several R1a1 excavated haplogroups (see above), 3) they reach to South Ural some 4,000 years bp, is where they built Arkaim, Sintashta (contemporary names) and the so-called "a country of towns" on South Ural 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 4,300±560 years bp chronologically corresponds to the events. Currently, some 16% of

Indian population, that is about 100 millions males, and the majority of the upper castes bear R1a1 haplogroup (Sengupta et al, 2006; Sharma et al, 2009).

R1a1 haplotypes, 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 differ by only one mutation, marked in bold. The first haplotype is the base one for seven haplotypes with 13 mutations in them, on average 0.186±0.052 mutations per marker, which gives 2,300±680 years to a common ancestor. The second haplotype is the base one for nine haplotypes with 26 mutations, an average 0.289±0.057 mutations per marker or 3,750±825 years to a common ancestor. Since a common ancestor of R1a1 haplotypes in Armenia and Anatolia lived 4,500±1,040 and 3,700±550 years bp, respectively (Klyosov, 2008b), it does not conflict with 3,750±825 years bp in the Arabian peninsula.

The Balkan ancient branch: the oldest trace of R1a1 haplogroup?

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 Fig. 9

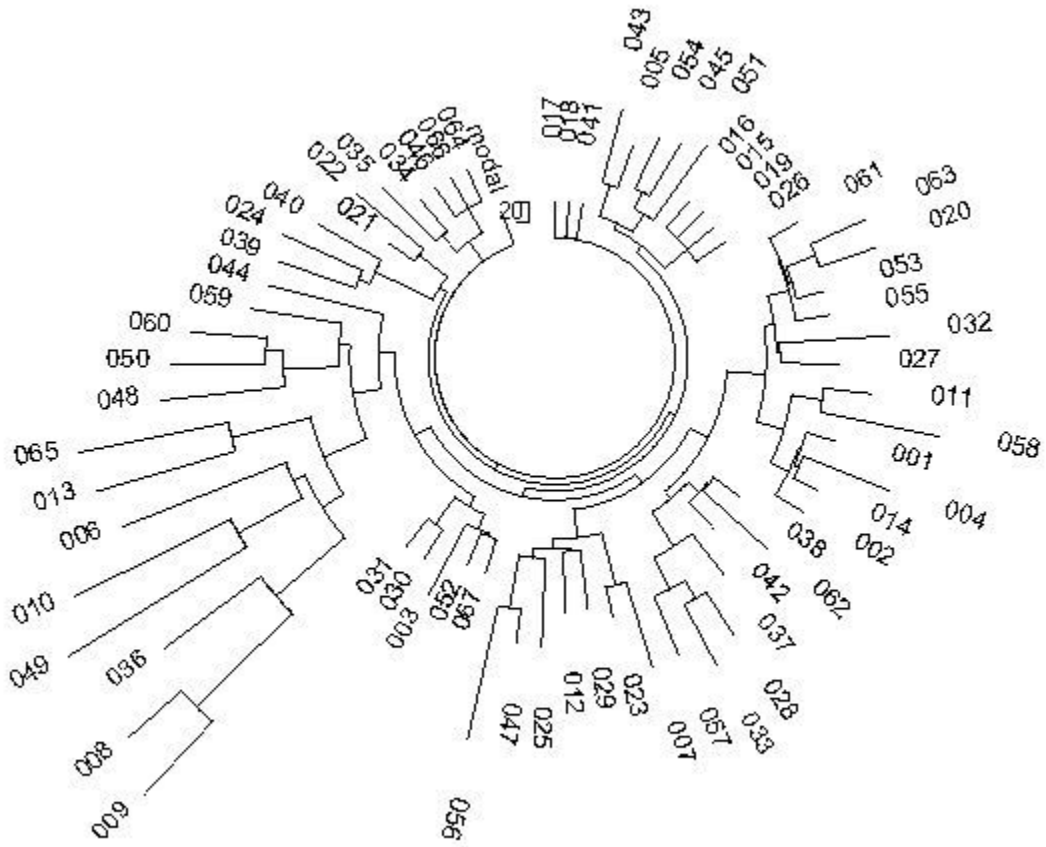

Figure 9. The 9-marker haplotype tree for the Balkans, haplogroup R1a1. The 67-haplotype tree was composed from data published (Barac et al., 2003a, 2003b; Pericic et al., 2005).

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 Fig. 9.includes a rather small branch of twelve haplotypes on top of the tree, which contains only 14 mutations. This results in 0.130±0.035 mutations per marker, or 1,850±530 years to a common ancestor. Its base haplotype

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

is exactly the same as that in Russia and Germany.

The wide 27-haplotype branch on the right contains 0.280±0.034 mutations per marker, which is rather typical for R1a1 haplotypes in Europe. It gives typical in kind 4,350±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 44 Russian-Ukrainian haplotypes (Fig. 7), 23 haplotype have “10”, and 21 have “11” in that locus. 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 haplotypes than to German ones.

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 determine 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±0.071 mutations on average per marker, and results in 11,425±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±0.081, which corresponds to 0.880/0.00189 = 466±62 generations or 11,650±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 (the preceding paper, Part 1).

A calculation of 11,650±1,550 years to a common ancestor is practically the same as 11,425±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^2 (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±62 generations or 11,650±1,550 years to a common ancestor of the R1a1 group in the Balkans.

The obtained data suggest that the first bearers of R1a1 haplogroup lived in the Balkans (Serbia, Kosovo, Bosnia, Macedonia) between 10 and 13 thousand years bp. It is unknown whether R1a1 appeared in the Balkans, presumably from R1 or R1a, or arrived from a yet unknown location. It was found (Klyosov, 2008a) that haplogroup R1b appeared about 16,000 years bp, apparently in Asia.

The data shown above suggests that only about 6,000-5,000 years bp bearers of R1a1 began to mobilize and migrate to the west toward the Atlantics, 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 at their common ancestors who lived around 4,800 to 4,500 years bp. On their way through the Russian plains and steppes the R1a1 tribe presumably sat up the Kurgan archaeological culture, apparently domesticated the horse, advanced to Central Asia and left the "Aryan population" which dated about 4,500 years bp. They then moved to the Ural mountains about 4,000 years bp and migrated to India as the Aryans circa

3,600-3,500 years bp. Presently, 16% of the male Indian population, or approximately 100 million people, bear R1a1 haplogroup's SNP mutation, with their common ancestor of 4,300±560 years bp, of times back to the Andronovo archaeological culture and the Aryans in the Russian plains and steppes . The current 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 points at some trails of R1a1 in India, an "Aryan trail".

The Chenchu R1a1 haplotypes

The Chenchus, an australoid tribal group from southern India, bear R1a1 haplogroup in 11 of 41 individuals tested (or 27% of total) (Kivisild et al., 2003). It is tempting to associate this with the Aryan influx into India which allegedly occurred some 3,600-3,500 years bp. However, questionable calculations of time spans to a common ancestor of R1a1 in India (Kivisild et al., 2003; Sengupta et al., 2006; Sahoo et al, 2006) using methods of population genetics rather than those of DNA genealogy have precluded an objective and balanced discussion of the events and their consequences.

Eleven R1a1 haplotypes of the Chenchus (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 years bp. Conversely, if these haplotypes are all mutated, and there is no base (ancestral) haplotype among them, a common ancestor

lived thousands years bp. 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 certain margin of error. If eleven of the 6-marker haplotypes are all mutated, it would mean that a common ancestor lived apparently earlier than 6 thousand years bp. Hence, even with such a poor set of haplotypes one can obtain useful and meaningful information.

Eleven Chenchu haplotypes have as many as seven identical (base) 6-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 6-marker format:

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

Actually, the author of this study, himself a Slav (haplotype R1a1) has the “Chenchu” base 6-marker haplotype.

These identical haplotypes are represented by a “comb” in Fig. 10. 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±1,900 years old and the other only 350±350 years old, starting from around the 17th century AD. The tree in Fig. 10 shows these two lineages.

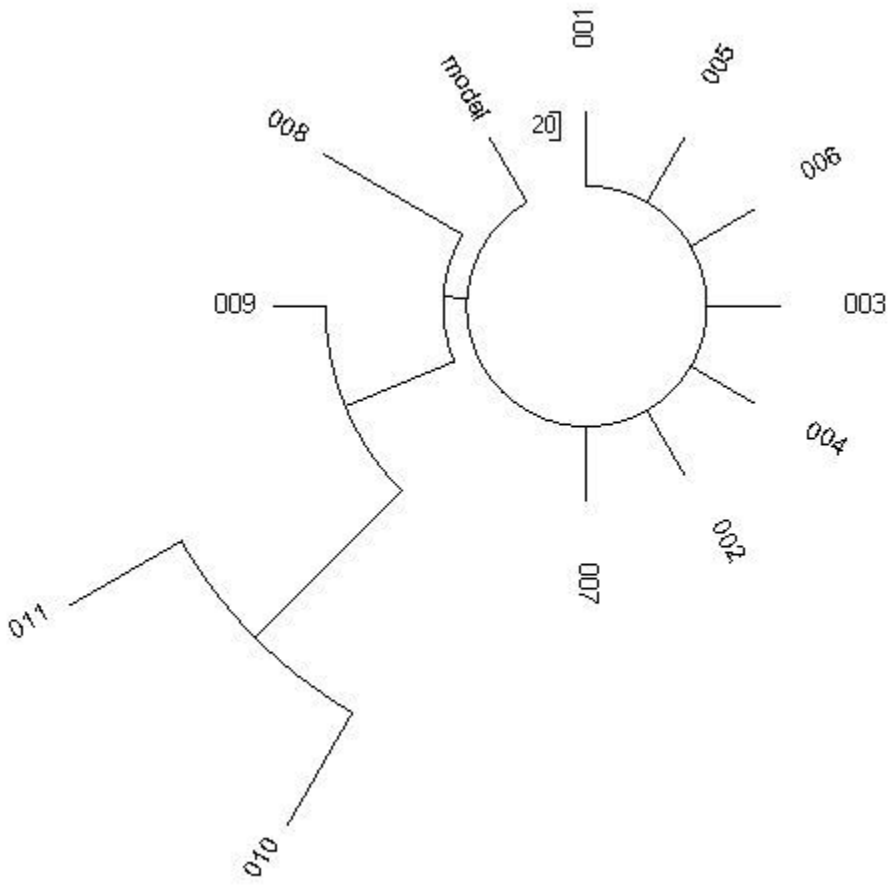

Figure 10. The 6-marker haplotype tree for the South Indian tribe Chenchu, haplogroup R1a1. The 11-haplotype tree was composed from data of Kivisild et al. (2003).

A quantitative description of these two lineages is as follows. Despite the 11-haplotype series contain 7 identical haplotypes, which – in case of one common ancestor

for the series – would have point at 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 figures without a correction for back mutations). Because there is a significant mismatch between these 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 (Fig. 10). A recent branch includes 8 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 8 individuals of the eleven lived only about 350±350 years bp, around the 17^{th} century. The old branch of haplotypes points at a common ancestor who lived 3/3/0.0088 = 114±67 generations BP, or 3,200±1,900 years bp with a correction for back mutations.

Considering that the Aryan (R1a1) wave to northern India took place about 3,600-3,500 years bp, it is quite plausible to refer the appearance of R1a1 in the Chenchu by 3,200±1,900 ybp to the Aryans.

The origins of the influx of Chenchu R1a1 haplotypes around the 17th century are found in this passage excerpted from (Kivisild et al., 2003): “Chenchus were first described as shy hunter-gatherers by the Mohammedan army in 1694”.

Native Americans haplotypes of haplogroup Q1a3a

Let us consider much more distant time periods to further examine and justify the timing methods of DNA genealogy developed in this study. One hundred seventeen haplotypes of Native American Indians, haplogroup Q-M3 (Q1a3a), were published (Bortolini et al., 2003), and a haplotype tree, shown in Fig. 11, was developed based upon their data.

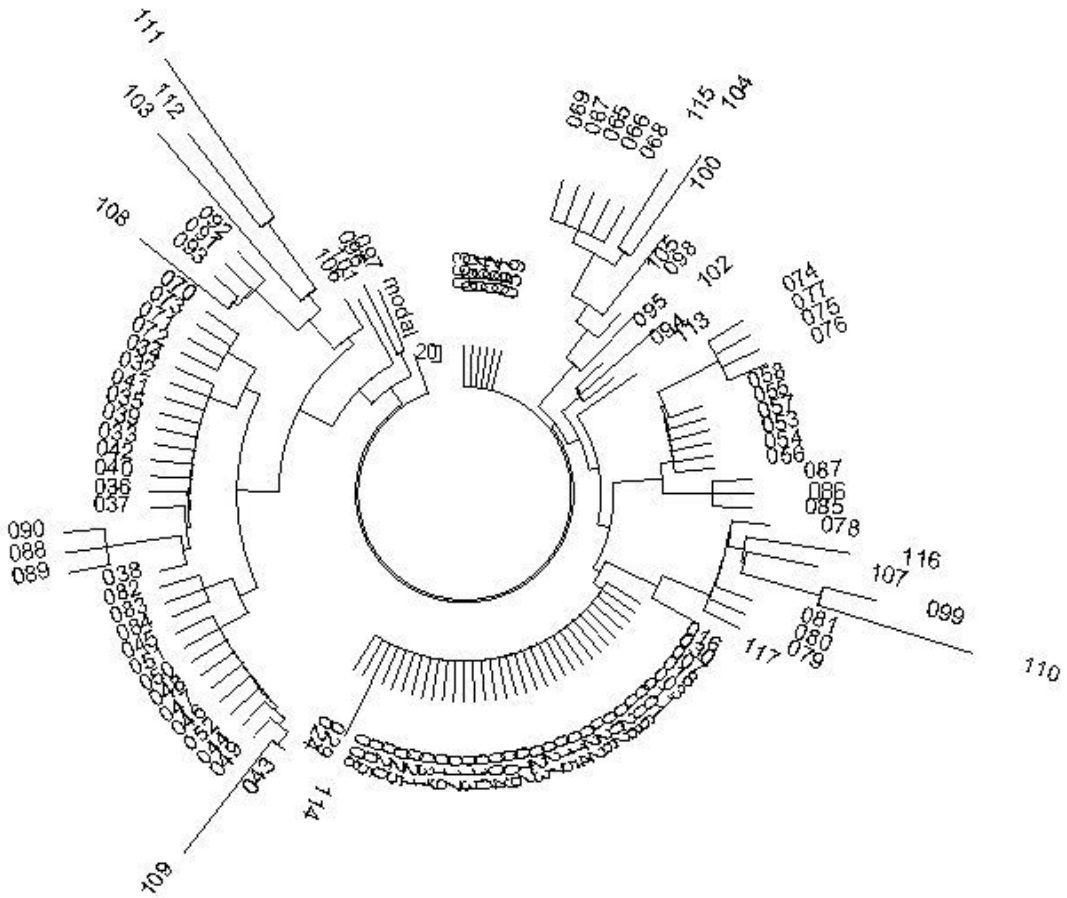

Figure 11. The 6-marker haplotype tree for the Native Americans, haplogroup Q-M3 (Q1a3a). The 117-haplotype tree was composed from data of Bortolini et al. (2003)

The tree contains 31 identical (base) haplotypes and 273 mutations from that "base" haplotype. It is obvious that the haplotypes in the tree descended from different common ancestors, since 31 base haplotypes out of 117 total would give ln(117/31)/0.0088 =151 generations to a common ancestor, though 273 mutations in all 117 haplotypes would give 265 generations (both 151 and 265 without corrections for back mutations). This is our principal criterion, suggested in this study, which points at multiplicity (more than one) of common ancestors in a given haplotype series. This in turn makes any calculations of a time span to a "common ancestor" a mute point, since a "common ancestor" in those cases is typically a phantom. Depending on relative amounts of descendants from different common ancestors in the same haplotype series, a timespan to a "phantom common ancestor" varies greatly, often by many thousands of years.

An analysis of the haplotype tree in Fig. 11 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±0.086 mutations per marker on average for 6-marker haplotypes, that is 12,125±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 a 6-marker haplotype. This corresponds to about 19,700 years of mutational difference between them and points out that THEIR common ancestor lived 16,300 ± 3,300 years bp.

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 years bp with the 95% confidence. This dating is in line with many independent data of archaeological, climatology, and genome study 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 years bp. (Fagundes et al, 2008). Others refer to archaeology data of Paleo-Indian people between 11,000 to approximately 18-22,000 years bp (Haynes, 2002, p. 52; Lepper, 1999, pp. 362–394; Bradley &

Stanford, 2004; Seielstad et al., 2003). At any rate, the time span of 16,000 years ago corresponds well 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 6-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 it splits into two principal lineages (Klyosov, 2008c), with 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 “Cohen Modal Haplotype”, if to follow its 6-marker notation, has the following 12-marker base haplotype (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 information from the President of the Cohanim Association of Latin America, Mr. Mashuah HaCohen-Pereira; also 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 actual 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±560 ybp. The presumably Cohanim-Sephardim base haplotype belonged to the common ancestor who lived 2,500±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.

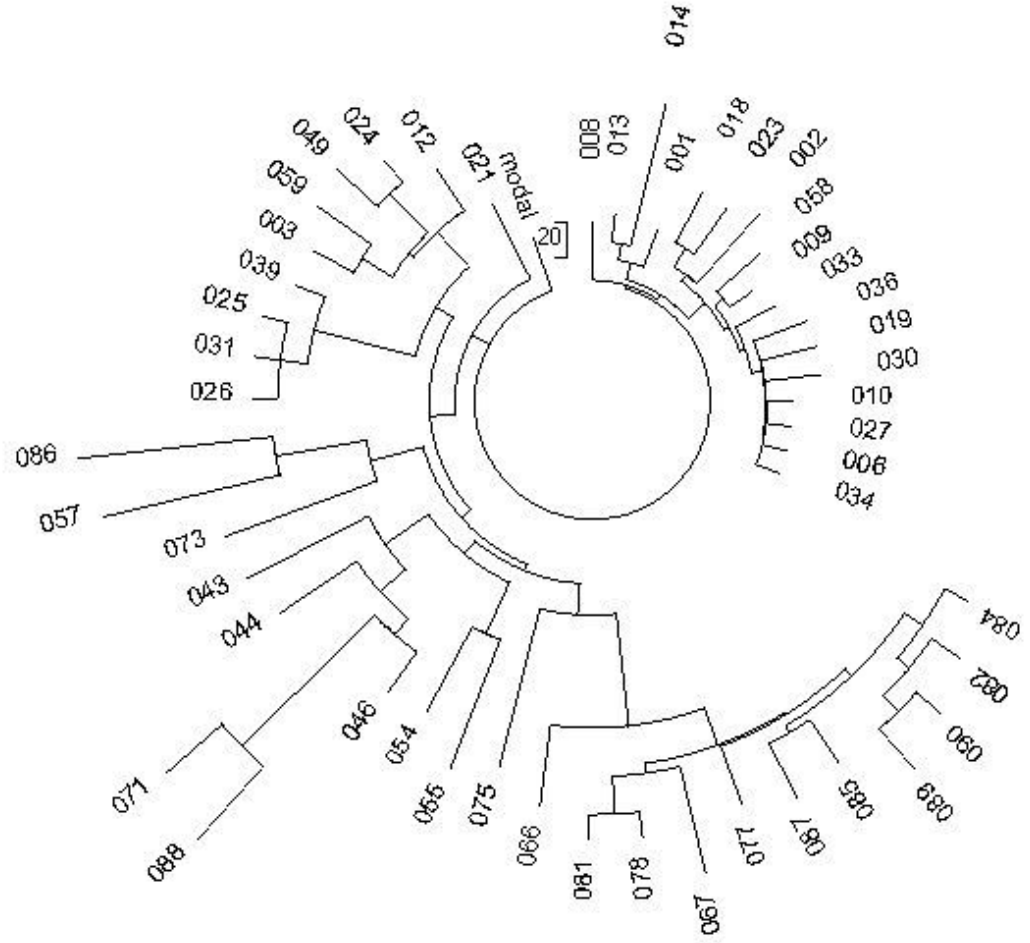

Figure 12. The 25-marker haplogroup J1 haplotype tree for 49 presumably Jewish haplotypes. The haplotypes were collected in YSearch database (Klyosov, 2008c).

A 25-marker haplotype three of 49 presumably Jewish J1 haplotypes is shown in Fig. 12. In this tree the both CMH branches are located on both sides at the top of the tree, the “recent CMH” (rCMH), more compact 17-haplotype branch on the right

(between haplotypes 008 and 034), and the 9-haplotype "older CMH" (oCMH) branch on the left (between haplotypes 012 and 028). 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 if to consider them in a round-up format, as shown above in bold. In fact, 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 figure later.

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

From the data obtained we can calculate that THEIR common ancestor lived about 4,225±520 years bp. That is when a common ancestor of the "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 detail the above information, as many as 85 of 37-marker “Cohen Modal” haplotypes were collected from both Jewish and non-Jewish descendants, and 33 haplotypes among them contained as many as 67 markers. Fig. 13 shows a 37-marker haplotype tree, in which all haplotypes belong to the CMH series, that is all of them have the 14-16-23-10-11-12 string in their haplotypes for DYS19-388-390-391-392-393. The left-hand side represents the “recent CMH”, the branch at the lower right represents the “older CMH” haplotypes.

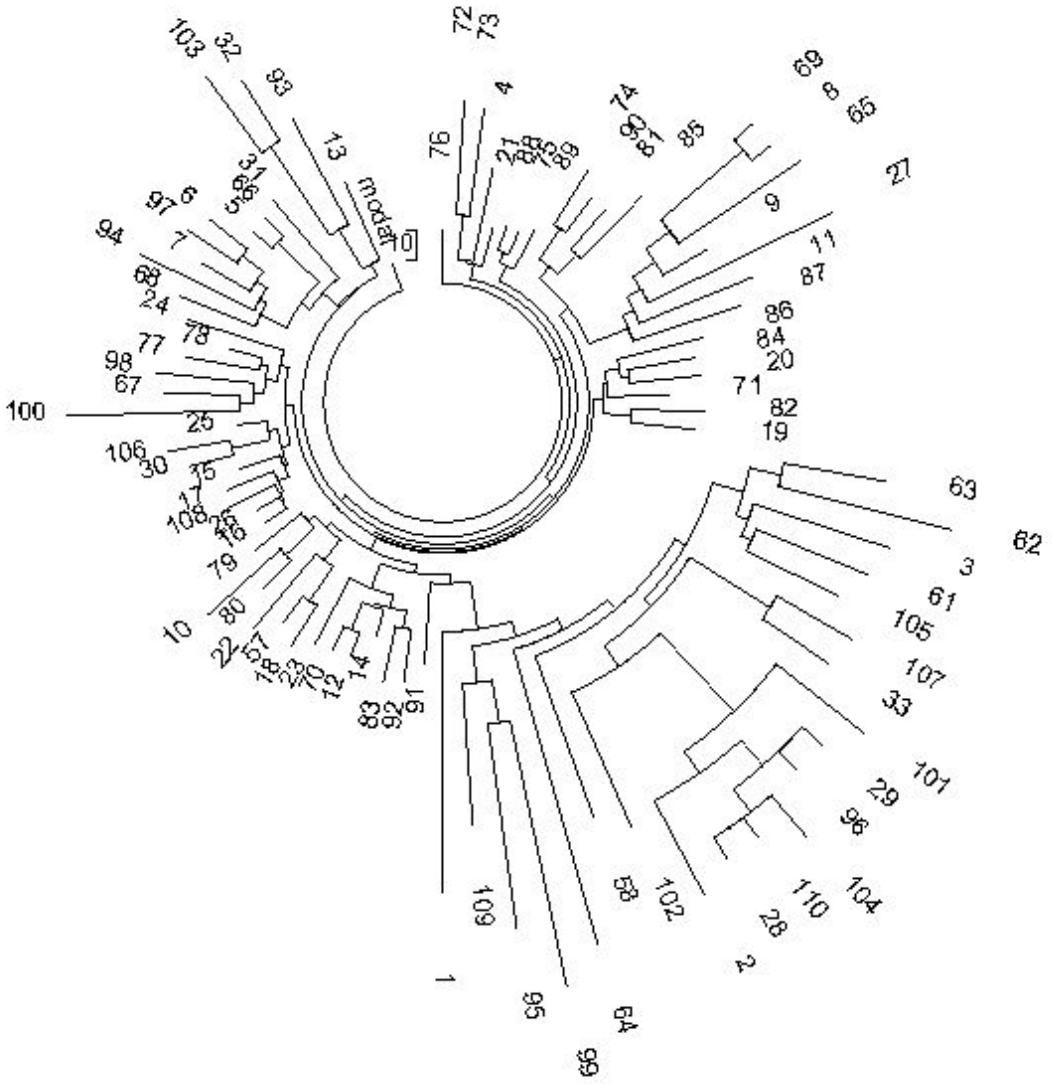

Figure 13. The 37-marker haplotype tree for the "Cohen Modal Haplotypes", haplogroup J1. The 85 haplotype tree was composed of haplotypes collected in YSearch database (Klyosov, 2008c) and private "Cohen Haplotype" projects, and provided by Dr. Alberto Aburto.

The “older CMH” 22-haplotype branch contains 126 mutations in 25-marker format and 243 mutations in 37-marker format, which results in 3,575±480 and 3,525±420 years from a common ancestor, respectively, on average 3,525±450 years.

The “recent CMH” branch corresponds to 975±135 and 1175±140 years to a common ancestor, respectively, on average 1,075±190 years bp.

Maximum high resolution CMH haplotypes are shown on the 67-haplotype tree (Fig 14).

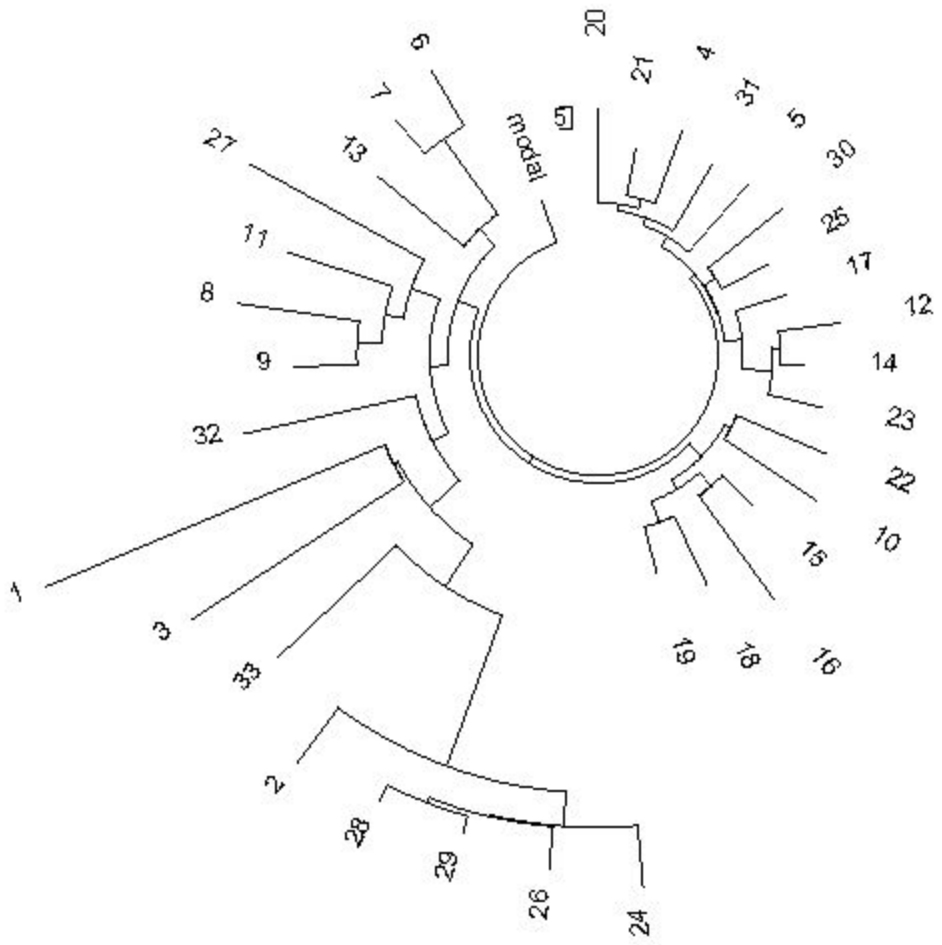

Figure 14. The 67-marker haplotype tree for the “Cohen Modal Haplotypes”, haplogroup J1. The 33 haplotype tree was composed of haplotypes collected from three

sources: 1) YSearch database (Klyosov, 2008c), 2) private "Cohen Haplotype" projects, and 3) provided by Dr. Alberto Aburto.

The tree 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 two principal "Cohen Modal Haplotype" branches, each one with its own common ancestor, who lived about 3,000 years apart. A common ancestor of the "older CMH", calculated from 25-,and 37-marker haplotypes, lived 4,150±580 and 3,850±470 years bp, on average 4,000±520 years bp. A common ancestor of the "recent CMH", also calculated from 25- and 37-marker haplotypes, lived 975±205 and 1,150±180 years to a common ancestor, respectively, on average 1,050±190 years bp, around the 9th to the11th century. This coincides with the Khazarian times; however, it would be a stretch to claim so.

The base (ancestral) 67-marker haplotype of the "older CMH" 16-haplotype branch (Fig. 14) 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:

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 bearers of the "older CMH" (10 individuals of the 16) in the respective 67-marker branch denied 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 contrary, 16 out of 17 haplotypes in the recent CMH branch claimed their Jewish heritage, and several claimed themselves to be descendants of Cohens. Their base haplotype is identical with the Jewish rCMH shown earlier.

To verify this concept, three haplotypes of inhabitants of the Arabian Peninsula with typical Arabic names, 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 a set of haplotypes shown in Fig. 13. All three Arabic haplotypes joined the lower, predominantly non-Jewish branch on the right (Fig. 13). After the addition of the Arabic haplotypes, all 25 of 25-marker haplotypes in the branch contained 162 mutations, which gives 4,125±525 years to a common ancestor for a collection of 88 Jewish and non-Jewish haplotypes, all of them belonging to the "CMH" family of haplotypes. This time period is close to that of the legendary Biblical split into the Jewish and the Arabic lineages, whether it is applicable or not to the results of this study.

The following section demonstrates that the "CMH" in fact appeared as long as 9,000 years bp or earlier on the Arabian Peninsula. The above time spans of about 4,000±520 or 4,125±525 (the "older CMH"), and 1,050±190 (the "recent CMH") years bp was generated as a result of drifts 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 9th and 11th centuries. At any rate, the bearer of the base "recent CMH" became a common ancestor to perhaps millions of present-day bearers of this lineage.

The Arabic "CMH", 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 Arabs from the Arabic Peninsula is shown in Fig. 15. The tree was composed from 19 of 37-marker haplotypes of haplogroup J1 listed in the "Arabian Peninsula YDNA Project" (2008).

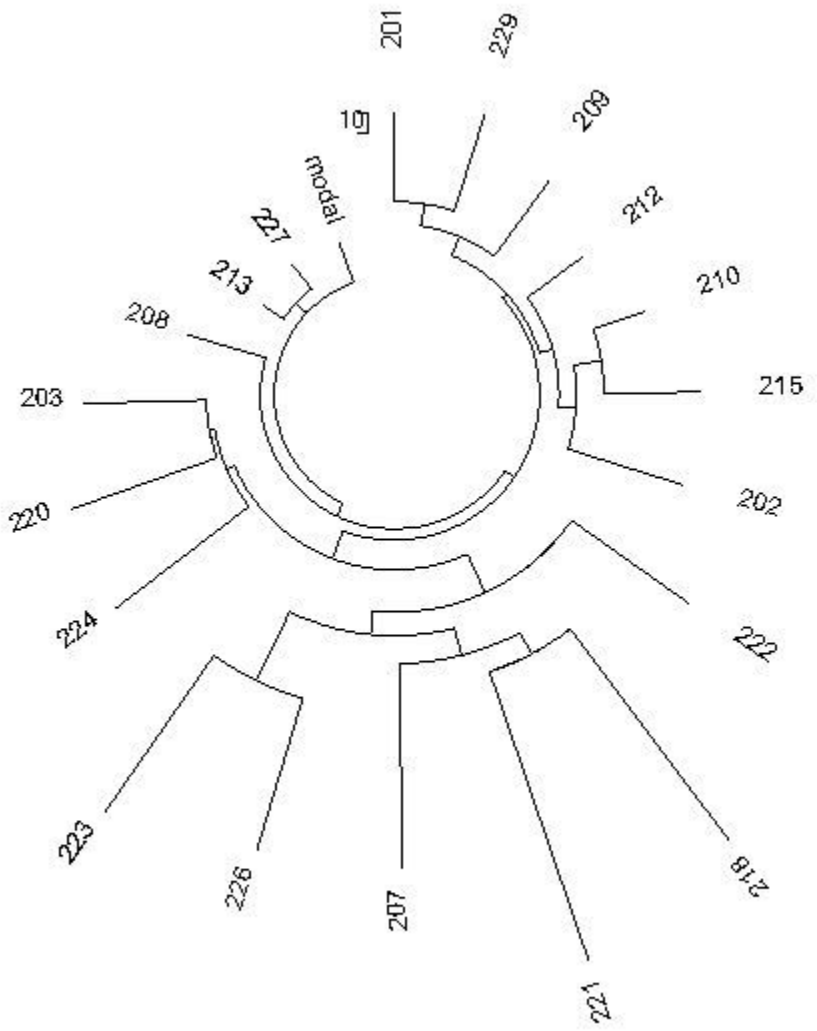

Figure 15. The 37-marker haplotype tree for Arabic haplotypes of haplogroup J1.

19 haplotypes were listed in the Arabic Peninsula YDNA Project (2008).

There are two branches of the tree, one having the "Cohen Modal Haplotype", (that is containing the string of alleles 14-23-16-10-11-12 in the loci DYS19-388-390-391-392-393) as an 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

Seventy three mutations in the six 25-marker haplotype branch, or 0.487±0.057 mutations per marker on average, gives 9,000±1,400 years to a common ancestor.

The second, a 7-haplotype branch (Fig. 15), has a significantly "younger" common ancestor, since it is much less extended from the "trunk" of the tree. It has the following base haplotype

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

All 7 haplotypes contain 27 mutations in their first 25 markers or 0.154±0.030 mutations per marker and 2,300±500 years to a common ancestor for this branch.

These two ancestral haplotypes differ by 10 mutations in their 25-marker format, which corresponds approximately to 7,000 years between them. The "younger" common ancestor who lived 2,300±500 years bp is very likely is a direct descendant of the "older" one who lived 9,000±1,400 years bp and had the "CMH" haplotype.

It seems that the "Cohen Modal Haplotype" was likely an ancestral haplotype between 7600 and 10400 years bp (with the 95% confidence) for both historical inhabitants of the Arabian Peninsula, along with the Arabs and the Jews of the Middle East. About 4,000±520 years ago the establishing Jewish population carried this "modal haplotype" although in its slightly drifted structure, which was coined as the "older

CMH". By around the 7th century AD, the "recent CMH" split from the "older CMH" and became the ancestral haplotype for a separate albeint recent Jewish lineage within haplogroup J1. If to consider only "CMH" haplotypes within this population, a common ancestor can be identified who lived 1,050±190 years bp can be identified.

CMH, Haplogroup J2

The Jewish "Cohen Modal Haplotype" of haplogroup J2 represents a rather compact group of haplotypes with a recent ancestor who lived in about 7th century AD (see below). As it was indicated above, this "J2-CMH" is unlikely to be associated with actual Cohanim, and represents just a string of alleles accidentally including the 14-16-23-10-11-12 sequence.

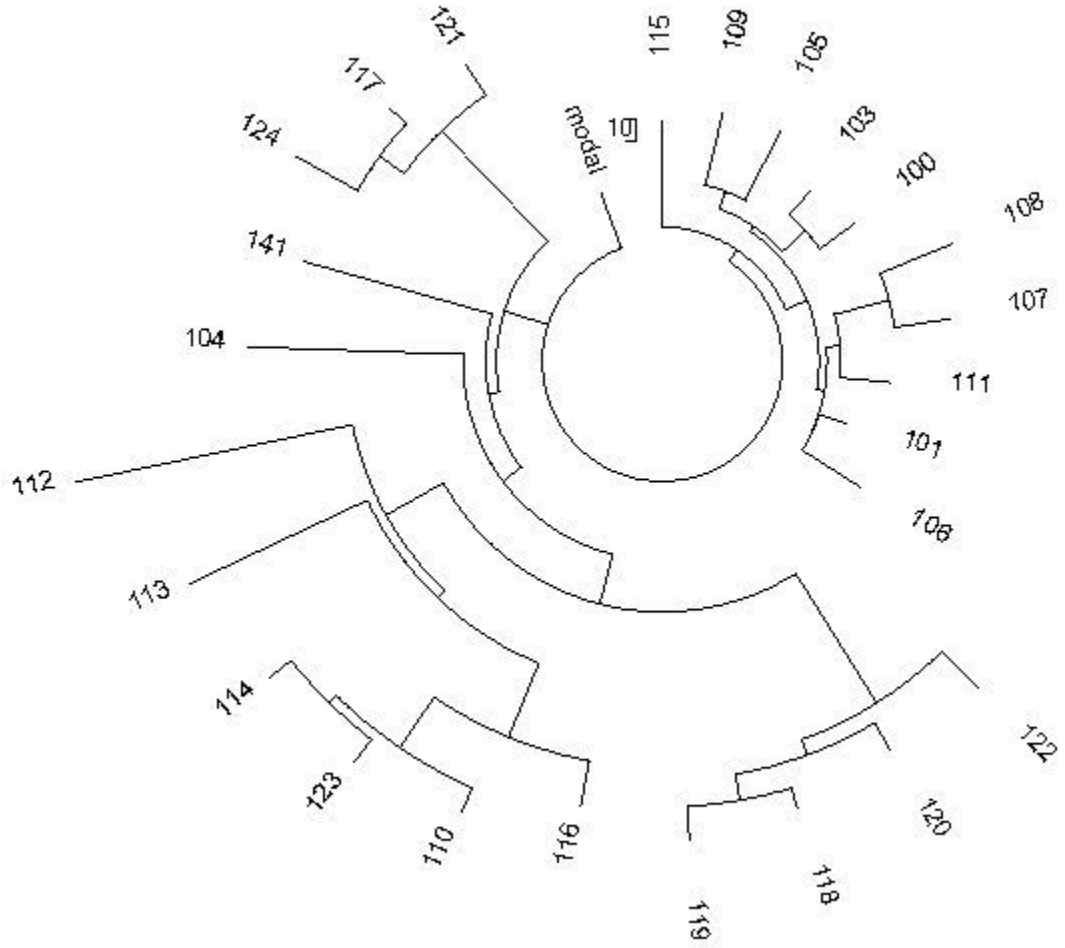

Figure 16. The 37-marker haplotype tree for J2 Jewish haplotypes. Twenty five haplotypes were collected in YSearch database (Klyosov, 2008c).

The compact group of ten haplotypes, located rather close to the trunk of the tree (Fig. 16), which indicates their "younger age", represents the J2-CMH. Their 37-marker base haplotype is as follows:

12-23-14-10-13-17-11-16-11-13-11-30-18-9-9-11-11-26-15-20-29-12-14-15-16-10-11-19-22-15-13-19-17-35-39-12-9

All 10 haplotypes contain 25 and 44 mutations in the 25- and 37-marker format, respectively. This gives 1,450±320 and 1,300±230 years to a common ancestor or 1,375±300 years bp on average, who lived around the 6th and 8th centuries AD.

The J2-CMH base haplotype differs from 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 Gypsies, haplogroup H1

-- Bulgarian Gypsies

According to old records, gypsies arrived in Bulgaria during the Middle Ages. Haplotypes of Bulgarian Gypsies, or Roma, have been determined by testing 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 Gypsies, perhaps a single tribe, came to Bulgaria between 500 and 700 years ago. Descendants of other unrelated tribes apparently did not survive. It cannot be excluded from consideration that a few close relatives were the patriarchs of the tribe that survived.

The most numerous tribe “Rudari” had the following 6-marker base haplotype which was represented by 62 identical haplotypes of the total amount of 67 haplotypes in 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 years bp.

Overall, all 179 haplotypes of Bulgarian Gypsies contained 146 identical (base) 6-marker haplotypes and 34 mutations compared to the base haplotype.

Considering base haplotypes, this gives ln(179/146)/0.0088 = 23 generations or 575 years to the common ancestor for all 179 members of all the 12 Gypsy tribes. Considering mutations, this gives 34/179/0.0088 = 22 generations or 550 years.

This quite a typical fit between the TSCAs obtained from two quite different modes of calculations points time and again that there is a fundamental basis for such fit. This basis was explained in detail in the preceding paper (Part I). Indeed, very often results are practically the same whether calculated using the logarithmic method or the “mutation-counting” method. “Formal” calculations of the standard deviation will give a 20%-SD (19.85%, to be more accurate) for the second figure above, that is 550±110 years to the common ancestor, and about the same margin of error for the first figure.

Thus, the Gypsies show a pretty straightforward, uncomplicated DNA genealogy family tree. Incidentally, these haplotypes belong to haplogroup H1, which is very typical for India. Beyond India it is met mainly among the Gypsies. The base 6-marker haplotype for H1 in India is exactly the same as that for the Gypsies shown above. However, its common ancestor in India lived several thousand years ago.

-- Croatian Gypsies

Another example of a population which apparently represents the Gypsies 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 Fig. 17.

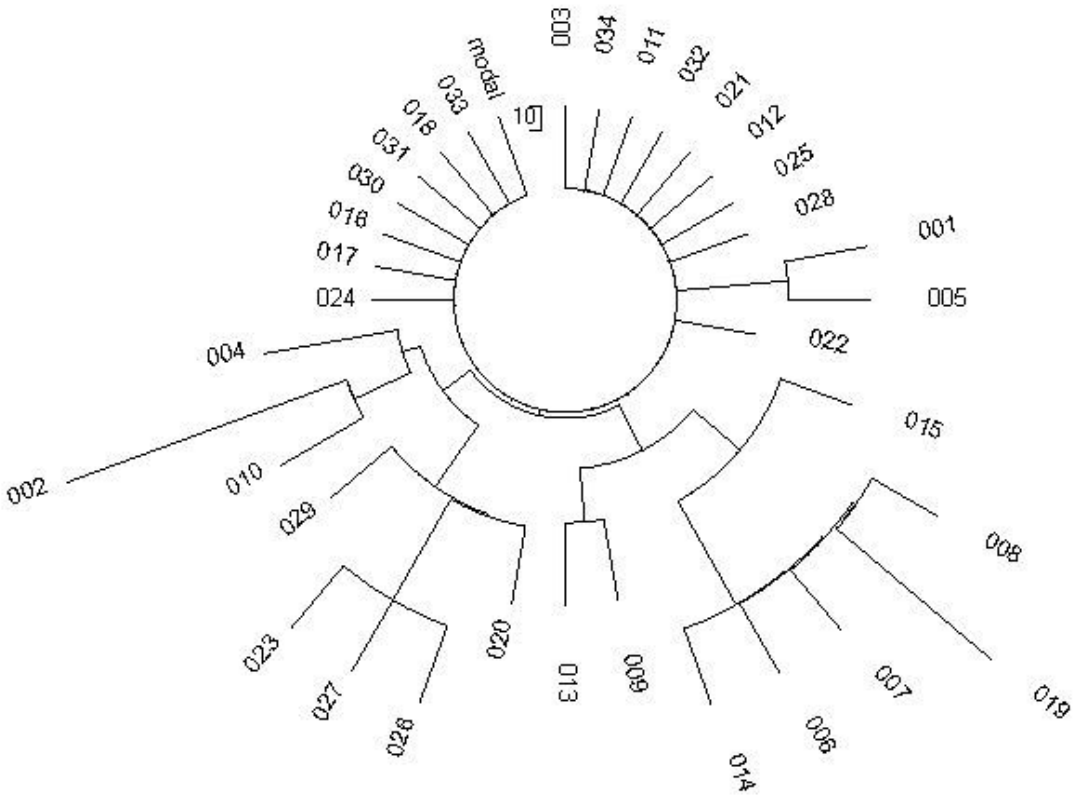

Figure 17. The 9-marker haplotype tree for 34 Croatian (apparently, Gypsy) haplotypes of haplogroup H1. The haplotype tree was composed from data (Barac et al., 2003a, 2003b; Pericic et al., 2005).

Said 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 to be the Gypsies. 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±10 generations ago (with a correction for back mutations), or 1125±250 years ago, between the 7^{th} and 12^{th} century AD. Obviously, the Gypsies arrived in Europe earlier than can be called “the middle centuries” as was previously suggested (Zhivotovsky et al, 2004).

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), 36 six-marker haplotypes were determined in these three populations, and the base haplotype for all of them follows:

16-15-20-10-12-14

There were 28 base haplotypes among 36 haplotypes total and only 10 mutations in the whole set with respect to the base haplotype.

Considering base haplotypes, this gives ln(36/28)/0.0088 = 29 generations (without a correction), or 30 generations (with a correction) to a common ancestor for all 36 Polynesians.

Considering mutations, this gives 10/36/0.0088 = 32 generations (without a correction), or 33 generations (with a correction) to a common ancestor.

It is not a poor fit, taking into account the relatively small set of short haplotypes, and gives 800±260 years to a common ancestor.

Incidentally, the theorized Polynesian expansion refers to between 650 and 1,200 years ago (cit. in Zhivotovski et al, 2004). One can see that the 800±260 years bp fits well into this range.

The South African Lemba haplotypes

Lemba is a South African tribe whose people live in Limpopo Province, Zimbabwe, Malawi, and Mozambique. A list of 136 Lemba haplotypes was published (Thomas et al.,

2000), and the authors alluded that some Lemba belong to the CMH Jewish lineage. We will demonstrate that it is very unlikely.

Forty one of the tested Lemba individuals had typical "Bantu" haplotypes belonging to the E3a haplogroup (by the author's definition) with a base haplotype:

15-12-21-10-11-13

All 41 haplotypes contains 91 mutations from the above haplotype, that is 0.370±0.039 mutations per marker or 8,300±1,200 years from a common ancestor.

Another 23 Lemba who were tested had the following base haplotype (a haplogroup was not identified, as well as of any haplogroups for the published haplotypes in Thomas et al., 2000):

14-12-23-10-15-14

All 23 haplotypes had only 16 mutations, which gives 2,150±580 years to a common ancestor for these individuals. However, their "modal" haplotype was quite different from the Cohen Modal Haplotype, with 10 mutations on the 6-marker haplotype. It corresponds to a mutational difference with the CMH equivalent to about 50 thousand years from a common ancestor.

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 (Fig. 18). Three base haplotypes, one for each branch, are shown below:

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 Fig. 18), 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.

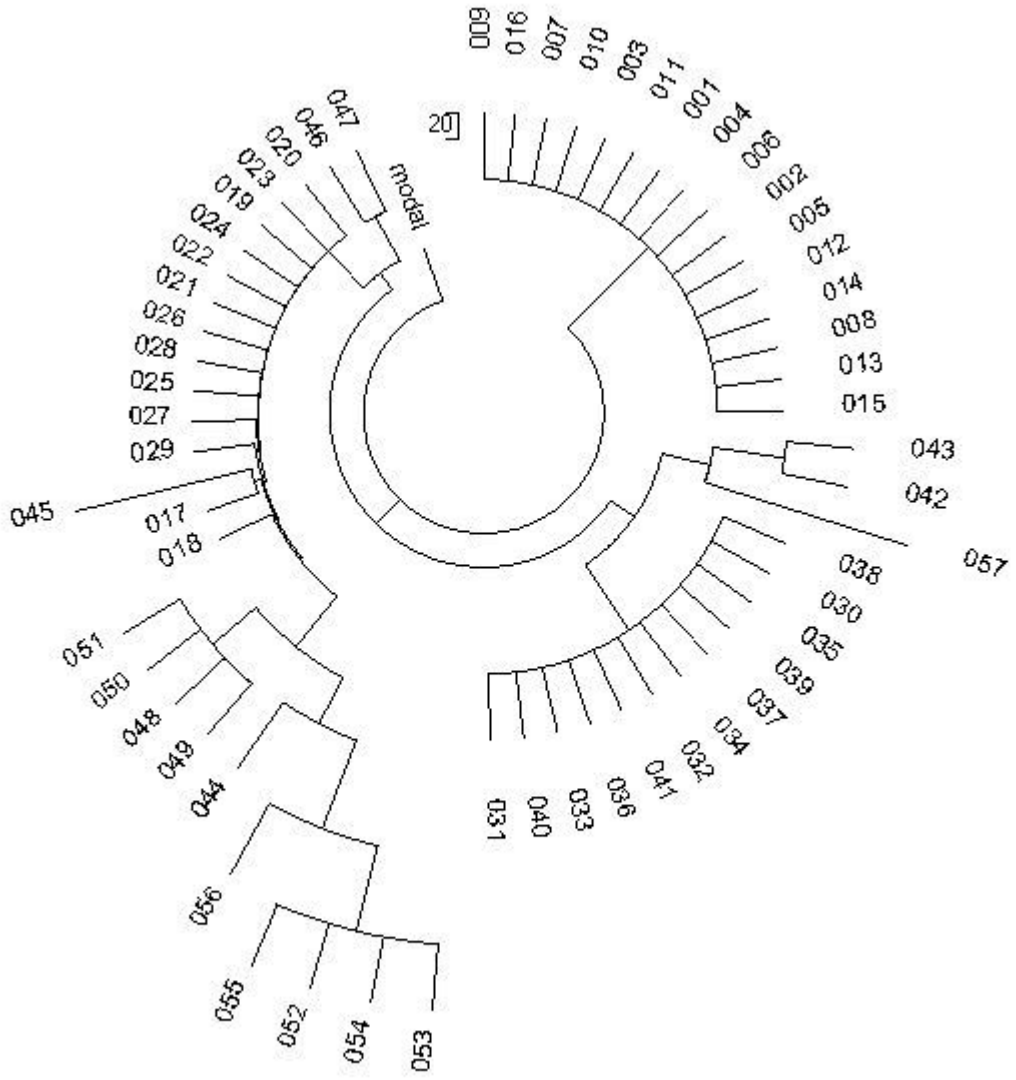

Figure 18. The Lemba 6-marker haplotype tree, apparently of haplogroup J. The 57 haplotype tree was composed of data published in (Thomas et al., 2000).

The second one, being a base haplotype for a 26-haplotype branch on the left-hand side in Fig. 18, is a rather common haplotype in the Arabic world, and belongs likely to haplogroups J and/or J2. The branch contains 21 mutations, which gives 2,550±610 years to a common ancestor, who most likely lived in the first millennia BC. It is clearly not the

"Cohen Modal Haplotype" and differs from the last by two mutations, which in the 6-marker format corresponds to about 7300 years..

The third base haplotype, which is the CMH in its 6-marker format, supports a branch of 15 haplotypes on the lower right-hand side. 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±200 years bp, around the 14^{th} 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 to which haplogroup they belong, J1 or J2.

Obviously, to call the Lemba haplotypes the "Cohen haplotype" is a huge stretch. They could have been Jewish and originated just a few centuries ago, or they could have been Arabic. 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 acquirement.

A conclusion

The two papers (Part I and II) present a rather consistent concept of dating both recent and distant common ancestors based upon appearance of haplotype trees, provide an approach to a verification of haplotype sets in terms of a singular common ancestor for the set or a multiplicity of them applying 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 testing 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, all of it is written in the DNA of each of us, the survivors. It is 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 advances quantitative descriptions in the field of DNA genealogy.

Acknowledgements I am indebted to Dr. A. Aburto for providing a series of the "Cohen Modal Haplotypes", to Theresa M. Wubben for valuable discussions, and to Dr. Whit Athey for multiple suggestions and critical consideration of the manuscript.

References

Adamov, D.S., and Klyosov, A.A. (2008). 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.

Arabian Peninsula YDNA Project. URL: http://www.familytreedna.com/public/ArabianPeninsulaDNAProject, 2008.

Barac, L., Pericic, M., Klaric, I.M., Janicijevic, B., Parik, J., Rootsi, S., and Rudan, P. (2003a). Y chromosome STRs in Croatians. *Forensic Sci. Internat*. *138*, 127-133.

Barac, L., Pericic, M., Klaric, I.M., Rootsi, S., Janicijevic, B., Kivisild, T., Parik, J., Rudan, I., Villems, R., and Rudan, P. (2003b). Y chromosomal heritage of Croatian population and its island isolates. *Europ. J. Human Genetics 11*, 535-542.

Bortolini, M.-C., Salzano, F.M., Thomas, M.G., Stuart, S., Nasanen, S.P.K., Bau, C.H.D., Hutz, M.H., Layrisse, Z., Petzl-Erler, M.L., Tsuneto, L.T., 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., and 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., and Stanford, D. (2004). The North Atlantic ice-edge corridor: a possible Paleolithic route to the New World. *World Archaeology 36*, 459 – 478.

Cadenas, A.M., Zhivotovsky, L.A., Cavalli-Sforza, L.L., Underhill, P.A., and Herrera, R.J. (2008). Y-chromosome diversity characterizes the Gulf of Oman. *Eur. J. Human Genetics*, *18*, 374-386.

Campbell, K.D. (2007) Geographic patterns of haplogroup R1b in the Brutish Isles. *J. Genet. Geneal.*, *3*, 1-13.

Fagundes, N.J.R., Kanitz, R., Eckert, R., Valls, A.C.S., Bogo, M.R., Salzano, F.M., Smith, D.G., Silva, W.A., Zago, M.A., 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. Human Genet*. *82*, 583–592.

Haak, W., Brandt, G., de Jong, H.N., Meyer, C., Ganslmeier, R., Heyd, V., Hawkesworth, C., Pike, A.W.G., Meller, H., and Alt, K.W. (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, 52.

Karlsson, A.O., Wallerstrom, T., Gotherstrom, A. and Holmlund, G. Y-chromosome diversity in Sweden – A long-time perspective. *Europ. J. Human Genetics, 14,* 963-970 (2006)

Kivisild, T., Rootsi, S., Metspalu, M., Mastana, S., Kaldma, K., Parik, J., Metspalu, E., Adojaan, M., Tolk, H.-V., 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, A.A. (2008a). The features of the "West European" R1b haplogroup (in Russian). *Proc. Russian Academy of DNA Genealogy* (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 Academy of DNA Genealogy* (ISSN 1942-7484), *1*, 400-477.

Klyosov, A.A. (2008c). Origin of the Jews via DNA genealogy. *Proc. Russian Academy of DNA Genealogy* (ISSN 1942-7484), *1*, 54-232.

Klyosov, A.A. (2009). Iberian haplotypes, and analysis of history of populations of Basques, Sephards and other groups in the Iberian peninsula. *Proc. Russian Academy of DNA Genealogy* (ISSN 1942-7484), *2*, 390-421.

Lepper, B.T. (1999). Pleistocene Peoples of Midcontinental North America, in: Ice Age People of North America (Bonnichsen, R. and Turnmire, K., eds): Oregon State University Press, 362–394.

McEvoy, B., Simms, K., Bradley, D.G. (2008). Genetic investigation of the patrilineal kinship structure of early medieval Ireland. *Amer. J. Phys. Anthropol., 136,* 415-422.

McEvoy, B., Bradley, D.G. (2006). Y-chromosomes and the extent of patrilineal ancestry in Irish surnames. *Human Genetics, 119,* 212-219.

Mertens, G. (2007) Y-Haplogroup frequencies in the Flemish population. *J. Genet. Geneal. 3*, 19-25.

Pericic, M., Lauc, L.B., Klaric, A.M., et al. (2005). High-resolution phylogenetic analysis of southeastern Europe traces major episodes of paternal gene flow among Slavic populations. *Mol. Biol. Evol. 22,* 1964-1975.

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. Natl. Acad. Sci. US, 103*, 843-848.

Seielstad, M., Yuldasheva, N., Singh, N., Underhill, P., Oefner, P., Shen, P., and Wells, R.S. (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, P.J., Lin, A.A., Arbuzova, S., Beckman, L.E., De Benedictis, G., Francalacci, P., Kouvatsi, A., Limborska, S., Marcikiae, M., Mika, A., Mika, B., Primorac, D., Santachiara-Benerecetti, A.S., Cavalli-Sforza, L.L., Underhilll, P.A. (2000) The genetic legacy of paleolithic *Homo sapiens* in extant Europeans: a Y chromosome perspective. *Science* 290, 1155-1159.

Sengupta, S., Zhivotovsky, L.A., King, R., Mehdi, S.Q., Edmonds, C.A., Chow, C.-E. T., Lin, A.A., Mitra, M., Sil, S.K., Ramesh, A., et al. (2006). Polarity and temporality of high-resolution Y-chromosome distributions in India identify both indigenous and

exogenous expansions and reveal minor genetic influence of Central Asian Pastoralis. *Amer. J. Human Genet*. 78, 202-221.

Sharma, S., Rai, E., Sharma, P., Jena, M., Singh, S., Darvishi, K., Bhat, A.K., Bhanwer, A.J.S., Tiwari, P.K., Bamezai, R.N.K. (2009) The Indian origin of paternal haplogroup R1a1* substantiates the autochthonous origin of Brahmins and the caste system. J. Hum. Genet. 54, 47-55.

Thomas, M.G., Skorecki, K., Ben-Ami, H., Parfitt, T., Bradman, N., and Goldstein, D.B. (1998). Origins of Old Testament priests. *Nature 394*, 138-140.

Thomas, M.G., Parfitt, T., Weiss, D.A., Skorecki, K., Wilson, J.F., le Roux, M., Bradman, N. and Goldstein, D.B. (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.

Wells, R.S., Yuldasheva, N., Ruzibakiev, R., Underhill, P.A. Evseeva, I., Blue-Smith, L., Jin, L., Su, B., Pitchappan, R., Shanmugalaksmi, S., Balakrishnan, K., Read, M., Pearson, N.M., Zerjal, T., Webster, M.T., Zholoshvili, I., Jamarjashvili, E., Gambarov, S., Nikbin, B., Dostiev, A., Aknazarov, O., Zallous, P., Tsoy, I., Kitaev, M., Mirrakhimov, M., Chariev, A., Bodmer, W.F. (2001) The Eurasian heartland: a continental perspective on Y-chromosome diversity. *Proc. Natl. Acad. Sci. US* 98, 10244-10249.

Zhivotovsky, L.A., Underhill, P.A., 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. Human Genet. 74*, 50-61.

Figure Legends

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

Figure 2. A tree of 218 Irish 19-marker haplotypes (haplotypes were published in McEvoy et al, 2008), presumably of R1b1 haplogroup.

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).

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 haplotype with DYS388=10 (all others primarily have DYS388=12, in two cases DYS388=14, haplotypes 003 and 034).

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

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 composed from YSearch database.

Figure 7. The 25-marker haplotype tree for Russia and Ukraine, haplogroup R1a1. The 58-haplotype tree was composed from data of YSearch database.

Figure 8. The 25-marker haplotype tree for India, Pakistan and Sri-Lanka, haplogroup R1a1. The 22-haplotype tree was composed from data of YSearch database.

Figure 9. 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).

Figure 10. The 6-marker haplotype tree for the South Indian tribe Chenchu, haplogroup R1a1. The 11-haplotype tree was composed from data of Kivisild et al (2003).

Figure 11. The 6-marker haplotype tree for the Native Americans, haplogroup Q-M3 (Q1a3a). The 117-haplotype tree was composed from data of Bortolini et al (2003).

Figure 12. The 25-marker haplogroup J1 haplotype tree for 49 presumably Jewish haplotypes. The haplotypes were collected in YSearch database (Klyosov, 2008c).

Figure 13. 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 14. 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.

Figure 15. The 37-marker haplotype tree for Arabic haplotypes of haplogroup J1. 19 haplotypes were listed in the Arabic Peninsula YDNA Project (2008).

Figure 16. The 37-marker haplotype tree for J2 Jewish haplotypes. Twenty five haplotypes were collected in YSearch database (Klyosov, 2008c).

Figure 17. The 9-marker haplotype tree for 34 Croatian (apparently, Gypsy) haplotypes of haplogroup H1. The haplotype tree was composed from data (Barac et al., 2003a, 2003b; Pericic et al., 2005).

Figure 18. The Lemba 6-marker haplotype tree, apparently of haplogroup J. The 57 haplotype tree was composed of data published in (Thomas et al., 2000).

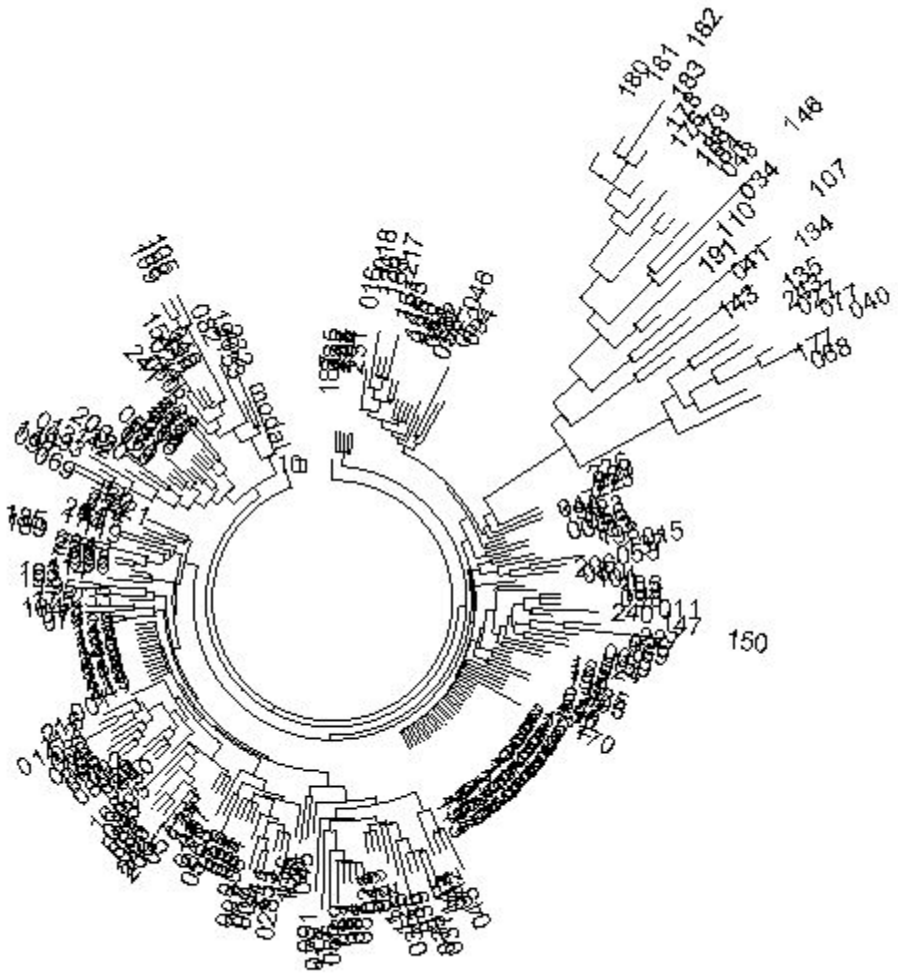

Figure 1

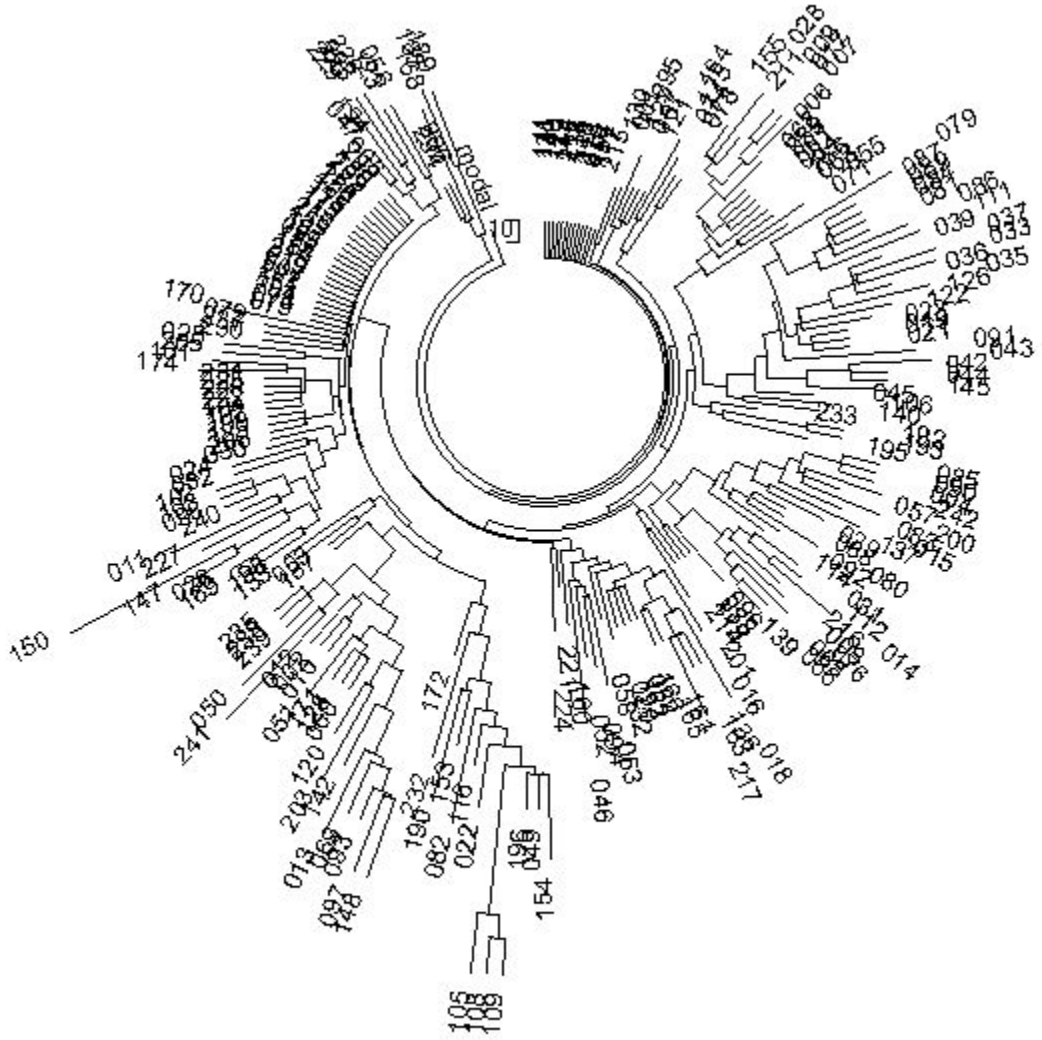

Figure 2.

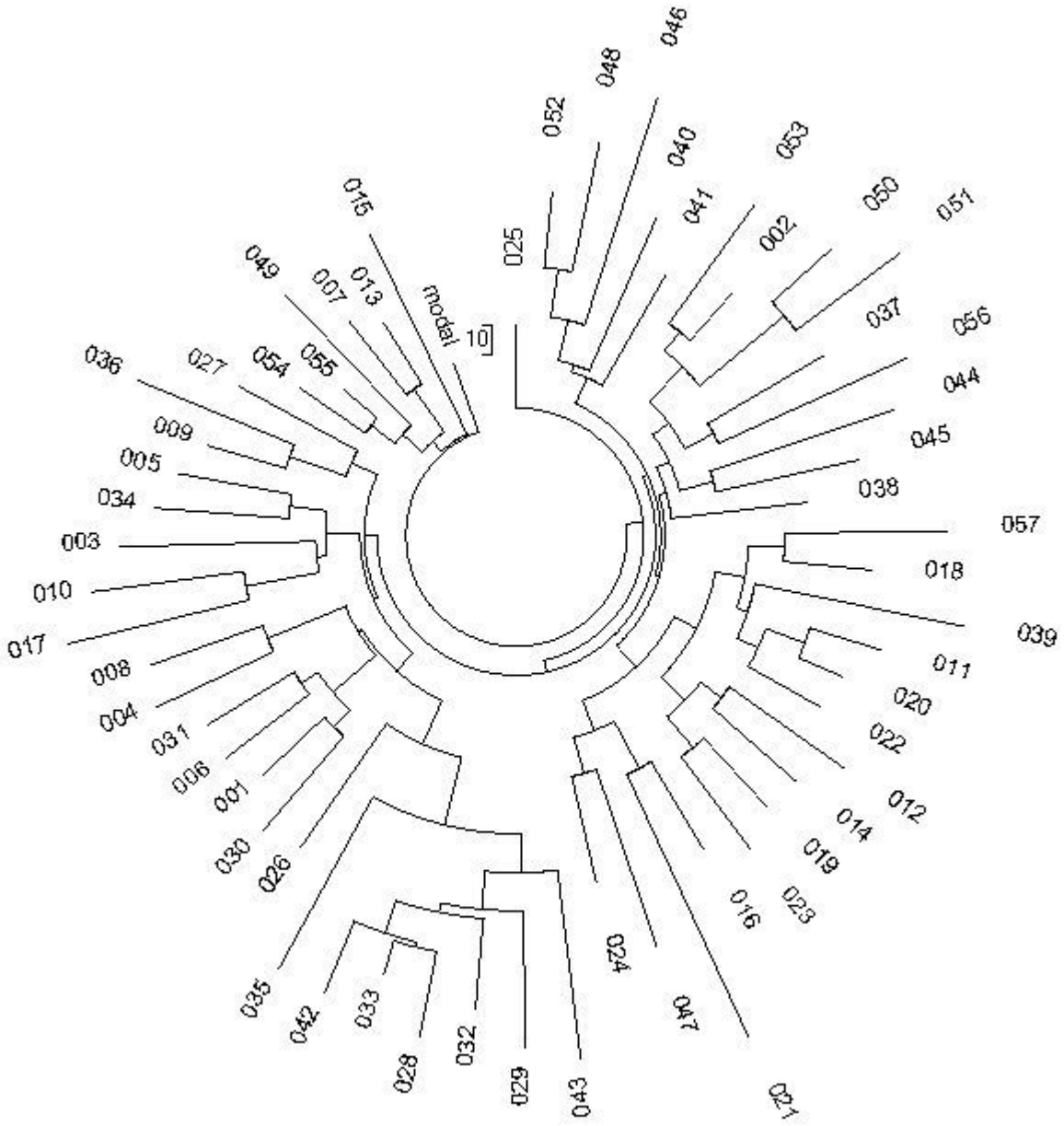
modal 10
015
049
007
013
025
052
048
046
040
041
053
002
050
051
037
056
044
045
038
057
018
039
011
020
022
012
014
019
023
016
024
047
021
043
029
032
028
033
042
035
030
026
001
006
031
004
008
017
010
003
034
005
009
036
027
054
055

Figure 3

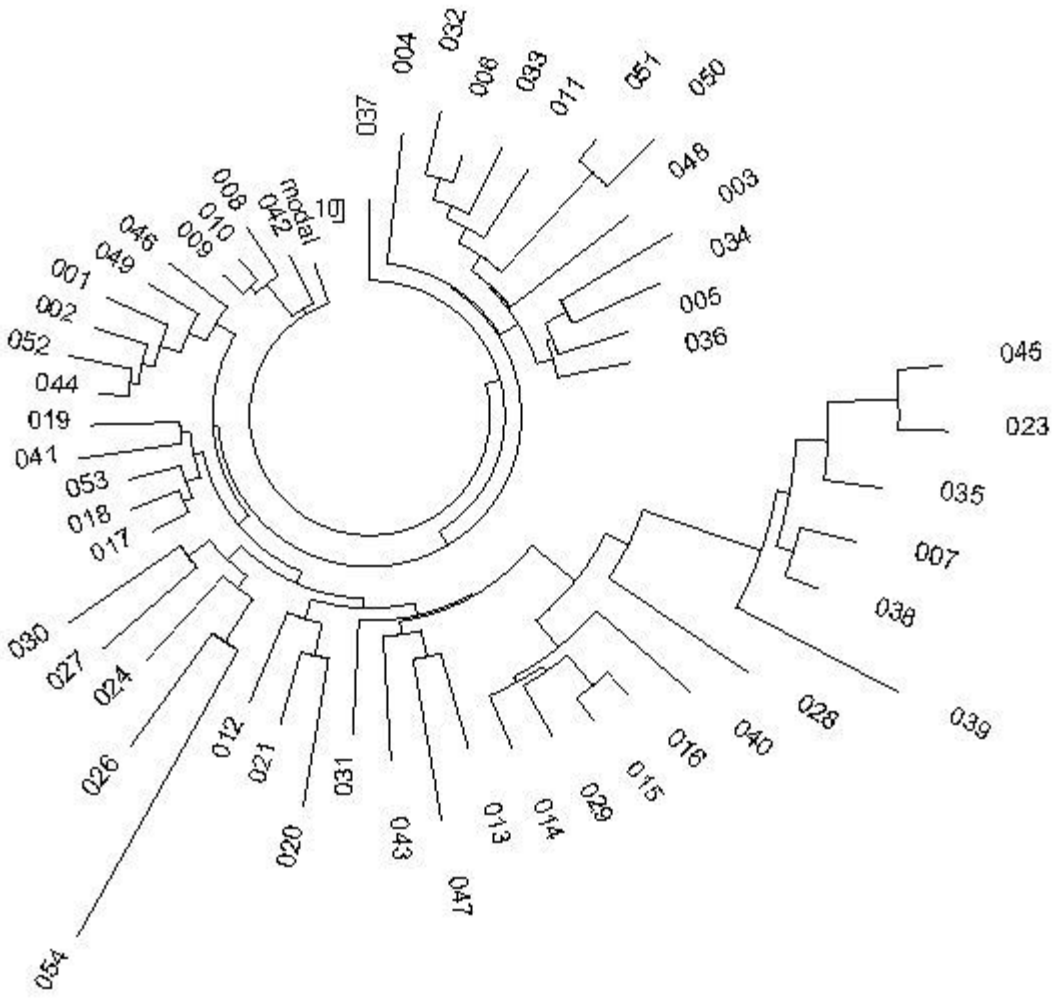
037
004
032
006
033
011
051
050
048
003
034
005
036
045
023
035
007
038
039
028
040
016
015
029
014
013
047
043
031
020
021
012
026
024
027
030
017
018
053
041
019
044
052
002
001
049
046
009
010
008
042
modal
1日

Figure 4

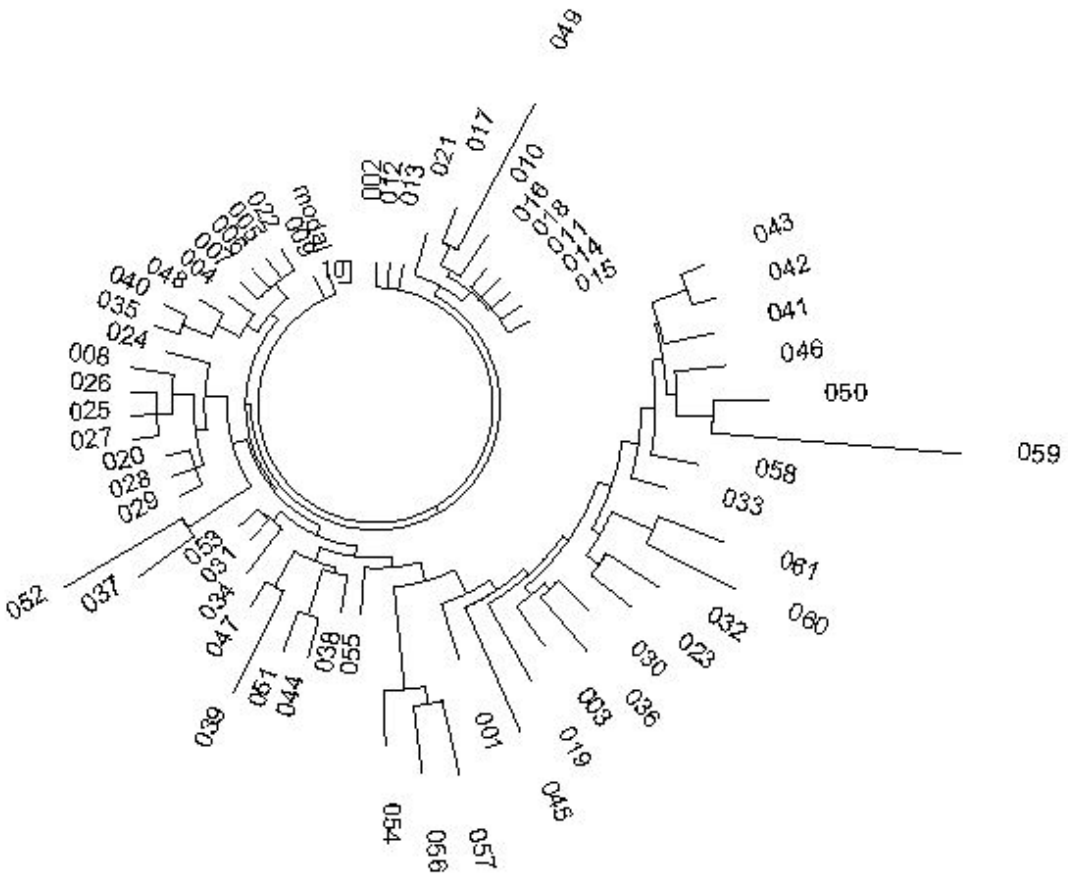
049
002
012
013
021
017
010
018
016
011
014
015
043
042
041
046
050
059
058
033
061
060
032
023
030
036
003
019
045
001
057
056
054
039
051
044
038
055
047
034
053
031
037
052
029
028
020
027
025
026
008
024
035
040
048
004
005
007
006
009

Figure 5

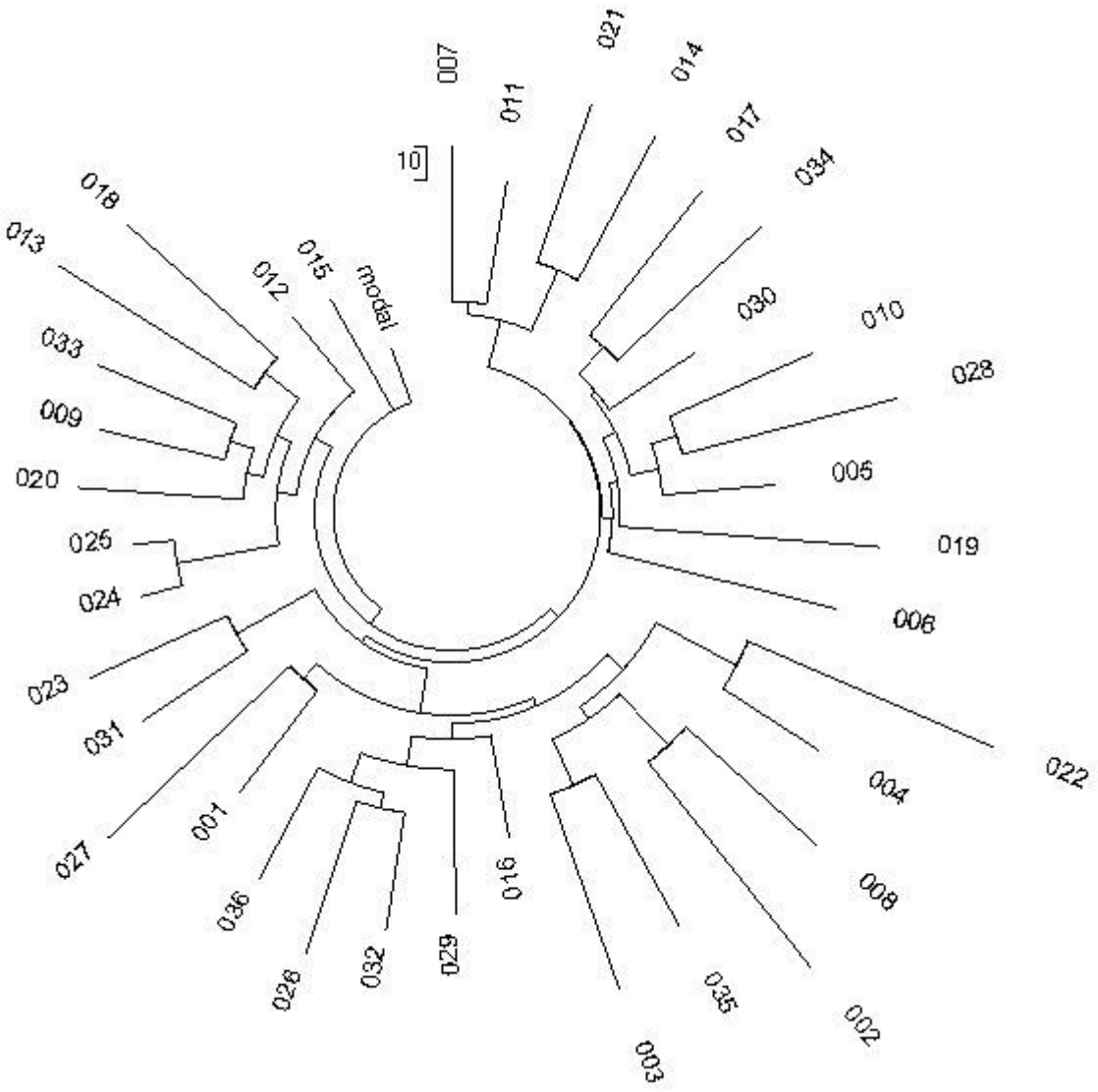
007
011
021
014
017
034
030
010
028
005
019
006
022
004
008
002
035
003
029
016
032
026
036
001
027
031
023
02A
025
020
009
033
013
018
012
015
modal
10

Figure 6

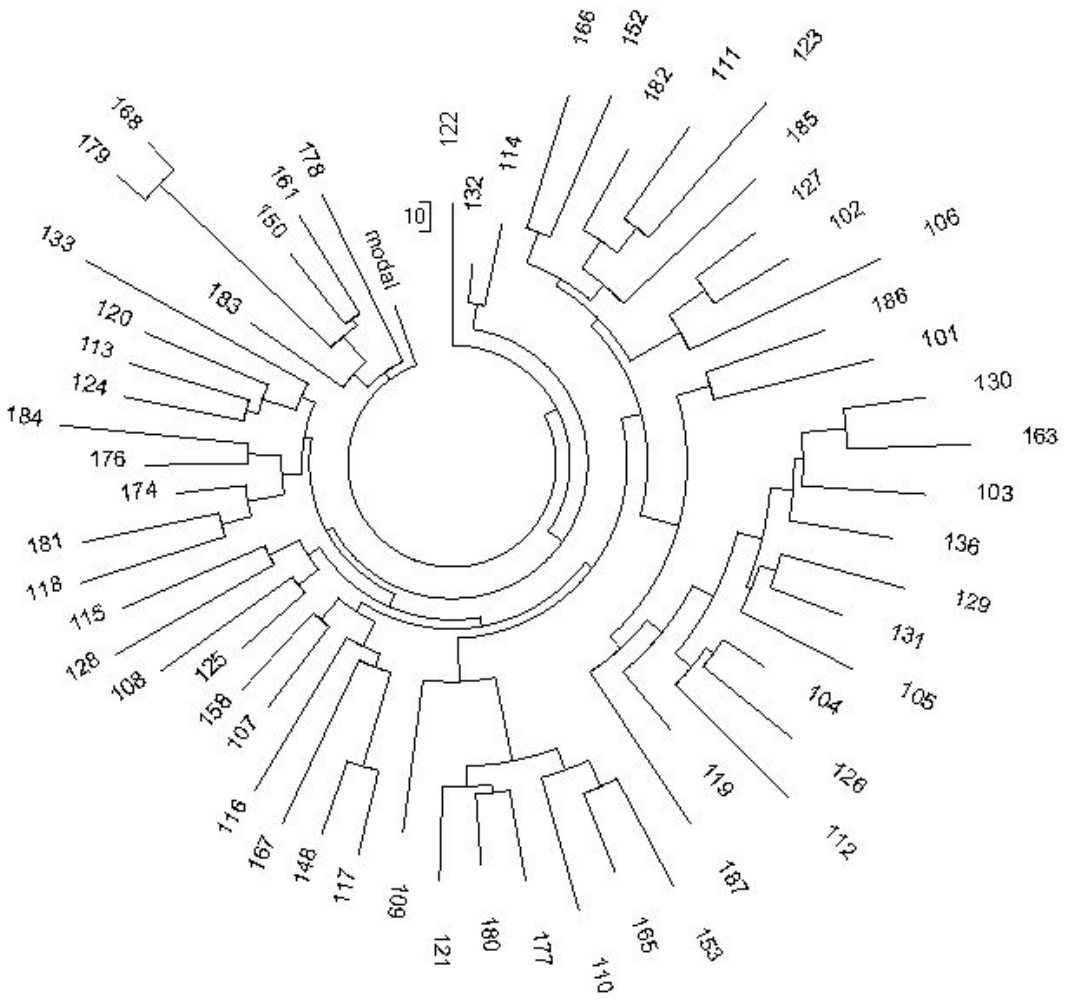
modal
10
122
132
114
166
152
182
111
123
185
127
102
106
186
101
130
163
103
136
129
131
105
104
126
112
119
187
153
165
110
177
180
121
109
117
148
167
116
107
158
125
108
128
115
118
181
174
176
184
124
113
120
183
133
150
161
178
179
168

Figure 7.


021
10
014
020
002
016
018
008
015
009
013
007
011
004
010
003
005
012
001
022
019
017
006
modal

Figure 8.

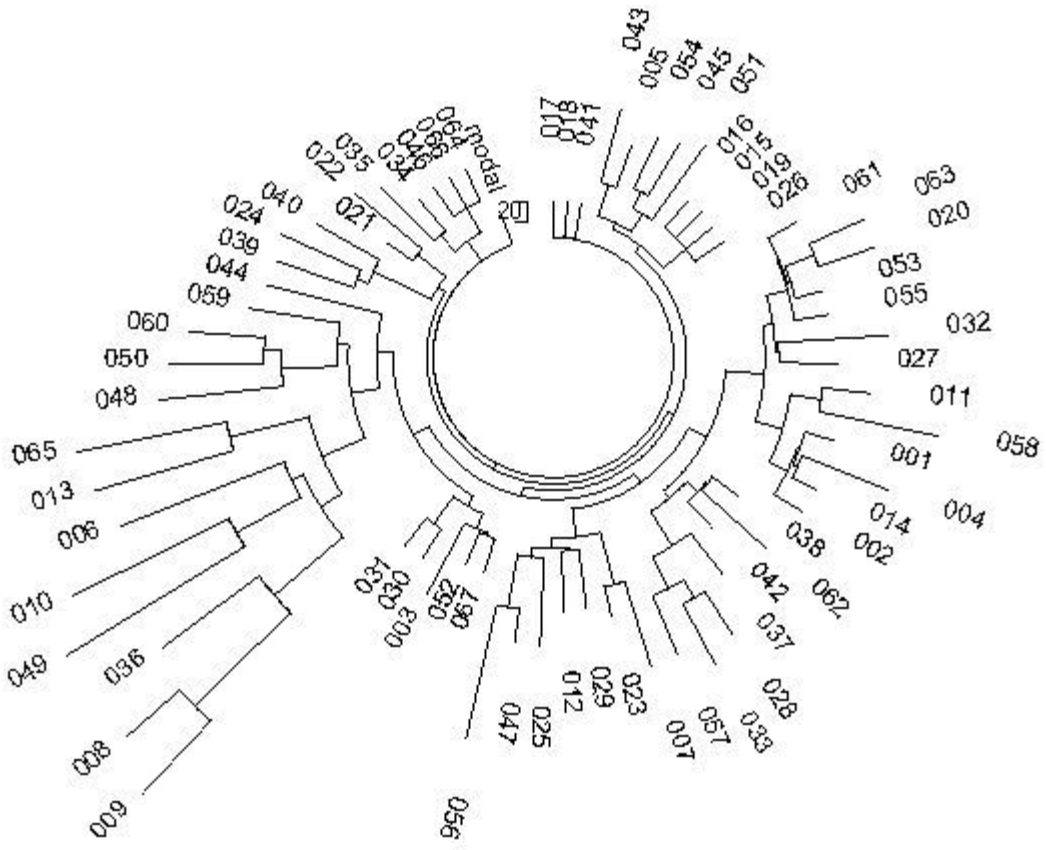
043
005
054
045
051
017
018
041
016
015
019
026
061
063
020
053
055
032
027
011
001
058
004
014
002
038
042
062
037
028
033
057
007
023
029
012
025
047
056
052
067
037
030
003
065
013
006
010
049
036
008
009
060
050
048
059
044
039
024
040
021
022
035
064
modal

Figure 9

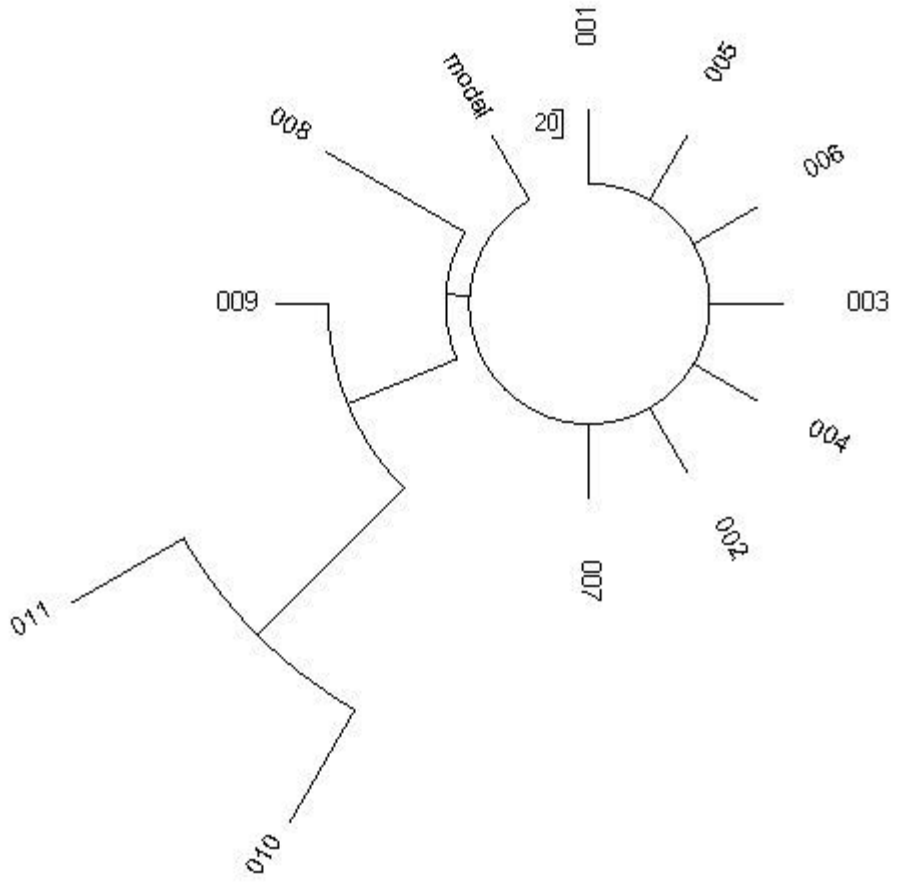

Figure 10

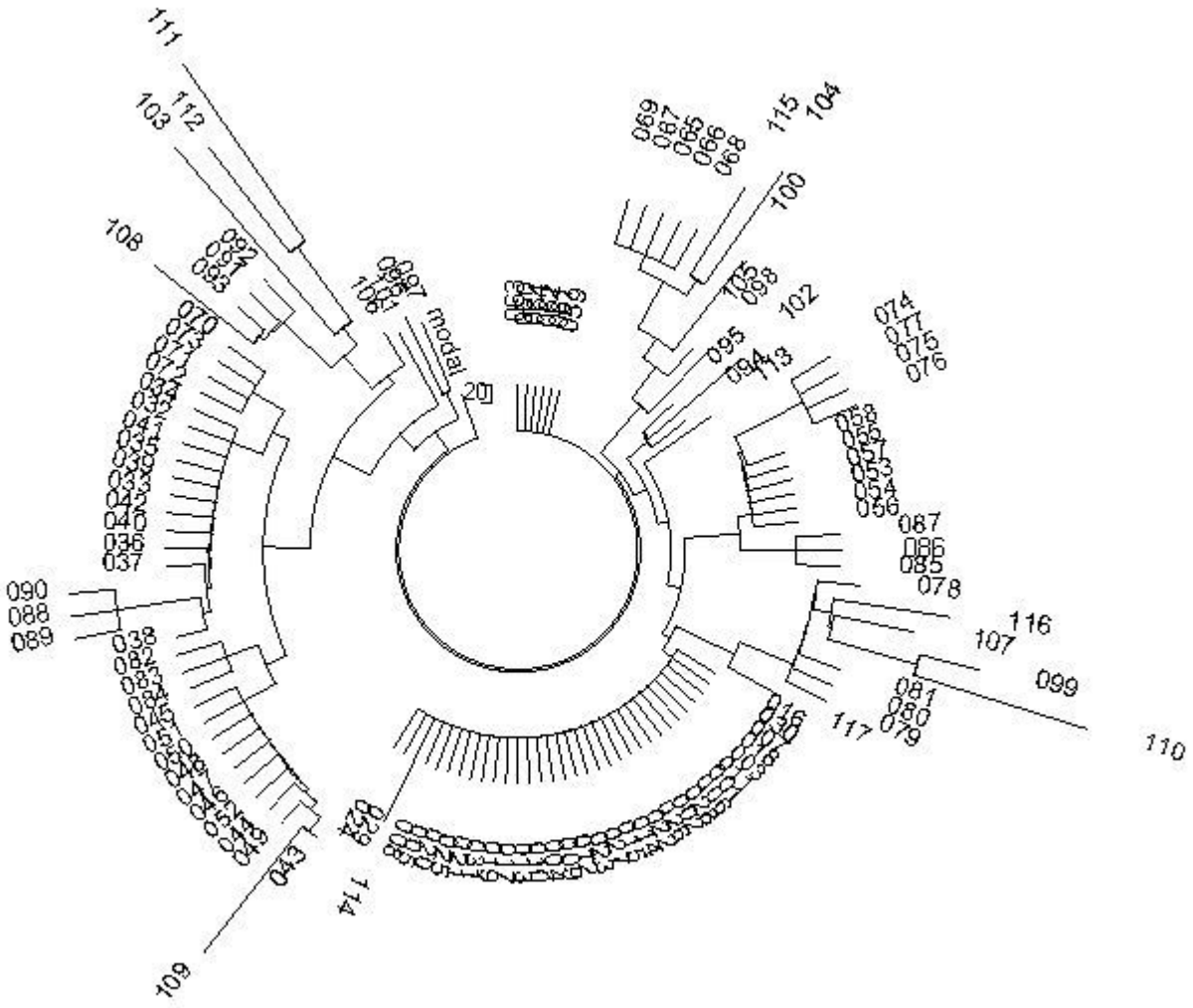
modal
25
111
112
103
108
109
090
088
089
043
114
029
117
110
099
116
107
078
087
086
085
115
104
100
102
074
077
075
076
113
095
094
105
098

Figure 11

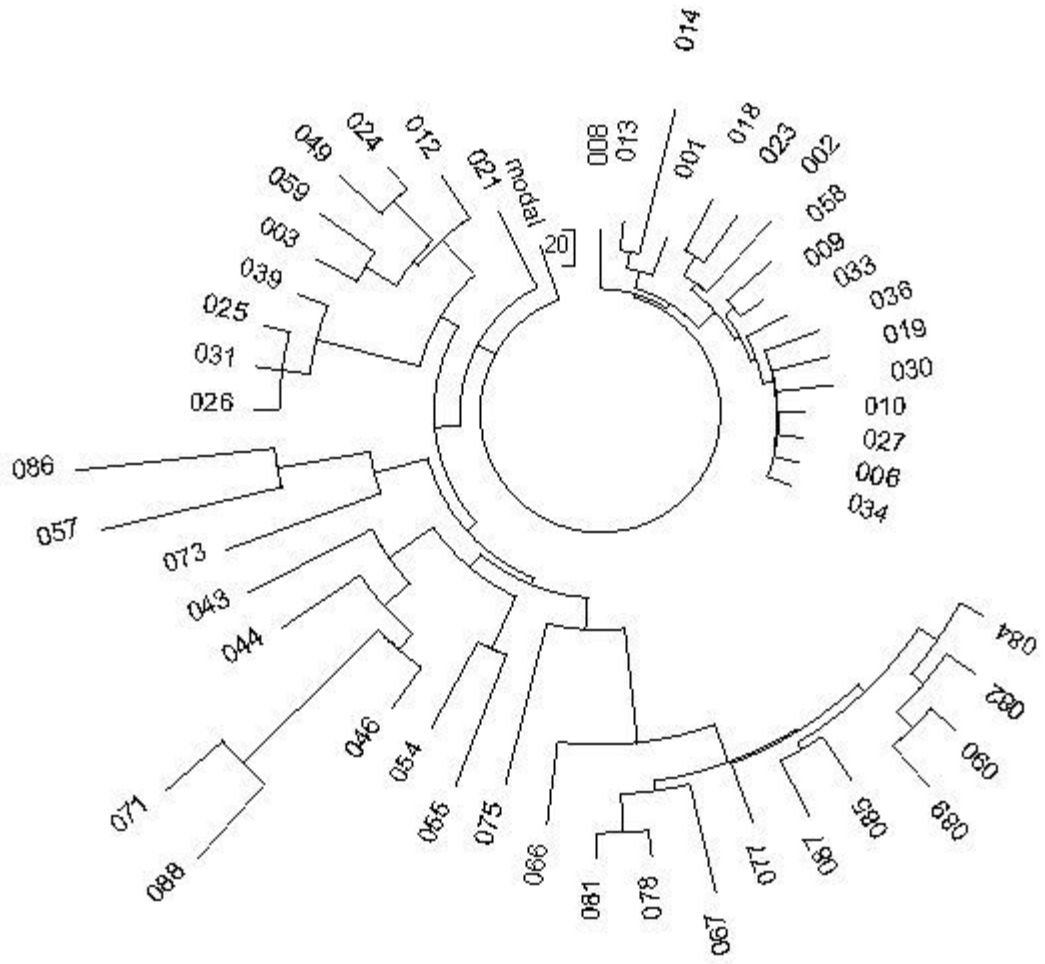
014
008
013
001
018
023
002
058
009
033
036
019
030
010
027
006
034
modal
20
021
012
024
049
059
003
039
025
031
026
086
057
073
043
044
046
054
065
075
066
081
078
067
071
088
077
080
085
089
090
082
084

Figure 12

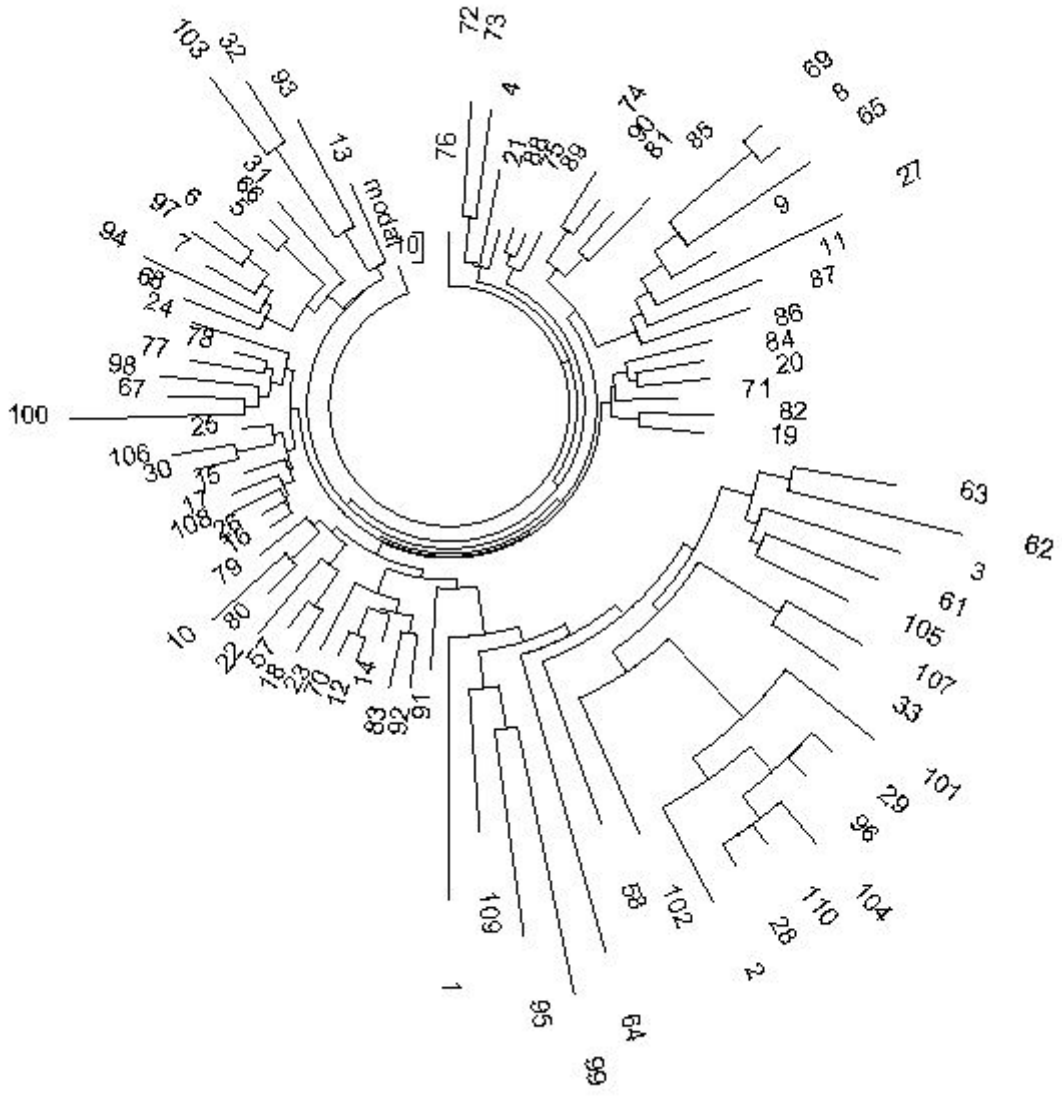

Figure 13

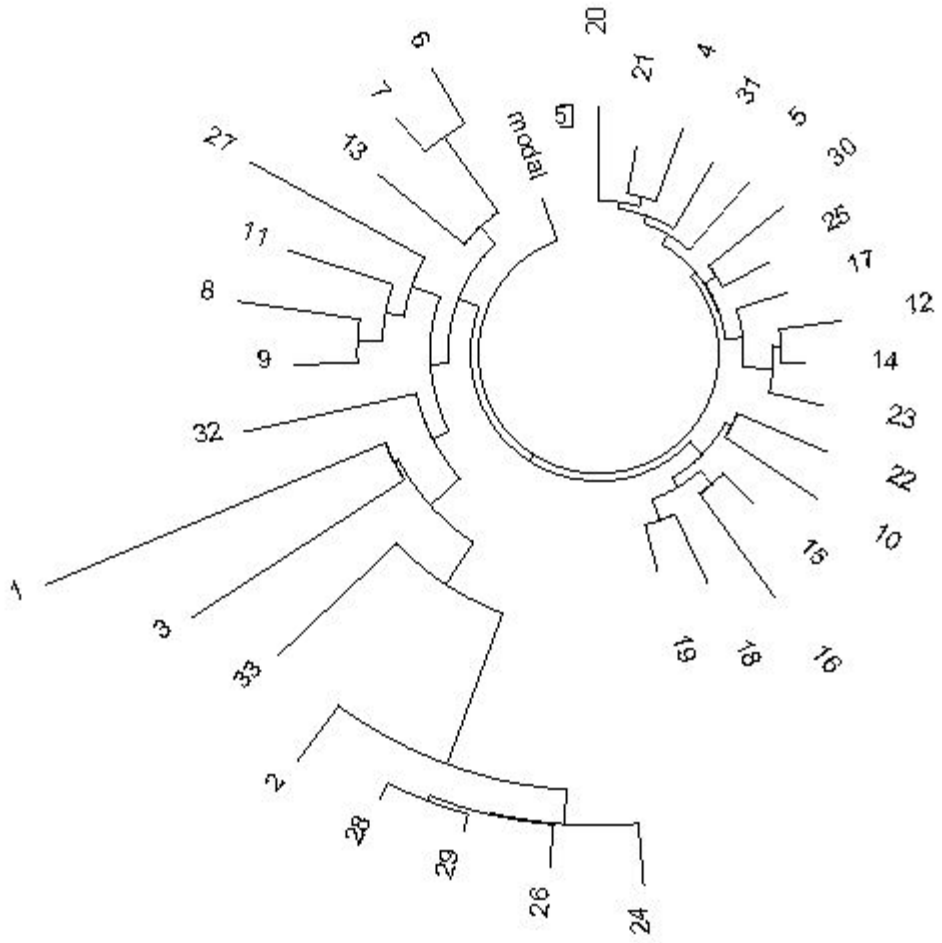
20
21
4
31
5
30
25
17
12
14
23
22
10
15
16
18
19
6
7
13
27
11
8
9
32
1
3
33
2
28
29
26
24
mdal

Figure 14

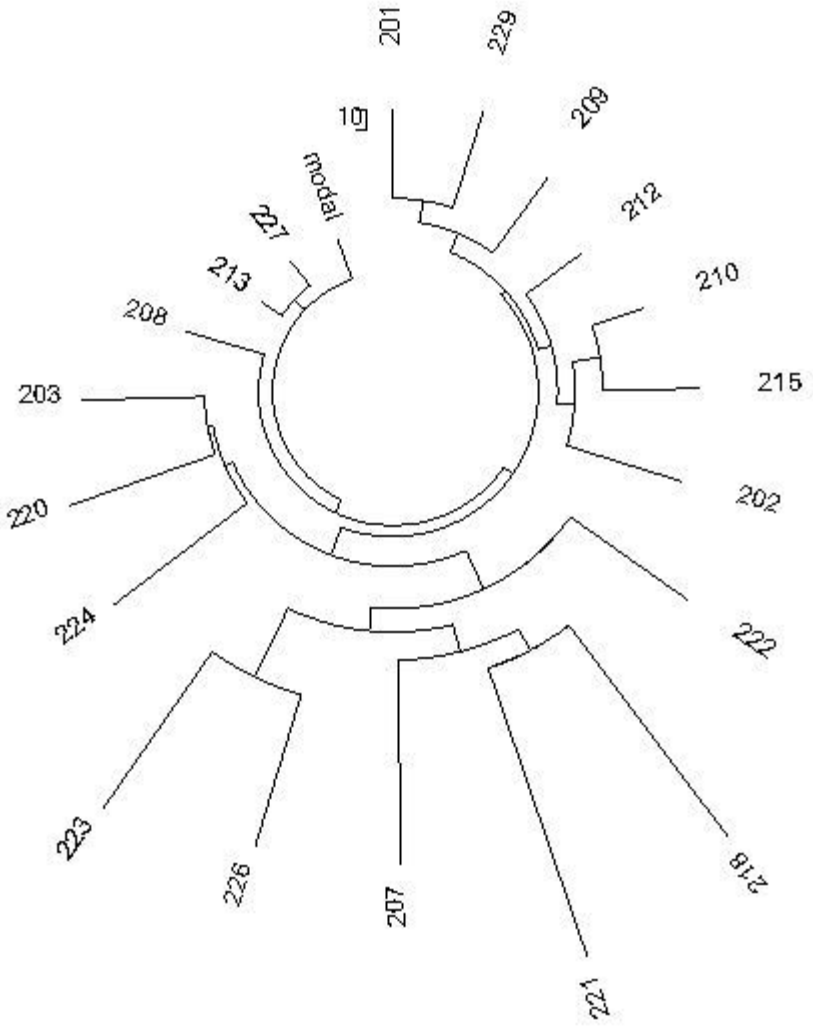
201
229
209
212
210
215
202
222
218
221
207
226
223
224
220
203
208
213
227
1g
lepdou

Figure 15

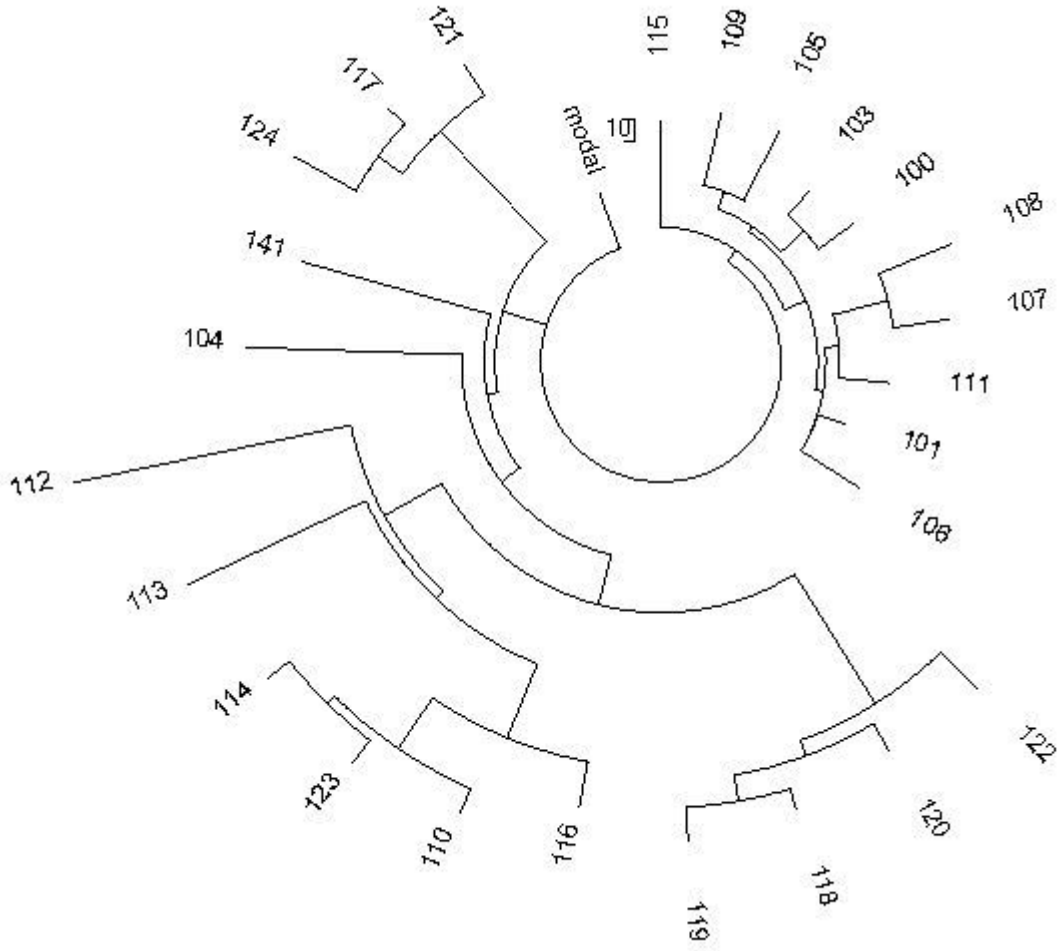
121
117
124
141
104
112
113
114
123
110
116
119
118
120
122
106
101
111
107
108
100
103
105
109
115
model
19

Figure 16

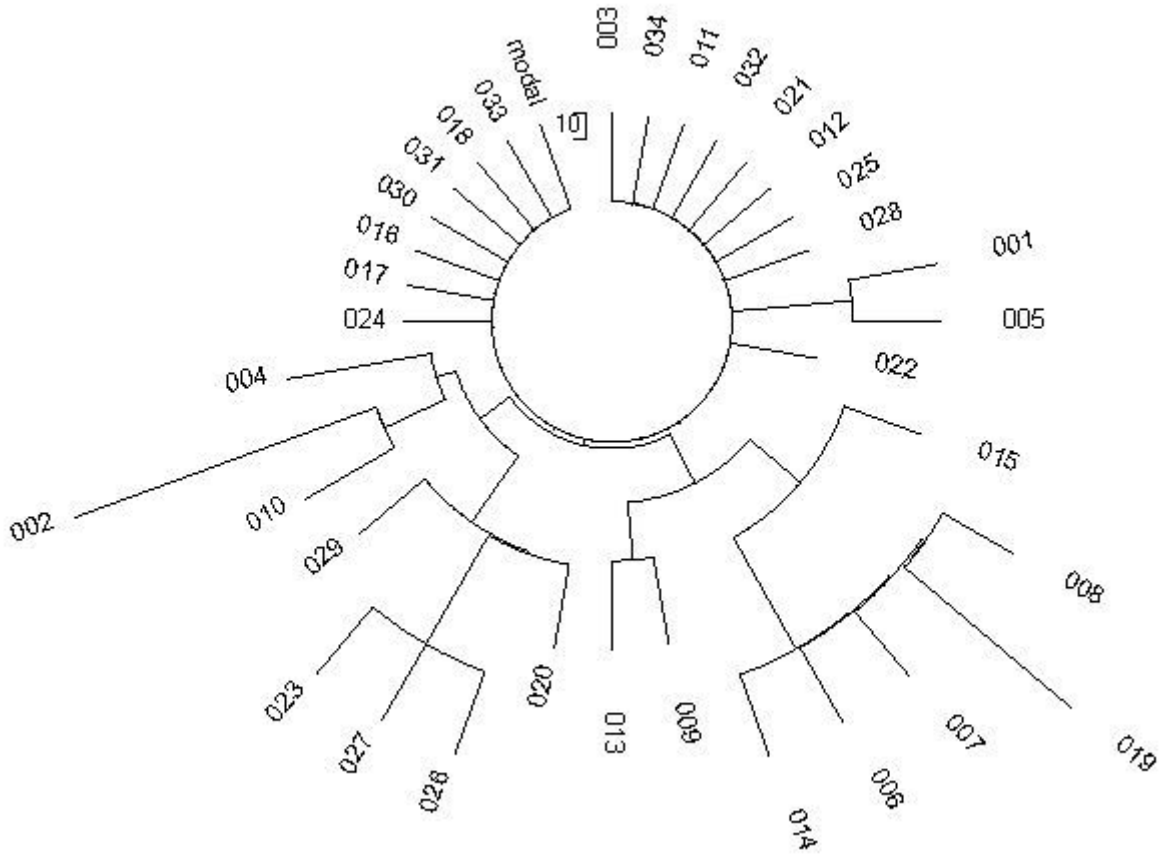
003
034
011
032
021
012
025
028
001
005
022
015
008
019
007
006
014
009
013
020
026
027
023
029
010
002
004
024
017
016
030
031
018
033
modal
1

Figure 17

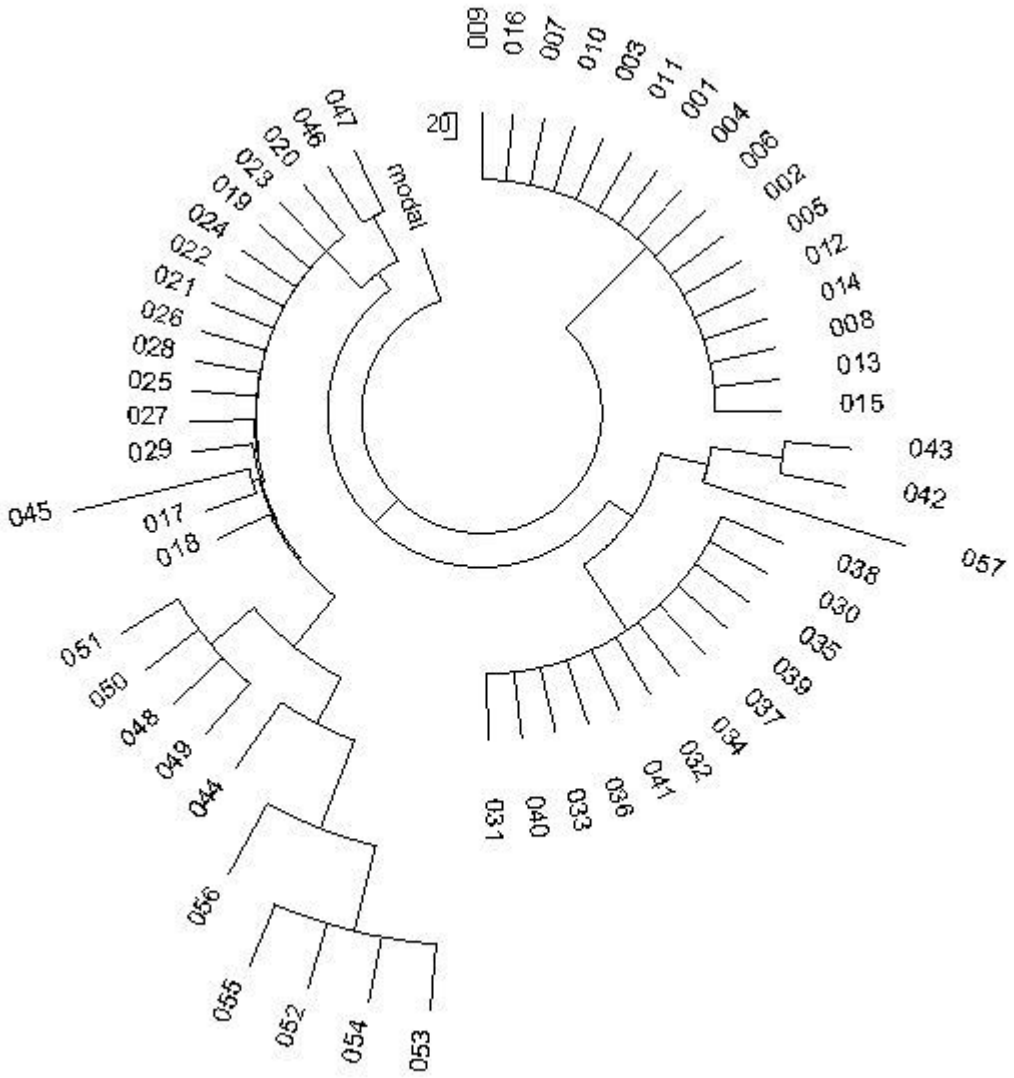
009
016
007
010
003
011
001
004
006
002
005
012
014
008
013
015
043
042
057
038
030
035
039
037
034
032
041
036
033
040
031
053
054
052
055
056
044
049
048
050
051
045
017
018
029
027
025
028
026
021
022
024
019
023
020
046
047
model

Figure 18