Ultima ratio

Proceedings of the Russian Academy of DNA Genealogy

Volume 1, No. 1

(Special Edition)

2008 June

Russian Academy of DNA Genealogy

ISSN 1942-7484

Proceedings of the Russian Academy of DNA Genealogy, Special Edition, Lulu inc., 2008.

Publisher
Russian Academy of DNA Genealogy

Production Coordinator
Michael Temosh

Address editorial correspondence to
temosh@gmail.com

Foreword

Dear Reader,

The very fact that you have opened this new edition tells you and me, that you either work in DNA genealogy, or have an interest in it. In both cases, you cannot be wrong.

This issue of the *Proceedings of the Russian Academy of DNA genealogy* launches the Academy. At the moment the Academy (RA-DNA) is just a small group of enthusiasts in the area. We represent different professions, in the fields of science, engineering, history, linguistics, computer programming, art, and so on. It makes our tentative (as a rule) conclusions sometimes a bit shallow, but we cover a wide area, which is much wider than any geneticist or a "DNA genealogist" can cover, leaving alone history and linguistics. Our goal is not re-write history, linguistics, archeology, of course – this would be silly – but to attract attention of professionals in those disciplines to a treasury and haplotypes and haplogroups, and to explain them how to handle it, what information we and they can extract from it, and to help them to interpret data.

This is the Academy is created for.

The *Proceedings* will publish multi-disciplinary papers related to DNA genealogy. The main criterion is that the papers should be related to haplotypes and haplogroups, and the data are new and/or thoughts are original. The articles will be published in Russian or in English, depending on the language of the submitted article.

The *Proceedings* will also contain the DISCUSSION section, where some selected topics are to be published, taken from http://www.dnatree.ru/index.php

A few words on the Academy itself. Fellows of the Academy are selected from people who have their haplotype/haplogroup tested, who take part in active discussions in http://www.dnatree.ru/index.php, who publish papers in this *Proceedings* and/or in professional editions in the field of DNA genealogy or other editions one can make a reference to, who is willing to become a Fellow of the Academy and who is elected by other Fellows of the Academy. Any individual in the world can become a Fellow, provided that the above requirements are accomplished.

The Academy is a non-profit organization, however, their Fellows can work in for-profit companies, and the Academy can apply to for-profit companies if it helps to find an answer to specific questions. Fellows of the Academy can provide consulting word for a fee, it does not present any conflict of interest with the Academy. The main *modus of operandi* of the Academy is exchanging of information and generating knowledge.

This particular issue of the *Proceedings* is a Special Edition, because due to technical reasons it contains articles written by one individual, and these articles were written in English. In plain language, other would-be Fellows did not make their papers ready for publication in this issue as yet.

Last but not least. A motto of the Academy is *Ultima Ratio*, that is *The Final Argument*. Hopefully, that is how historians, linguists, archeologists and alike should (and will) look at haplotypes and haplogroups, provided, of course, that the haplotypes collected and processed properly. The story goes that these words – *Ultima Ratio* – were engraved on French cannon by order of Louis XIV. We would go that far.

Regards,

Anatole Klyosov

June 2008

Basic Rules of DNA Genealogy (Y-Chromosome), Mutation Rates and their Calibration

Anatole A. Klyosov

Newton, Massachusetts 02459, U.S.A.
http://aklyosov.home.comcast.net

Summary

An overview of Ten Commandments of DNA genealogy is presented:

1. You shall have no genes before non-recombinant DNA fragments
2. You shall not make wrongful calculations of mutations in haplotypes
3. You shall not use wrongful mutation rates
4. Honor your common ancestors and calculate timespans to their times
5. You should not make for yourself a haplotype of a certain length an idol
6. You shall not commit calculations of time spans to ancestors based on just a pair of haplotypes. You shall honor statistics.
7. You shall not covet a neighbor haplogroup when calculate mutations in haplotypes
8. Remember the base (ancestral) haplotype and keep it holy
9. You shall not mix STRs and SNPs
10. You shall verify if a given haplotype set is derived from one common ancestor or from multiple ones. I am giving you an imperative how to do it.

A new approach for kinetics of haplogroup mutations is suggested, examined and verified using many examples of haplotype sets related to haplogroups R1a1, R1b, R1b1c, I1, I2b1, R2, F, H1, C2, Q, J*, J1, J2, N3, N3a, N3a1, G and others. The principal approach is based on a comparison of a number of generations (time spans) to an alleged common ancestor calculated from (a) a number of mutations in the haplotype set, and (b) a number of base (ancestral) haplotypes in the set. In the second case a number of mutations in haplotypes is not required for calculations of a time span to a common ancestor. Only when the two figures (numbers of generations) match each other, they indicate a single common ancestor for the given haplotype set, and the calculations are correct. If the two figures do not match each other, the haplotype set is derived from more than one "common ancestor" and results of the calculations are erroneous.

A list of mutation rates is presented and calibrated – using known historical events and "classical" genealogy – for 6-, 9-, 10-, 12-, 25, 37-, and 67-marker haplotypes. Many examples of calculated time spans to common ancestors for various haplotype sets of the above haplogroups from different territories are provided, and some revelations to their history are provided.

Introduction

Origin of peoples in the context of DNA genealogy is an assignment of each of them to a particular tribe or its branch (lineage) initiated – in a genealogical sense – by a particular ancestor who had a base haplotype (an "ancestral" haplotype). This also includes an estimation of a time span between the common ancestor and its current descendants. If information obtained this way can be presented in a historical context, and supported, even arguably, by other independent archeological, linguistic, historical, ethnographic, anthropological and other related considerations, this can be called a success. This is a step ahead in understanding our roots in the historical and genealogical context.

Since in the subsequent study, published in this issue, I will operate by terms of DNA genealogy, I will remind and systematize the principal statements of this discipline. The main reason for it is explained by the golden rule of science: extraordinary conclusions require extraordinary justifications.

Also, because of the subsequent study, published in this issue, deals with the Jewish DNA genealogy, a number of illustrations and examples in this paper consider Jewish haplotypes.

Let us consider the main principles, or statements of DNA genealogy.

The main Statements of DNA Genealogy (in the context of this paper)

First – Fragments of DNA considered in this work have nothing to do with genes.

Second – copying of the Y chromosome from father to son results in mutations of two kinds, SNP (single nucleotide polymorphism, or "snips", which are certain inserts and deletions in Y chromosomes) and mutations in STR (short tandem repeats), which make them shorter or longer by certain blocks of nucleotides. A

DNA Y chromosome segment (DYS) containing an STR is called a locus, or a marker. A combination of certain markers is called a haplotype.

Third – All people have a single common ancestor, who lived – on various estimates – between 80,000 and 200,000 years ago. This time is required to explain variations of haplotypes in all tested humans.

Fourth – Haplotypes can be practically of any length. Typically, the shortest haplotype considered in DNA genealogy is a 6-marker haplotype. It is the most common in publications on DNA genealogy. We will call it a "scientific" notation, since this 6-marker format is typically used in peer-review publications. For example, a typical
Jewish 6-marker haplotype (in haplogroups J1 and J2) is

14-16-23-10-11-12

which is also called the "Cohen Modal Haplotype" (CMH). It shows a number of tandem repeats (as alleles) in markers of DYS, numbered as DYS#19, 388, 390, 391, 392, and 393. When this or any other haplotype is transmitted from father to son, a mutation in it can happen with a probability of about 1/100 (for a 6-marker haplotype), that is one mutation per hundred births, such as a mutation in the third marker

14-16-**24**-10-11-12

or in the first marker

15-16-23-10-11-12

or any other mutation (one-step, as a rule, but not always) in any marker of this or any other 6-marker haplotype.

12-marker haplotypes are also often considered in DNA genealogy, however, they are rather seldom in professional publications. For example, a common Jewish 12-marker haplotype (a 12-marker CMH in this particular case) is

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

In this case the order of markers is a different one compared to that in the above 6-marker haplotypes, and it corresponds to the so-called FTDNA standard notation: DYS 393, 390, 19, 391, 385a, 385b, 426, 388, 439, 389-1, 392, 389-2.

Moving to 37-marker haplotypes, here is an example of a haplotype which belongs to a descendant of an individual with a last name Cohen, who lived about 150 years ago in St. Petersburg, Russia:

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

Naturally, the longer the haplotype, the easier to find a close relative living today, provided that close relatives have their haplotypes tested. Notice the mutation in the ninth marker DYS#439 (from 11 to 12, marked in bold) compared to that in the 12-marker haplotype given above. In fact, that 12-marker haplotype above (with DYS#439 = 11) belongs to a Jewish person who currently lives in England, many generations since the birth of his ancestor in 1799, also in England. His complete 37-marker haplotype is as follows:

12-23-14-10-13-15-11-16-**11**-13-11-30-17-8-9-11-11-26-14-21-27-12-14-16-**16**-11-10-22-22-15-14-**17**-**17**-31-35-**14**-10

Compared to the descendant of Mr. Cohen from St. Petersburg, his British distant relative has seven more mutations (besides the one in the 12-marker haplotype), total eight, namely 12→11, 17→16, 21→17, 18→17, and 13→14. Eight, because one mutation is a four-step one; it could have happened as four successive mutations in one marker, but not necessarily.

In a simplified way, eight mutations between a pair of 37-marker haplotypes correspond to a time span of about 100 generations between the haplotypes, or about 2,500 years between them. In other words, their alleged common ancestor lived half-way of this time span, or about 1,250 years BP, around the 8th century AD. By the way, one mutation in a 12-marker haplotype corresponds to a time span of approximately 1,100 years, hence, a rather fair match between these two estimates.

However, this match should be considered as largely coincidental one, due to the next statement:

Fifth – one cannot calculate time to a common ancestor based on just a pair of haplotypes. Mutations are ruled by statistics, and they can be analyzed quantitatively (often, semi-quantitatively) only using numerous haplotypes. The smaller the number of haplotypes in a set, the less reliable the result. With a pair of haplotypes, one or two mutations can appear in the latest or a recent generation. This would mistakenly lead to an addition or subtraction of a thousand years compared with just one generation. In a large set of haplotypes

recent mutations are balanced by statistics across all the mutations, both recent and old.

How often haplotypes mutate? For STR, tandem repeat mutations, an average mutation rate in a 6-marker haplotype is 0.0096 mutations per haplotype per a generation, as it will be shown and calibrated below. It means that a mutation in a 6-marker haplotype occurs – on average – once per about 100 generations, that if once in 2500 years, if to assume 25 years for a generation. In fact, it would be not 100 generations, but 116 generations, that is 2,900 years, if to take so-called back mutations into account. This will be explained below. In a 12-marker haplotype an average tandem mutation rate is 0.024 per haplotype per a generation, that is a mutation happens once over a time span of 40 generations, a thousand years. Again, if to consider back mutations, it will be 43 generations, that is about 1,075 years. SNP mutations, snips, occur once and practically forever. They serve as tags for haplogroups (see below).

That leads us to the following statement –

Six – an average number of STR (tandem) mutations per haplotype can serve for calculations of a time span elapsed from the common ancestor for all haplotypes in the set, if they all derived from the same common ancestor. That ancestor had a so-called base, or ancestor haplotype. However, very often haplotypes in a set are derived not from one common ancestor. I will present here criteria which can tell if it is true or not (see below).

Since descendants retain the base haplotype, which is relayed along the lineage, from father to son, and mutations in haplotypes happen on average once in about 2900 years (for 6-marker haplotypes) or about once in 1075 years (for 12-marker haplotypes), then even after 5000 years descendants retain 20% of the base, ancestral, unchanged 6-marker haplotype. If to consider 12-marker haplotypes, 20% of them will still be the base ones after 70 generations, that is 1750 years.

Therefore, an ancestral haplotype can be identified after many centuries and even after several millennia, depending on a number of markers in the haplotype. And its look/appearance, structure, sequence of alleles might provide a hint where this haplotype came from.

This follows from the next statement:

Seven – Haplotypes often (but not always) are bound to certain territories. In ancient times people commonly migrated by tribes. A tribe was a group of people typically related to each other. They have the same or similar haplotypes, since

one mutation – on average – happens once in centuries, or even thousands of years. Sometimes a tribe population was reduced to a few, or even to just one individual, passing though a bottleneck, as it is said, and if the tribe survived, remaining people (or just an individual from the tribe), having certain mutations in their haplotypes, have passed their mutations to the offspring. Many tribe members were leaving the tribe voluntarily or by force – as prisoners, escapees, through journeys, military expeditions, and survivors continued, or rather initiated a new tribe on a new territory. As a result, a world map – in terms of DNA genealogy – is rather spotty, and each spot often has its own prevailing haplotype, sometimes a new, mutated haplotype. The most popular haplotype on a territory is called a modal haplotype. It often, but not necessarily, represents the ancestor's haplotype.

SNP mutations, snips, are practically permanent. Once they appear, they stay. Theoretically, some other mutations can happen at the same spot, in the same nucleotide, changing the first one. However, with millions of nucleotides such an event is very unlikely, but not impossible. There are more than three million chromosomal SNP's (The International HapMap Consortium, 2007), and DNA genealogists have employed a few hundreds of them. As it was said, they are called snips. Anyway, those who had left their tribes voluntarily or by force, always had snips in their Y chromosome, along with many of different ones. All those snips were passed on to their sons, if any. From those several hundred snips which were employed in DNA genealogy, more than 200 were found to be useful labels of certain populations on the globe. These populations were coined by letters from A to T in an assumed order of their appearance in mankind (though many revisions are foreseen), and called haplogroups and sub-groups. Thus, the next statement:

Eight – people can be assigned to their original tribes (that is, tribes of their ancestors) not only (and not much) based on their haplotypes, but based on their snips, which in turn lead to their haplogroups. For example, haplogroups A and B – African, the oldest ones, haplogroups C is a mongoloid one, as well as of a significant part of Native American Indians, descendants of mongoloids; haplogroups J – a Middle Eastern one, with J1 is mainly Jewish and Arabic, and J2 is mainly a Mediterranean one, including many Jews; haplogroups N is carried by many Siberian peoples as well as those from the North of Europe as well as many American Indians who have originated from Siberians carrying those haplogroups. Haplogroup R1b and its subgroups are observed mainly (but not only) in Western Europe, R1a1 mainly in Eastern Europe and Asia, with about a third of the population in India, where R1a1 was brought from Eastern Europe/Western Asia. Haplogroup R1a1 also marks some areas in Central Asia

(particularly in Kirgizstan and Tadzhikistan) where R1a1 was brought thousands of years ago from Eastern Europe.

In other words, each man has snips from a certain set, recognized by DNA genealogy, which assigns the man to a certain ancient tribe. Snips, typically employed in DNA genealogy, range – time wise – from 40-50 thousand years for "old" snips to 10-15 thousand years for "young" ones. It is assumed, for example, that the J1 haplogroup (SNP M267) has appeared some 12-14 thousand years back, though not proven as yet. There will be more about it in the subsequent paper in this issue.

Ninth – a time measuring unit in DNA genealogy is commonly a generation. A generation in the context of DNA genealogy is an event that happens four times a century. Numerically and time-wise, it is close to a generation in its common sense, but not necessarily equals to it. After all, a "common" generation cannot be defined precisely in years, and floats in its duration in real life. Hence, in this study 25 years for a generation it set by definition. This was a time span used for the calibration of mutation rates in this study. If someone prefers 30 years for generation, or any other figure, they can re-calibrate the mutation rates and as a result will get exactly the same number of years to a common ancestor. Hence, a number of years per generation in this case, if properly calibrated, is practically irrelevant.

Tenth - this statement is a new one and it was developed and explained in this work. Since the author has a background in chemical and enzyme kinetics, this new statement is rather obvious to him. The core of the statement is that the following two figures should be used for estimates of time span to a common ancestor: (1) a number of mutations in a set of haplotype with respect to the base haplotype, and (2) a fraction of the base haplotypes in the whole set, despite a number of mutations in the set.

Only if a number of generations to a common ancestor calculated using these two approaches is the same (within a reasonable error margin) it can be used as a time span to the common ancestor. If the two numbers are principally different (often by 2-3 times), the set of haplotypes should be dissected (for example, by constructing a tree of haplotypes and identifying separate branches of the tree), and the same two approaches should be applied to calculations.

Below is a brief example to illustrate this important principle. Let us consider two sets of 10 haplotypes in each:

14-16-24-10-11-12	14-16-24-10-11-12
14-16-24-10-11-12	14-16-24-10-11-12
14-16-24-10-11-12	14-16-24-10-11-12
14-16-24-10-11-12	14-16-24-10-11-12
14-16-24-10-11-12	14-16-24-10-11-12
14-16-24-10-11-12	14-16-24-10-11-12
14-17-24-10-11-12	14-16-25-9-11-13
15-16-24-10-11-12	14-16-25-10-12-13
14-15-24-10-11-12	14-17-23-10-10-13
15-17-24-10-11-12	16-16-24-10-11-12

The first six haplotypes in each set are base (ancestral) haplotypes. Other four are mutated base haplotypes. A number of mutations in the two sets with respect to the base haplotypes are 5 and 12, respectively. If to operate only with mutations, the apparent number of generations to a common ancestor in the sets is equal to 5/10/0.0096 = 52 generations and 12/10/0.0096 = 125 generations, respectively. However, in the both cases a ratio of base haplotypes gives us a number of generations equal to ln(10/5)/0.0096 = 53 generations (principles of calculations are described below). Hence, only the first set of haplotypes gave matching numbers of generations (52 and 53) and represents a “clean” set, having formally one common ancestor. The second set is “distorted”, it has apparently more than one common ancestor, and cannot be used for such simplified calculations of a number of generations to the ancestor.

Multiple examples of calculations with actual sets of haplotypes of various haplogroups are given below in this and subsequent paper.

Kinetics of haplogroup mutations and their calibration and verification

As it follows from the preceding section, in case of one common ancestor, hence, with a symmetrical haplotype tree, a transition of the base haplotype into mutated ones is described by the first-order kinetics

$\ln(B/A) = k*t$

where:
B = a total number of haplotypes in a set,
A = a number of unchanged (not mutated) base haplotypes in the set

k = an average mutation rate (frequency), which is, for example, 0.024 mutations per haplotype per generation for a 12-marker haplotype
t = a number of generations to the common (for the whole set of haplotypes) ancestor.

Let us examine this equation, using haplotypes for which a common ancestor is known and in some cases well documented.

Donald R1a1 family

The founding father for said family was John Lord of the Isles, who died in 1386, that is 26 generations ago if to take 25 years per generation. 68 haplotypes of his direct descendants are available (DNA-Project Clan-Donald, see ref.)

A list of 68 of 6-marker haplotypes contains 53 of the base haplotypes (in a "scientific" notation, that is in the order of DYS markers 19, 388, 390, 391, 392, 393):

15-12-25-11-11-13

All the 68 haplotypes contain 17 one-step mutations total from the above base haplotype. Using the above formula and assuming 26 generations from the common ancestor, we get

$\ln(68/53)/k = 26$, and

$k = 0.0096$ mutations per a 6-marker haplotype per generation.

If to consider all mutations in all the 68 haplotypes, we get

$17/68/k = 26$,

$k = 0.0096$ mutations per a 6-marker haplotype per generation, that is exactly the same value as above.

It suggests that the Donald haplotype tree is "clean", simple, symmetrical, that is not distorted one.

In fact, it gives an average mutation rate of 0.0016 mutations per marker per generation, which is close to a well-known number of 0.0020 mutations per marker per generation for 12-marker haplotypes. Overall, values of the mutation

rates vary in the literature, but generally they group around the value of 0.0019±0.0003 mutations per marker per generation, averaged per marker for 12-marker haplotypes. Hence, for a 6-marker haplotype the average mutation rate would be 0.0114±0.0018 per haplotype, that is within the error margin with the above value 0.0096.

Therefore, the value of 0.0096 mutations per haplotype per generation, obtained using two different approaches, fits the literature data. This means that a number of mutations in haplotypes and a fraction of base haplotypes in the whole set should be firmly connected. This gives us a principal criterion for a "rightness" of the set of haplotypes in terms of a single common ancestor and "rightness" of a number of calculated generations (years) to the common ancestor.

Let us verify these mutation rates using other sets of 6-marker haplotypes.

Bulgarian Gypsies (haplogroup H1)

Gypsies have arrived to Bulgaria – according to old records – in the Middle Ages. Haplotypes of Bulgarian Gypsies, or Roma, have been determined by testing 179 males from 12 local tribes (Zhivotovsky et al, 2004). All the haplotypes were pretty much similar and apparently originated from the same rather recent common ancestor. It seems that a very narrow circle of Gypsies, maybe a single tribe, had come to Bulgaria 500-700 years ago. The rest (if any) apparently did not pass through a genealogical bottleneck, and a single patriarch gave a survived offspring. It cannot be excluded that a few close relatives were the patriarchs.

The most numerous tribe "Rudari" had the following 6-marker base haplotype

15-12-22-10-11-12

which was represented by 62 identical haplotypes of the total amount of 67 haplotypes in the tribe. The same base haplotype was represented by 12 members of the tribe "Kalderash" of the total tested 13 individuals, 9 out of 26 members of the tribe "Lom", all four tested members of the tribe "Torgovzi" ("Traders"), 20 out of 29 from the tribe "Kalaidjii", 12 of 19 from the tribe "Musicians". Other haplotypes also contained very few mutations. It is obvious that the ancestral haplotype was rather "young", not 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)/0.0096 = 21 generations, that is 525 years to the common ancestor for all 179 members of all the 12 Gypsy tribes.

Considering mutations, this gives 34/179/0.0096 = 20 generations, that is 500 years.

It is practically the same time. Thus, the Gypsies show a pretty straightforward, uncomplicated DNA genealogy family tree, in which there is a rather firm fit between diminishing (with time) base haplotypes, on the one hand, and accumulation of mutations in the resulting haplotypes, on the other. Incidentally, these haplotypes belong to haplogroup H1, which is very typical for India, and beyond that country it is met mainly among the Gypsies. The base 6-marker haplotype for H1 in India is exactly the same as that shown above, however, its age there is much older, several thousand years.

Gypsy presence in Bulgaria was allegedly recorded some 700 years ago [cit. in (Zhivotovsky et al, 2004)], however, the records are not necessarily exact, the Gypsies could be of different tribes, and those who arrived first could have left descendants who did not survive. The haplotypes show that the lineage originated some 500-525 years ago.

One can see that the base haplotype/mutations dual procedure for calculations of a time span to a common ancestor works well with haplogroups R1a1 and H1, described above. There is no reason to believe that mutation rates are different for haplotypes in other haplogroups, however, I will consider in this paper a number of different haplogroups (for example, C2, N3, N3a, N3a1, R2, R1a1, F, H1, Q, R1b, R1b1c, J*, J1, J2, I1, I2, G and some other populations).

The Polynesians (haplogroup C2)

The Polynesians, such as the Maoris, Cook Islanders, Samoans, often have haplogroup C2. In a published study (Zhivotovsky et al) 36 haplotypes were determined in these three populations, and the base haplotype for all of them was as 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)/0.0096 = 26 generations to the common ancestor for all 36 Polynesians.

Considering mutations, this gives 10/36/0.0096 = 29 generations.

It is not a poor fit, taking into account a relatively small set of haplotypes. Incidentally, the so-called Polynesian expansion is referred to 650-700 to 800 and even to 1200 years ago [cit. in (Zhivotovsky et al)]. One can see that the 650-700 years BP shows a good fit to our data of 26-29 generations, that is 650-725 years BP; the 1200 years BP is apparently on some extreme side for the haplotypes presented here.

* * *

One can see, that so much different data series, such as those including haplotypes R1a1 (the Donald family), C2 (Polynesians), and H1 (the Gypsies), show a good fit in terms of average mutation rates in their haplotypes between a fraction of their residual base haplotypes and a number of mutations in their haplotypes. Their sets of haplotypes also show that their bearers have descended from a single common ancestor in each case.

Other examples of calculations of time spans to common ancestors are given in the second half of this paper. Before that, let us consider calibration of mutation rates in more detail.

Mutation rates and haplotype trees for short and extended haplotypes

It was shown above that the Donald family presents a “clean” set of haplotypes, at least in terms of 6-marker haplotypes (see above), and the dual procedure (considering a fraction of base haplotypes and total number of mutations) gives the same time span to a common ancestor. In this section we further examine 12-, 25-, 37- and 67-marker haplotypes and calculate the respective mutation rates for these extended haplotypes. Besides, we consider the respective haplotype trees, in order to have representative examples of “clean” trees, uncomplicated by overlapping ancestors of different lineages. This will help us further on in this and the subsequent paper in which we consider more complicated haplotype trees.

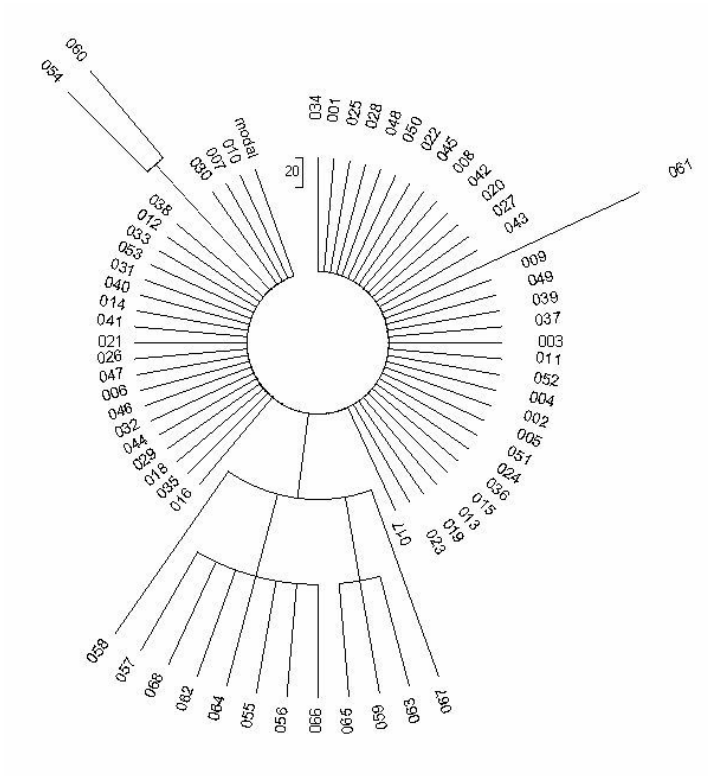

Figure 1. The 6-marker haplotype tree for the Clan Donald USA, haplogroup R1a. There are 68 haplotypes in the set. Source – see References

An example of an uncomplicated, single-common ancestor 6-marker haplotype tree is shown in Fig. 1. It contains 68 haplotypes. This and other trees in this paper were constructed using the program Phylip, the Phylogeny Inference Package (Felsenstein, 2005). A "comb" around the wheel (a "trunk") in this particular case consists of 53 identical, base haplotypes. For only 26 generations passed after the common ancestor, the base haplotypes are still in the majority, and a few mutated branches are situated not far away from the trunk itself. As it was shown above, the average mutation rate for 6-marker haplotypes is equal to 0.0096 mutation/haplotype/generation.

The base 12-marker haplotype of Donald family is as follows

13-25-15-11-11-14-12-12-10-14-11-31

An order of DYS markers here is as follow: 393, 390, 19, 391, 385a, 385b, 426, 388, 439,389-1, 392, 389-2. It is called "the FTDNA format".

In the “scientific” notation it is reduced to the same 6-marker haplotype shown earlier:

15-12-25-11-11-13

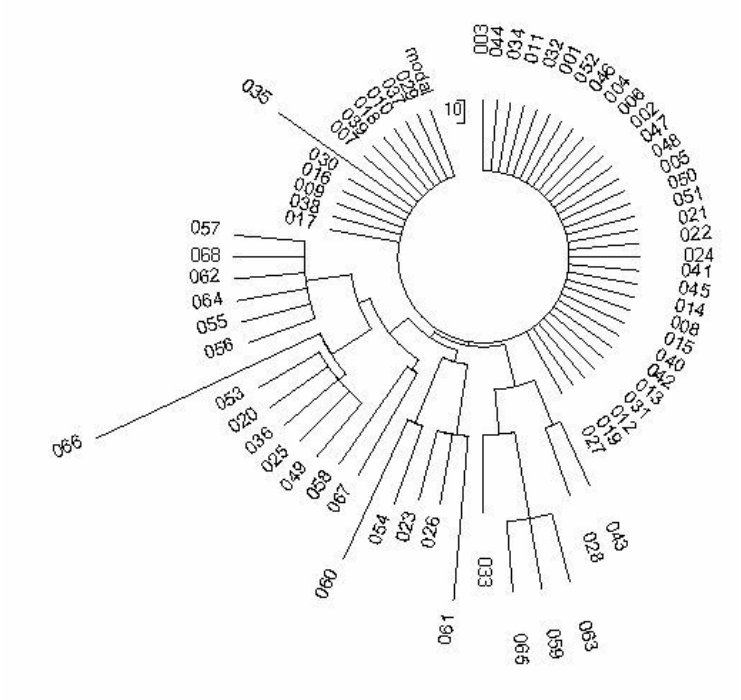

Figure 2. The 12-marker haplotype tree for the Clan Donald USA, haplogroup R1a. There are 68 haplotypes in the set. Source – see References

The respective 12-marker tree is shown in Fig. 2. It still contains as many as 42 base haplotypes, forming a comb around the tree in its upper part and on the right-hand side. The whole set of 12-marker haplotypes contains 44 mutations.

The fraction of base haplotypes gives ln(68/42)/0.024 = 20 generations.

The mutations give 44/68/0.024 = 27 generations.

The second number is reasonably close to the 26 generations established earlier. The first number deviates from the 26 generations, and probably reflects an excessive number of base haplotypes in this particular case. It will be shown a bit

later that an average mutation rate of 0.024 mutation/haplotype/generation is indeed an appropriate one for 12-marker haplotypes.

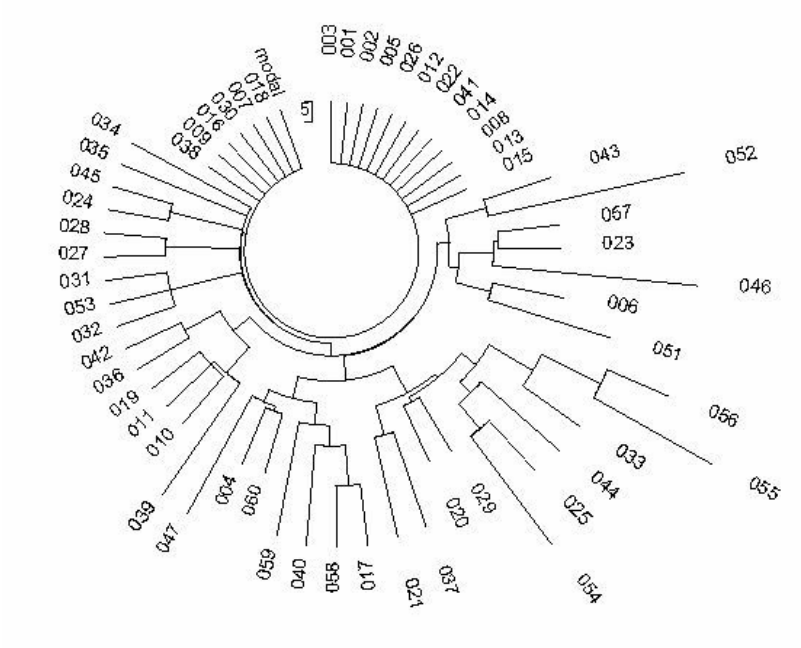

Figure 3. The 25-marker haplotype tree for the Clan Donald USA, haplogroup R1a. There are 60 haplotypes in the set. Source – see References

Fig. 3 shows a 25-marker haplotype tree for the Donald family. If has a smaller number of haplotypes since not everyone has conducted the respective DNA testing. The base 25-marker haplotype is as follows:

13-25-15-11-11-14-12-12-10-14-11-31-16-8-10-11-11-23-14-20-31-12-15-15-16

The set contains 18 base haplotypes and 69 mutations off the base haplotype in all 60 haplotypes. It gives for a fraction of base haplotypes ln(60/18)/0.046 = 26 generations, and for mutations 69/60/0.046 = 25 generations. The values are practically the same.

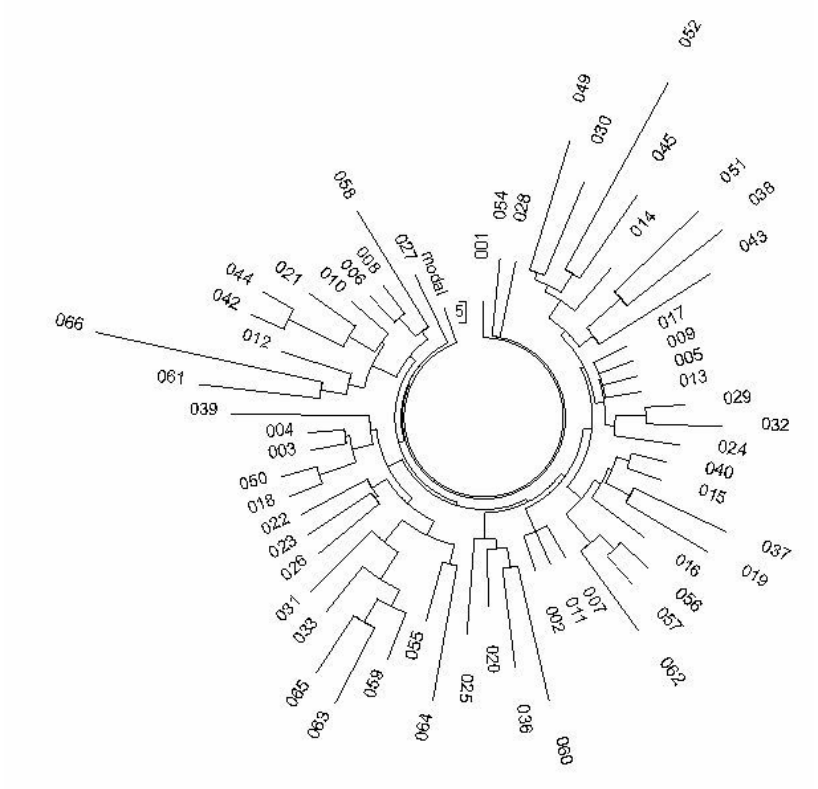

Figure 4. The 37-marker haplotype tree for the Clan Donald USA, haplogroup R1a. There are 59 haplotypes in the updated set. Source – see References

The 37-marker haplotype tree for the Donald family is shown in Fig. 4. It contains 59 haplotypes. The base haplotype is

13-25-15-11-11-14-12-12-10-14-11-31-16-8-10-11-11-23-14-20-31-12-15-15-16-11-12-19-21-17-16-17-18-34-39-12-11

It represents only one haplotype in the whole set of haplotypes. Clearly, it could not have been determined "visually", it was deduced with a computer. These 59 haplotypes contain 178 mutations with respect to the base haplotype, however, the one-fifth of all these mutations, namely 36, were provided by an apparently anomaly 35th marker, DYS#CDYb. Such a pattern for CDYb does not happen often in other haplotypes. Therefore, I have calculated two mutation rates for a 37-marker haplotype, considering an unusually high mutation rate of CDYb, and without it. It gives us 0.12 and 0.09 mutations per 37-marker haplotype,

respectively. Applying the both mutation rates to the 59 Donald 37-marker haplotypes, we obtain: 178/59/0.12 = 25 generations, and 142/59/0.09 = 27 generations. It is a good fit in both cases. Based on multiple calculations of hundreds – literally – haplotypes and their sets, I have chosen 0.09 mutations per 37-marker haplotype per generation as a better-fit value.

Finally, let us take a look at the 67-marker haplotypes (Fig. 5). There are only 26 of them on the tree. The tree is rather symmetrical, and contains two principal branches, one of them in its turn splits into four branches, each of them having its own ancestor. All 26 haplotypes have a deduced base haplotype

13-25-15-11-11-14-12-12-10-14-11-31-16-8-10-11-11-23-14-20-31-12-15-15-16-11-12-19-21-17-16-17-18-34-38-12-11-11-8-17-17-8-12-10-8-11-10-12-22-22-15-11-12-12-13-8-14-23-21-12-12-11-13-11-11-12-12

and all the 26 haplotypes have 98 mutations from this base haplotype. It corresponds to an average mutation rate for 67-marker haplotypes equal to 0.145 mutations per haplotype per generation.

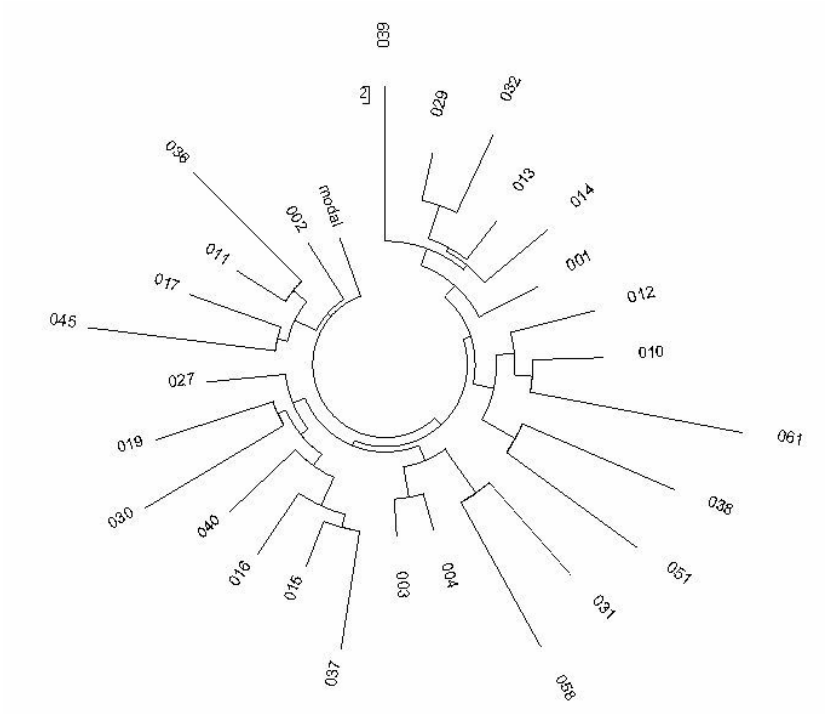

Figure 5. The 67-marker haplotype tree for the Clan Donald USA, haplogroup R1a. There are 26 haplotypes in the set. Source – see References

Overall, the following list compares (a) a haplotype length (a number of markers), (b) the respective average mutation rate per haplotype per generation, and (c) the respective mutation rate per marker in a haplotype of the given length per generation.

Haplotype	Average mutation rate	
	Per haplotype	Per marker
6-marker	0.0096	0.0016
9-marker	0.025	0.0028
10-marker	0.026*	0.0026
10-marker	0.025**	0.0025
12-marker	0.024	0.0020
25-marker	0.046	0.0018
37-marker	0.090	0.0025
67-marker	0.145	0.0022

Here 12- and higher- marker haplotypes are those with their marker sequences in accord with the FNDNA and YSearch notation.
-- 6-marker haplotypes are those of the "scientific" notation, that is in the order of DYS## 19, 388, 390, 391, 392, 393 (though the order does not effect outcomes of calculations)
-- 9-marker ones DYS##19, 389-1, 389-2, 390, 391, 392, 393, 385a, and 385b, or DYS##19, 388, 389-1, 389-2, 390, 391, 392, 393, and 439,
-- 10-marker haplotypes are those with DYS##19, 388 (or 439), 389-1, 389-2, 390, 391, 392, 393, 385a, and 385b (*), or DYS##19, 389-1, 389-2, 390, 391, 392, 393, 385a, 385b, and 439 (*), or DYS##19, 388, 389-1, 389-2, 390, 391, 392, 393, 426, and 439 (**).

I will commonly represent haplotypes with more than 6-markers in the 12-marker notation, replacing missing alleles with X, Y, Z symbols, is in the following order of DYS: 393, 390, 19, 391, 385a, 385b, 426, 388, 439, 389-1, 392, 389-2.

These values are employed in this and the subsequent paper.

Some comparative values for average mutation rates per marker given in the most recent studies are as follows (in mutation/marker/generation) (Chandler, 2006; Athey, 2007):

12-marker – 0.0019±0.0003
25-marker – 0.0028±0.0004
37-marker – 0.0049±0.0007

12-marker – 0.0039; 0.0024
25-marker – 0.0044; 0.0027
37-marker – 0.0058; 0.0041±0.0002

These values of the average mutation rates are either within error margins with those obtained in this paper, or would not fit the calibration presented above.

Mutation rates and back mutations in calculations of a time span to a common ancestor

If we assume that all mutations in any direction, back and forth, occur with the same probability, that is with the same mutation rates, which is the most likely, than the following flowchart can be considered

D ↔ B ↔ A ↔ C ↔ E

where A is a base, or an ancestor marker, and all other are its mutated forms.

With back mutations one sees less of overall mutations compared with their total number of back and forth events. The larger the time span passed from the ancestor, the higher the discrepancy, and the larger the actual time span compared to that calculated from a number of mutations. For 12-marker haplotypes, for example, if mutations would show 8,000 years from the common ancestor, in reality it corresponds to 11,800 years. In another words, the average effective mutation rate constant would be not 0.0020 mutations per marker per generation, but about 0.0014 mutation/marker/generation instead. Let us consider how significant would be a correction factor for different situations, and which time span should be considered as a threshold for a necessity of such corrections.

A kinetic equation for the above flowchart can be solved in a numerical form. Some programs can be utilized in such calculations (a program kindly provided by Mikhail Semenov and arranged in a tabulated format by Vadim Urasin was used in this work).

Let us compare average mutation rates and number of mutations with and without back mutation effects for several time span values to a common ancestor. We

will consider three models for calculations of a number of mutations, based on a number of generations passed from an ancestor. For these exemplary calculations we will take an average mutation rate equal to 0.002 mutation/marker/generation, that is a mutation occurs – on average – once per 500 generations, or 12,500 years, with 25 years per generation. This mutation rate was validated for 12-marker haplotypes, as shown above.

The following models will be considered:

(a) A linear model. This simple model is applicable for a relatively small number of generations, or a small number of mutations in a base haplotype:

$t = n/N/k$

where:
t = time span to a common ancestor, in generations,
n = a number of mutations in all N haplotypes of the set,
k = the average mutation rate, in mutation/marker/generation

(b) A probability model, which considers a certain likelihood for a mutation in one or an other direction. When a probability of a mutation in either direction in one generation equals to 0.002, the probability that there would be no mutation in one generation equals 0.998, and the probability that there would be no mutations after x generations equals to 0.998^x.

(c) A back mutation model. A back mutation probability is included into the model. Numerical data for generations and the respective number of mutations were calculated using a computer program generated in Excel and employed mutation rates of 0.002 mutation/marker/generation in either direction.

Several examples for different time spans are given in the Table.

Table. Conversion of mutations in 12-marker haplotypes to a number of generations for the three models, described in the text.

An average number of mutations per marker in 12-marker haplotypes*	A number of generations (years) to a common ancestor			The intrinsic mutation rate (in mutation/marker /generation) at different time spans to a common ancestor
	A linear model	A probability model (without back mutations)	A back mutation model	
0.020	10 (250)	10 (250)	10 (250)	0.0020
0.052	26** (650)	27 (650)	27 (675)	0.0020
0.080	40 (1,000)	42 (1.050)	42 (1,050)	0.0019
0.12	60 (1,500)	64 (1,600)	64 (1,600)	0.0019
0.20	100 (2,500)	111 (2,775)	111 (2,775)	0.0018
0.40	200 (5,000)	255 (6,375)	250 (6,250)	0.0016
0.80	400 (10,000)	804 (20,100)	675 (16,875)	0.0012
1.16	580 (14,500)	>4,000 (>100,000)	1,700 (42,500)	0.0007
1.20	600 (15,000)	>5,000 (>125,000)	2,200 (55,000)	0.00055

* In 6-, 25-, 37- and 67-marker haplotypes introducing of a correction factor is needed (see the mutation rates in the preceding section)
** The number of generations employed for the calibration of the mutation rates in the preceding section

The Table shows that even after 1,000 years of mutational evolution there is a slight deviation of the linear model from more realistic probability model and the back mutation model. Until about 0.2 mutations per marker, that is 100-110 generations to a common ancestor, the linear model can be used and gives up to a 10% underestimation regarding a number of generations to a common ancestor.

At about 6,000 years to a common ancestor the linear model yields a significant deviation, up to 25% in the error margin, and hardly can be used. At this time span to a common ancestor the probability model begins to show a deviation, since it does not consider back mutations.

After about 10,000 years to a common ancestor the probability model starts to overestimate the time span progressively, reaching 20% deviation at 10,000 years BP, and soon launches into quite unrealistic values of generations. The linear model does not work at those time spans at all, since it underestimates a number of generations by more than twice, and even up to three or four times.

The intrinsic mutation rate decreases accordingly, as it is shown in the Table. However, in order to reach the value of 0.0007 mutations per marker per generation, a common ancestor should have been living some 40 thousand years BP. Unfortunately, this coefficient (Zhivotovsky et al, 2004) is mistakenly used in a number of publications with practically any time span considered, often at 1,000-5,000 years to a common ancestor [a specific case is considered below, in the section "Russians (haplogroup N31a)"]. Hence, time spans to a common ancestor are often overestimated in the literature by about 3 to 4 times.

As one can see, effects of back mutations on time spans often considered in DNA genealogy have a fundamental significance, and might practically double time spans to a common ancestor compared to the linear model of calculations.

Multiple effects of back mutations in particular situations are given below.

More examples of calculations of time spans to common ancestors

6-marker haplotypes

6-marker haplotypes are a poor choice for calculations of a time span to a common ancestor. However, in some cases there is no any other choice. In some cases, however, 6-marker haplotypes provide pretty good estimates of time spans, as can be proven using more extended haplotypes (see this and the subsequent paper in this issue). It happens when a number of mutations in the first 6-marker panel statistically fit to a number of mutations in next panels of markers, such as 12-, 25-, 37-marker panels, and so on. However, it is unpredictable. They might fit, they might show non-proportionally less mutations, or they can show non-proportionally more mutations. That is why it is always a good idea to use more extended haplotypes for a verification of results

obtained with 6-marker haplotypes. Or better not to use 6-marker haplotypes at all, when more extended haplotypes are available.

Examples with 6-marker haplotypes shown in the following sections sometimes give "reasonable" estimates, though they should be verified with more extended haplotypes, when possible.

Indians and Pakistanis (haplogroup R2)

In a recent paper (Sengupta et al, 2006) a set of 68 haplotypes of Indians and 13 those of Pakistanis having haplogroup R2 was listed. 21 of them had the identical, base haplotype

14-12-23-10-10-14

All 81 haplotypes had 105 mutations in comparison with the base haplotype. Therefore, based on a fraction of base haplotypes in the total set, we obtain ln(81/21)/0.0096 = 141 generations to a common ancestor, and based on mutations we obtain 105/81/0.0096 = 135 generations (without a correction for back mutations). This is a good fit, and points at a single common ancestor for this set of haplotypes. With the correction, it gives 160±6 generations, that is 4,000±150 years to a common ancestor.

Principles of correction for back mutations are considered in detail earlier in this paper. Typically, a contribution of back mutations, which "reduce" calculated estimates of time spans to common ancestors, becomes measurable after about 25-30 generations, quite noticeable after about 100 generations, and makes time estimate plain wrong after about 300-400 generations.

Chenchu (haplogroup R1a1)

The Chenchus, an australoid tribal group from southern India, bear R1a1 haplogroup in about 27% of their members (in 11 of 41 individuals tested) (Kivisild et al, 2003). It is tempting to associate it with the Aryan influx into India which allegedly occurred some 3,400 years BP, or maybe earlier considering possibly several waves of the Aryan influx to the Hindustan. However, incorrect calculations of time spans to a common ancestor of R1a1 in India (Kivisild et al, 2003; Sengupta et al, 2006; Sahoo et al, 2006) 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 a good statistics, however, they can allow a reasonable estimate of a time span to a common ancestor for these 11 individuals. There is a simple logic behind it – if these haplotypes are more or less identical, a common ancestor lived within a thousand years BP; 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)/0.0096 = 178 generations, or 4,450 years to a common ancestor, which, corrected to back mutations, would result in 216 generations, or 5,400 years to a common ancestor. If all eleven 6-marker haplotypes are mutated, it means that a common ancestor lived earlier than 5-6 thousand years BP. Hence, even with such a poor set of haplotypes one can obtain useful and meaningful information.

11 Chenchu haplotypes have seven (!) base 6-marker haplotypes:

16-12-24-11-11-13

They are only one mutation away compared with the base Russian (Eastern Slavs) haplotypes, described below:

16-12-25-11-11-13

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

For the Chenchus, it gives ln(11/7)/0.0096 = 47 generations to a common ancestor. Those 11 haplotypes contain 9 mutations from the base haplotypes, that gives 9/11/0.0096 = 85 generations. Clearly, the haplotype set reflects several (minimum two) common ancestors. Indeed, a haplotype tree (to be published, in a paper concerning R1a1 haplotypes in India) shows two distinct branches, one recent, which includes 8 haplotypes, seven of them are the base haplotypes, and only one mutation, and an older one, containing three haplotypes, having three mutations from their base haplotype

15-12-25-10-11-13

The recent branch results in ln(8/7)/0.0096 = 14 generations, and 1/8/0.0096 = 13 generations, from the residual base haplotypes and a number of mutations, respectively. It shows a good fit between the two estimates, and confirms that a single common ancestor for 8 individuals of 11 lived only about 350 years BP, in the 17^{th} century. The old branch of haplotypes points at a common ancestor who lived 3/3/0.0096 = 104 generations BP. With a correction for back mutations, it

gives 116 generations, that is 2,900 years to a common ancestor. Considering that the Aryan (R1a1) wave to northern India took place about 3,400 years BP, it is quite plausible to refer the appearance of R1a1 in the Chencha to the Aryans.

Regarding an origin of the influx of Chenchu R1a1 haplotypes in the 17th century, I would just quote a passage from (Kivisild et al, 2003): "Chenchus were first described as shy hunter-gatherers by the Mohammedan army in 1694".

Koyas (haplotypes F and H1)

A set of haplotypes of the Koyas, another southern Indian tribe, was given in (Kivisild et al, 2003). The Koyas show only a minor amount of haplotypes of R1a1 haplogroup (in fact, only one among 41 tested), and bear more traditional Indian haplogroups such as F and H1.

Among 11 haplotypes of haplogroup F, two identical of them can be considered as the base haplotypes

16-13-21-11-11-14

This would give ln(11/2)/0.0096 = 178±40 generations to a common ancestor. The standard deviation was introduced here since there are only two base haplotypes in the set, and such a small number of base haplotypes is sensitive to deviations; three base haplotypes which could have occurred in the set by chance, would have given 135 generations to a common ancestor.

Since these 11 haplotypes have 15 mutations from the base haplotype, it would give 15/11/0.0096 = 142±10 generations to a common ancestor. These two figures are about the same within error margin. This indicates that a single common ancestor for all 11 Koyas individuals with haplogroup F lived (with a correction for back mutations) 4,800±700 years BP.

There are 25 haplotypes of Koyas of haplogroup H1 published in the same study (Kivisild et al, 2003). There are as many as 11 base haplotypes in the 25-haplotype set:

15-12-22-10-11-12

The whole set contains 22 mutations from the above base haplotype. Based on a fraction of residual base haplotypes, we obtain ln(25/11)/0.0096 = 86±10 generations to a common ancestor. Based on a number of mutations, we get

22/25/0.0096 = 92±4 generations BP. These two figures are about the same within error margin. This indicates that a single common ancestor for all 25 Koya individuals with haplogroup H1 lived (with a correction for back mutations) 2,400±200 years BP.

Koragas (haplotypes H1)

A set of haplotypes of the Koragas, another southern Indian tribe, was given in (Cordaux et al, 2004). The paper presents only 5-marker haplotypes, given below in the order of DYS##389-1, 389-2, 390, 391, 393.

A calibration performed using the Donald Clan haplotypes as described above has resulted in mutation rate of 0.023 mutations per marker per generation.

The tested Koragas were almost exclusively bearers of H1 haplogroup, and their 28 haplotypes included as many as 10 identical, base haplotypes:

11-27-22-10-12

These 28 haplotypes contained 27 mutations from the above base haplotype.

Clearly, 10 identical, base haplotypes out of 28 haplotypes total is itself a good indication that the tested population is not too old, in terms of their common ancestor. A fraction of the residual base haplotypes give ln(28/10)/0.023 = 45 generations from a common ancestor, while mutations give 27/28/0.023 = 42 generations. This is a pretty good fit for a series of haplotypes so short in length. It indicates that the ancestor was very likely a single one, and that the haplotype set does not have admixtures of other common ancestors.

A correction for back mutations provides a time span to a common ancestor which is 91 generations, that is about 2,300 years BP.

American Indians (haplogroup Q)

117 haplotypes of Native American Indians, haplogroup Q-M3, were published in (Bortolini et al, 2003). They clearly were derived from different common ancestors, since 31 identical base haplotypes out of 117 haplotypes would give 138 generations to a common ancestor, though 273 mutations would give 243 generations, and with corrections for back mutations it would result in 321 generations.

Indeed, a haplotype tree (will be published elsewhere) showed at least six common ancestors for the 117 individuals. Four of them turned out to be quite recent 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 haplotypes derived from an ancient ancestor. His base (ancestral) haplotype was as follows:

13-13-24-10-14-14

The 11-haplotype branch contains 32 mutations, that give 436 generations, or 10,900 years to a common ancestor.

Armenians (haplogroup R1b)

A set of 238 haplotypes was provided in (Weale et al, 2001). It included haplotypes from six geographical regions in Armenia, Karabakh, Iran, etc. A consideration of those haplotypes have shown that all of them descended from both old and rather recent ancestors. In other words, all six regions represent a heterogeneous mix of populations, and for all of them calculations based on residual base haplotypes, on the one hand, and mutations, on the other, give different numbers of generations to common ancestors. In all six regions, though, a base haplotype is the same:

14-12-24-11-13-12

It has one mutational deviation from the Atlantic Modal Haplotype in its 6-marker format:

14-12-14-11-13-13

For example, for Iranian region a number of generations calculated from a fraction of residual base haplotypes and a number of mutations are 157 and 179, respectively. For the Karabakh region they are 140 and 156 generations. For the Eastern Turkey they are 168 and 203 generations. For the northern Armenia they

are 160 and 206 generations. For the Southern Armenia they are 107 and 138 generations. The highest, albeit distorted number of generations were from the Ararat region, which gives 168 and 260 generations, respectively, that is about 4,200 and 6,500 years to an "effective" common ancestor. Typically, the first figure is inclined to a more recent ancestor, and the second figure – to a more ancient ancestor. However, both of them are distorted ones, and a degree of distortion depends on a number of factors, including a relative number of descendants and a number of generations to each of the common ancestors.

In order to resolve the issue, a haplotype tree has been constructed (it will be shown in future publications in this Proceedings). The above suggestion turns out to be correct. For example, a 52-haplotype combined "Ararat" and "Northern Armenian" series of haplotypes split into five branches. One branch is obviously very ancient, and gave 456 generations (!) to a common ancestor, that is about 11,400 years BP. Another branch gave 217 generations, that is 5,400 years to a common ancestor. A third branch gave 200 generations to a common ancestor, that is 5,000 years BP. Yet another branch gave 150 generations to a common ancestor, that is 3,750 years BP. The last branch could not be resolved since it contained only base haplotypes (11 of them), apparently from a very recent ancestor.

Obviously, the Armenian R1b common ancestors are significantly older compared to the West European ones, which lived between 3,100 and 3,900 years BP, as it will be shown below with more extended haplotypes. Another ancient common R1b AMH ancestor, of 5,000±200 years BP was identified among the Jews (see the subsequent paper in this issue), obviously, also geographically far away from the European continent. It seems indeed that the ancient R1b population in Europe was wiped out some 4,000 years BP, and the continent was re-populated again with R1b bearers, probably by a new influx from Asia, Armenia, or elsewhere. That is why the Armenian and the Jewish R1b common ancestors are noticeably older compared with West European ones.

Armenians (haplogroup R1a1)

A set of 37 haplotypes was provided in the same work as that listed above (Weale et al, 2001). It included haplotypes from five geographical regions in Armenia, Karabakh, Iran, etc. Again, as it was with R1b haplotypes, the set shows more than one common ancestor. Indeed, a fraction of residual base haplotypes for the Armenian R1a1 haplotypes gave 64 generations, while a number of mutations in the whole 37-haplotype set gave 96 generations to a common ancestor. Typically, in case of two common ancestors, the first figure is an overestimation (in reality a

common ancestor lived much more recently), which the second figure is an underestimation (the common ancestor in reality lived much earlier).

A more close consideration of the set showed that haplotypes from three of said regions contained a high amount of a quite recent base haplotype

15-12-25-11-11-13

which could not even be dated in terms of its common ancestor. It is too recent in order to be dated. There are 20 of those identical (base) haplotypes in the whole 37-haplotype set. Only Iranian and Eastern Turkey haplotype sets do not contain them.

A haplotype tree (will be shown in future publications) confirms that the set indeed contains two distinct subgroups. 20 base haplotypes form one of them, and 17 mutated haplotypes form another branch.

The recent (base) haplotype was removed from the set, and remaining 17 haplotypes were considered. Their base haplotype was one mutation away from the above base haplotype

16-12-25-11-11-13

It is a "classical" old Slavic haplotype (see below), which in a 12-marker format looks as follows:

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

A common ancestor of the above Slavic (Eastern Slav) haplotypes lived 185 generations, that is 4,625 years BP, 27th century BC (see below). The Armenian haplotypes gave 164 generations, or 4,100 years to a common ancestor.

It is generally understandable why this ancient Slavic ancestral R1a1 haplotype presents in all Armenian populations, while the quite recent base haplotype presents only in the former Soviet Armenia haplotype sets, but not in Iranian and Turkish territories.

Jews (haplogroups J* and J1)

A set of 194 haplotypes of Jewish 6-marker haplotypes of haplogroup J* was published (Behar et al, 2003), 91 of them were identical to each other:

14-16-23-10-11-12

This is so-called the Cohen Modal Haplotype (CMH). The whole set contains 263 mutations from the above base (presumably ancestral) haplotype. A fraction of the residual base haplotype gives ln(194/91)/0.0096 = 79 generations to a common ancestor, that is about 2,000 years BP, while a number of mutations give 263/194/0.0096 = 141 generations, that is about 3,500 years BP. The discrepancy indicates, that there is more than one common ancestor for these 194 individuals. The haplotype set is not "clean", not a uniform one.

Another set of 90 Jewish 6-marker haplotypes of haplogroup J1 was extracted from YSearch database (see the subsequent paper in this issue). It contained 41 base haplotypes, which are exactly the same CMH as shown above. The whole set contains 123 mutations from the CMH. A fraction of the base haplotype gives ln(90/41)/0.0096 = 82 generations to a common ancestor, that is about 2,000 years BP, while a number of mutations give 123/90/0.0096 = 142 generations, that is about 3,500 years BP. These are practically the same figures as those obtained from a "scientific set" (Behar et al, 2003), though the second series of haplotypes was taken from a "commercial" database. The discrepancy again indicates that there is more than one common ancestor for these 90 individuals.

A detailed study has shown that the Jewish J* and J1 haplotypes include at least eight lineages, each of them derived from a separate common ancestor (see the subsequent paper in this issue). These ancestors lived 11,100±600, 6,200±400, 5,100±500, 3,500±200, 1,800±200, 775±50, 1,650±350 and 1,350±100 years ago. The last two lineages are of the "Cohen Modal Haplotypes", which happened to be from two separate common ancestors, both of them lived in the first millennium AD. Some of the ancient lineages descended from common ancestors, who lived 13,400±2,400 and 9,000±1,000 years BP.

Jews (haplogroup J2)

A set of 88 haplotypes of Jewish 6-marker haplotypes of haplogroup J2 was published (Behar et al, 2004), 25 of them were identical to each other:

14-16-23-10-11-12

Again, it is so-called the Cohen Modal Haplotype (CMH), the same in the 6-marker format as one in haplogroup J1. However, now it is in haplogroup J2. The whole set contains 160 mutations from the above base (presumably ancestral)

haplotype. A fraction of the residual base haplotype gives ln(88/25)/0.0096 = 131 generations to a common ancestor, that is about 3,300 years BP, while a number of mutations give 160/88/0.0096 = 189 generations, that is about 4,700 years BP. Again we have a discrepancy, which indicates, that there is more than one common ancestor for these 88 individuals. Again, the haplotype set is not "clean" and not a uniform one.

Another set of 75 Jewish 6-marker haplotypes of haplogroup J2 was taken from YSearch database (see the subsequent paper in this issue). It contained 15 base haplotypes, of exactly the same CMH as shown above. The whole set contains 135 mutations from the CMH. A fraction of the base haplotype gives ln(75/15)/0.0096 = 168 generations to a common ancestor, that is about 4,200 years BP, while a number of mutations give 135/75/0.0096 = 188 generations, that is about 4,700 years BP. Despite some similarity between these figures resulted from the "scientific" and the "commercial" sets, these data are plain wrong. They do not refer to a single common ancestor at all.

A detailed study has shown that the Jewish J2 haplotypes descended from at least five different common ancestors (see the subsequent paper in this issue). Ancestors, descendants of which are presented the most haplotypes in the 75-haplotype set, lived 6,100±1,000, 900±50, and 1,200±200 years BP.

These examples show again, that one cannot calculate a time span to a common ancestor from a haplotype set based on mutations only, without applying a clearly defined criteria, such as one suggested and illustrated here, that the set is "clean" and "uniform", and is derived from a single ancestor.

Jews (haplogroup R1a1)

There is a set of Jewish R1a1 haplotypes available in the literature (Behar et al, 2003), which consists of 42 haplotypes, 25 of them are identical to each other (base haplotypes):

16-12-25-10-11-13

It is of interest, that the Jewish base (ancestral) R1a1 haplotype is different from other R1a1 base haplotypes, considered in this study:

Mac-Donald, Swedish and Armenian R1a1 base haplotypes -

15-12-25-11-11-13

Russian (Eastern Slavs) and Indian R1a1 base haplotypes –

16-12-25-11-11-13

and the Chenchu R1a1 base haplotype –

16-12-24-11-11-13

All of them form a pathway for R1a1 base haplotypes, each of them differs by just one mutation from the next one in chain, such as

West European, Armenian ←Slavic→ Indian → Chenchu
↓
Jewish

that is

15-12-25-11-11-13 ←16-12-25-11-11-13→ 16-12-24-11-11-13
↓
16-12-25-10-11-13

Coming back to the Jewish R1a1 haplotypes, their whole 42-haplotype set contains 25 base (ancestral) haplotypes and 22 mutations with respect to said base haplotype. A fraction of the base haplotype gives ln(42/25)/0.0096 = 54 generations to a common ancestor, and the number of mutations give 22/42/0.0096 = 55 generations (54.6, to be absurdly precise). This indicates that the Jews of R1a1 haplogroup descended from a single common ancestor who lived about 1,350 years BP, in the 7th century AD, during the Diaspora.

In fact, 12-marker haplotypes of the R1a1 Jews, extracted from YSearch database (there are no available data on 12-marker Jewish haplotypes in the literature), form a set of 44 haplotypes, 14 of them are base haplotypes

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

and the whole set contains 57 mutations. This gives ln(44/14)/0.024 = 48 generations, and 57/44/0.024 = 54 generations to a common ancestor. Hence, the 6- and 12-marker haplotypes give pretty compatible figures, on average 53±3 generations, or 1,325±75 years to a common ancestor of the Jews in haplogroup

R1a1 in the said haplogroup set. It is described in more detail in the subsequent paper in this issue.

It is remarkable that the Jewish R1a1 base haplotype differs by only one mutation from both the 6-marker Russian (Eastern Slavic) base haplotype and the 12-marker haplotype as well (in the Slavic base haplotype the mutated allele is shown in bold):

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

For the record, the 12-marker base haplotype of the Donald family

13-25-15-11-11-14-12-12-10-14-11-31

differs from the Slavic R1a1 base haplotype by three mutations, and from the Jewish R1a1 base haplotype by four mutations.

* * *

As a reader can see, the approach offered in this study, works well on 6-marker haplotypes, and "well" means that the data obtained look reasonable, not counter-intuitive. Of course, all these data need to be confirmed with more extended haplotypes or verified independently.

As it was said above, 6-marker haplotypes is often a poor choice for calculations of a time span to a common ancestor. In the above cases it was seemingly the only choice. Other available sets of haplotypes can include 9-, 10-, 12-, 25-, 37-, 67-marker haplotypes, and whatever in between and beyond. Each set of haplotypes of a certain length/composition, shown below, was uniform in terms of length of the haplotypes throughout the set, it was calibrated using the approach described above, and the obtained respective mutation rates are given above.

Here I will continue with examples with more extended haplotypes, in order to show an overall picture and give a reader a feeling of different time spans to common ancestors in different haplogroups, obtained with haplotypes of various origins. When a time span exceeds (or approaches) 40 generations, that is about 1,000 years to a common ancestor, a correction for back mutation is introduced, as described above.

Descriptions of population haplotypes, given below, and time spans to common ancestors of quoted haplotype sets will be discussed in more detail in studies

published in subsequent issues of the Proceedings. Here it is just a set of examples of calculations.

Baltic Sea region populations (haplogroup N3a)

A set of 249 haplotypes of N3a haplogroup with a base haplotype (in 6- and 12-marker formats)

14-12-24-11-14-14

14-24-14-11-11-13-11-12-10-14-14-30

was collected in YSearch database (Vladimir Volkov, personal communication), and had identical 51 of the 12-marker base haplotypes and 558 mutations in the whole set of haplotypes with respect to the above haplotype. The two calculations – based on base haplotypes and on mutations, gave 66 and 93 generations respectively to a common ancestor. Clearly, these 249 haplotypes are derived from at least two different ancestors, since these two figures did not match. Indeed, the respective haplotype tree revealed a distinct separate branch consisting of 126 haplotypes with the following number of base haplotypes and mutations with respect to the base haplotypes:

6-marker haplotype	77 base	54 mutations
12-marker haplotype	48 base	134 mutations

For the 6-marker haplotypes
ln(126/77)/0.0096 = 51 generations, 54/126/0.0096 = 45 generations

For the 12-marker haplotypes
ln(126/48/0.024 = 40 generations, 134/126/0.024 = 44 generations

An average of all the four values gives 45±5 generations to the common ancestor, that is between 1,000 and 1,250 years BP, 8-10th century AD. It is interesting that far most of those 126 individuals were Finnish (116 of the 126). Two were from Russia and Germany each, and six from Sweden. 123 haplotypes from other branches were of individuals mainly from Poland, Lithuania, Belarus, Germany and some other countries, forming at least nine different branches, with common ancestors up to 110 generations BP.

Russians (haplogroup N3a1)

A set of 37 twelve-marker haplotypes was listed in (Derenko et al., 2007). However, DYS##437 and 438 were from two different panels of the 37-marker standard FTDNA notation, hence, only 10-marker haplotypes were analyzed in this study. Consideration of haplotypes have shown that they certainly derived from more than one common ancestor. When the main criterion was applied, namely comparing a number of generations obtained from residual base haplotypes and that obtained from a number of mutations, they gave 64 and 112 generations, respectively (using all 7 base haplotypes and 97 mutations in the whole set). Obviously, there were too many base haplotypes for that too many mutations. The base haplotypes were clearly derived from a rather recent ancestor.

A haplotype tree confirmed that conclusion (the tree will be considered in detail elsewhere). There were at least four principal branches in the tree. The first one was obviously "young", with the following base haplotype

14-23-14-11-11-13-X-Y-10-14-14-30

It deviated by just one mutation (in DYS#390) from the "Baltic N3a" base haplotype shown above. This 9-haplotype branch contains all seven base haplotypes, which corresponds to $\ln(9/7)/0.026 = 10$ generations to a common ancestor. Since the branch contains only 2 mutations from the above base haplotype, it gives $2/9/0.026 = 9$ generations, or about 225 years to a common ancestor.

Another 9-haplotype branch is rather distant from the above-described branch, and contains 5 base haplotypes

14-23-15-11-11-14-X-Y-10-13-14-29

and only 4 mutations. This gives $\ln(9/5)/0.026 = 23\pm(8 \text{ to } 5)$ generations and $4/9/0.026 = 17\pm4$ generations from a common ancestor. Since statistics here is poor, the figures can be considered as a fair fit. The deviations above are shown as it would have been a ±1 change in base haplotypes or mutations. Hence, the number of generations are within the error margin. It seems that the common ancestor of those 9 individuals lived just around 500 years BP.

Since the two above base haplotypes differ by as many as four mutations in their 9-marker format, it translates to 206 generations between them, or about 5,200

years. Taking into account the age of their separate common ancestors, their common ancestor lived about 3,000 years BP.

The "oldest" branch on the tree is one containing 8 haplotypes, which in turn have 19 mutations. It gives 100 generations, that is 2,500 years from a common ancestor.

It should be noted that the authors (Derenko et al, 2007), using "evolutionary coefficient" of 0.00069 mutations per generation (Zhivotovsky et al, 2004) obtained an "evolutionary time estimate" for the same Russian N3a1 haplotypes, equal to 8.21±2.01 thousand years. These figures have, of course, quite a different meaning compared to the estimation of a time span to a common ancestor for each lineage, described in this study, let alone the notation "8.21±2.01" which is more than questionable (NOTE: 8±2 would be more appropriate in that case) .

The Russians (Eastern Slavs) and Indians (haplogroup R1a1)

Let us try to go deeper in time, to 4,000-5,000 years BP, and consider sixteen 25-marker haplotypes of the Russians, so-called Eastern Slavs, of haplogroup R1a1. Between 50% and 70% of the current Russian population consists of the R1a1 Eastern Slavs. The haplotypes were collected in YSearch database, sections R1a and R1a1 combined. There was no difference between haplotypes in these two sections.

Two imaginary base haplotypes among sixteen 25-marker haplotypes would have referred to ln(16/2)/0.046 = 45 generations to a common ancestor, and one imaginary base haplotype among 16 would have given 60 generations to a common ancestor, that is 1,500 years BP. No surprise, that there was no any single base haplotype among those 16 Russian haplotypes, since their common ancestor lived well before those times.

In fact, those 16 haplotypes indeed contained no actual base haplotypes, and 115 mutations from the deduced base haplotype

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

Without a correction for back mutations this would give 115/16/0.046 = 156 generations, with the correction it gives 185 generations to a common ancestor, that is 4,625 years BP, 27th century BC.

Since about 3400 years BP these haplotypes – along with the R1a1 haplogroup – were brought to India by the same people, one can expect to see similar haplotypes of haplogroup R1a1 in India as well, with a time span to a common ancestor similar or a bit less, compared to that shown above.

Here are the 25-marker base haplotype of Indians-Pakistani deduced from their haplotypes in YSearch database:

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

There are only 2 mutations between these 25-marker base haplotypes, which shows a very close common ancestor for the both haplotypes, the Russian (Eastern Slavs) and Indian-Pakistani. Twelve 25-marker Indian-Pakistani haplotypes, collected from YSearch database contain 86 mutations from the above base haplotype, which translates to 167 generations, that is 4,175 years from their common ancestor, as it was generally predicted above.

When these two sets were pooled together, these 28 Indian-Russian 25-marker haplotypes showed their common ancestral R1a1 haplotype

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

which is identical with the Russian ancestral haplotype. All 28 haplotypes contain 213 mutations, which translates to 198 generations to a common ancestor, that is 4,950 years BP.

R1a1 in the Gulf of Oman

Sixteen 17-marker haplotypes of R1a1 haplogroup in males from United Arab Emirates and Qatar have been published (Cadenas et al, 2008). Only ten markers have been used in this study to make data truly comparable with other calculations of sets of R1a1 haplotypes. Their deduced and actual base haplotype (the only one that did not have mutations at the mutation minimization procedure) showed little discrepancy between a number of generations, calculated from its single presence in the set, ln(16/1)/0.026 = 107 generations to a common ancestor, and 46 mutations in the whole haplotype set, and gave 46/16/0.026 = 111 generations to a common ancestor. Despite the (apparent) similarity of the two figures, there were two concerns: (a) an instability of calculations at just one base haplotype, and (b) a high number of mutations at DYS##19 and 389-1. It looked as the set tends to split into two subsets. A haplotype tree (will be

published elsewhere) confirmed this thought, since it splits indeed into two distinct branches, with the following 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**

They differ by two mutations (shown in bold). It is of interest that each of them differs by just one mutation from the base Eastern Slav R1a1 haplotype, but by two mutations from the Donald family base R1a1 haplotype (see above). It looks like we are walking in footsteps on the ancient Aryans.

The first Arabic R1a1 haplotype set has 13 mutations in all the 7 haplotypes, which results in 77 generations from a common ancestor, who lived about 1,900 years BP, in the 1st century AD. The set shows only one base haplotype, as shown above, however, it is too few for calculations (although formally it gives 75 generations from a common ancestor, which can be merely a plain coincidence). The second set has 26 mutations in all the 9 haplotypes, which results in 125 generations, or 3,125 years to a common ancestor. Now it brings us to the Aryan times, the 12th century BC.

R1a1 in Anatolia

36 of 10-marker haplotypes of haplogroup R1a1 were listed in (Cinnioglu et al, 2004). The locus DYS#A7.2 (DYS#461) did not belong to the first 37 loci in the standard FTDNA format and was eliminated from this study. The whole set contained 5 identical (presumably base) haplotypes and 80 mutations from that base haplotype. This would give ln(36/5)/0.025 = 79 generations and 80/36/0.025 = 89 generations to a presumably common ancestor. The discrepancy was noticeable, particularly in view of the fact that the "base haplotype" contained too many mutations in DYS##19 and 389-2. They made the base haplotype splitting into at least two different ones.

The haplotype tree (will be considered in detail elsewhere) confirmed the observation. There were at least three distinct branches on the tree, with base haplotypes

13-25-16-10-X-Y-Z-12-10-13-11-30

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

13-25-16-11-X-Y-Z-12-11-13-11-30

The first base haplotype was deduced from an 11-haplotype branch, which contained 26 mutations. This translated to 117 generations, or 2,900 years to a common ancestor. This base haplotype differs by just one mutation from the Easter Slav base haplotype and from a "more recent Gulf of Oman base haplotype" (see above), and by four mutations from the Donald base haplotype. Clearly, the Western R1a1 haplotypes are less related to the Middle East and Asia Minor compared with the Eastern Slav haplotypes.

The second base haplotype was deduced from a 10-haplotype branch, which contained 17 mutations. This translated to 82 generations, or 2,050 years to a common ancestor. This base haplotype differs again by only one mutation from the Easter Slav base haplotype and by two mutations from the Donald base haplotype, and exactly fits with a "more recent Gulf of Oman base haplotype", which is 1,900 years old. It looks like it belongs to the same common ancestor with the last one.

The third base haplotype is obviously the youngest one. It belong to a 11-haplotype branch which contains only 12 mutations, that translates to 51 generation, or about 1,300 years to a common ancestor, who lived in the 8th century AD. With its DYS#439 = 11 in the base haplotype it has a rather distinct sequence among R1a1 base haplotypes. It is apparently a product of recent mutations.

R1a at the Adriatic coast

Many Croatian eight-marker haplotypes were published in (Barac et al, 2003; Barac et al., 2003), and 67 nine-marker R1a1 haplotypes were additionally provided to me personally by Dr. Marijana Pericic.

Consideration of the R1a1 haplotypes showed that they are certainly derived from more than one common ancestor. The total set contains three apparently base haplotypes, that would indicate 124 generations to a "common ancestor", while a number of mutations, which amounts to 172 total, points at 172 generations to a "common ancestor".

This prompted me to draw a haplotype tree, which – of course – was not a uniform one (it will be described and discussed in future publications). It contained at least four distinct branches, of which the "youngest" one, containing 12 haplotypes, included all those three base haplotypes

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

and had 14 mutations in the branch. This gives ln(12/3)/0.025 = 55 generations and 14/12/0.025 = 47 generations to a common ancestor. With a correction for back mutations the last figure gives 55 generations, that is 1,375 years to a common ancestor, who lived in the 7th century AD.

The above base haplotype is only one mutation away from the Eastern Slav base haplotype:

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

which belonged to a common ancestor who lived 3,600 years BP (see above).

Another "old" base haplotype in the tree was deduced from a wide 27-haplotype branch

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

It precisely fits to the above Eastern Slav base haplotype, and contains 68 mutations in all 27 haplotypes. This results in 126 generations, that is 3,150 years from a common ancestor.

However, the oldest branch on the tree provided with the most striking results. It 13 haplotypes contained 77 mutations, that translates to 356 generations, or 8,900 years to a common ancestor. This is the most ancient common ancestor for R1a1 haplotypes observed thus far.

Swedes (haplogroups R1a1, R1b1c, I1, I2b1, N3, G, and J)

A set of 383 Swedish haplotypes was listed in (Karlsson et al, 2006). It included 48 haplotypes of haplogroup R1a1, 76 haplotypes of haplogroup R1b3, 133 haplotypes of haplogroup I1 (former I1a), 15 haplotypes of haplogroup I2b1 (formerly I1c), 7 haplotypes of haplogroup G, 72 haplotypes of haplogroup N3, and 11 haplotypes of haplogroup J. These sets of haplotypes are of a particular interest since they provide comparisons of time spans to common ancestors which turned out to be mainly single common ancestors (from a viewpoint of DNA genealogy) and moved to Sweden in different and unrelated to each other times.

R1a1

R1a1 base haplotype in Sweden was as follows:

13-25-15-11-11-14-X-Y-Z-13-11-30

A haplotype tree (will be published in later issues of the Proceedings) strongly pointed to a single common ancestor for all 48 haplotypes. It is of interest to compare the Swedish R1a1 base haplotype with a Slavic (Russian) one, given above:

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

There is only one mutation between their 9-marker haplotypes (with other three markers missed in the Swedish study most likely the same), which places these two base haplotypes only about a thousand years apart.

These 48 Swedish R1a1 haplotypes contain 137 mutations compared to their base haplotype. There are only three base haplotypes among the whole set. This gives ln(48/3)/0.025 = 111 generations to a common ancestor. The number mutations results in 137/48/0.025 = 114 generations from a common ancestor. Indeed, we have a "clean" haplotype set, pointing at a single common ancestor. However, since the number of generations exceed a hundred, we need to introduce a correction for back mutations, which will increase the time span by about 30% in this particular case.

Indeed, the set gives 0.317 mutations per marker in a 9-marker haplotype, which with a correction for back mutations gives 145 generations, or 3,625 years to a common ancestor, the 17th century BC. The Russian R1a1 set gives 185 generations to a common ancestor, that is 4,625 years BP, the 27th century BC. It is a thousand years older, as it follows from the base haplotypes themselves.

Evidently, it took about a thousand years for ancestors of the Eastern Slavs to get to Scandinavia. Their descendants now amount about 13% of the Swedish haplotype set.

R1b1c

R1b1c base haplotype in Sweden was as follows:

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

This is a classical Atlantic Modal Haplotype, which in a 12-marker format is (see also the subsequent paper in this issue):

13-24-14-11-11-14-12-12-12-13-13-29

The 76 Swedish R1b1c haplotypes contain 187 mutations compared to their base haplotype. There are as many as 7 base haplotypes in the set, that is those which do not have mutations compared as the above haplotype. This gives ln(76/7)/0.025 = 95 generations to a common ancestor. A number of mutations results in 187/76/0.025 = 98 generations from a common ancestor. Again we have a "clean" haplotype set, pointing at a single common ancestor. However, since the number of generations is approaching a hundred, we need to introduce a correction for back mutations, which will increase the time span by about 25% in this particular case.

The set gives 0.273 mutations per marker in a 9-marker haplotype, which with a correction for back mutations gives 123 generations, or about 3,100 years to a common ancestor.

Again, a haplotype tree (will be published later) strongly confirmed a single common ancestor for all 76 haplotypes in the set.

Interesting that 3,100 years is exactly the same time span as to a common ancestor for Atlantic Modal Haplotype in Europe. 104 haplotypes randomly picked from YSearch database produced a haplotype tree split to two large branches both of which had the same AMH ancestor who lived 122 and 124 generations BP, respectively, that is the same 3,100 years ago.

One of the oldest Jewish R1b* haplotype, which happened to be the Atlantic Modal Haplotype, is derived from an ancestor who lived 200±5 generations, that is 5,000±200 years BP (see the subsequent paper in this issue). It is significantly older than West European R1b common ancestors from which present bearers of R1b haplotypes descended.

A close figure of 136 generations, or 3,400 years to a common ancestor of the haplogroup R1b (with the AMH as the base haplotype) was obtained for the Flemish population, as it is described below. This places a Flemish R1b common ancestor only 300 years before a Swedish R1b common ancestor, within a certain error margin, of course.

It is interesting why the Jewish AMH is significantly older that the European AMH. It seems that the descendants of the European AMH have passed a severe

population bottleneck around 3,500 years BP, and started the lineage practically all over again. That is why the AMH is so numerous nowadays, and not eroded with mutations. It is just still "fresh". This bottleneck was not related to the Jews in the Middle East and Africa those times, hence, their R1b haplotypes are much older. The same is related to Armenian R1b haplotypes, as it is shown below.

I1

Haplotypes of haplogroup I1 split to several lineages (branches in the haplotype tree) and cannot be calculated as one combined set. This is illustrated by many mutations along with a high number of the base haplotypes

13-23-14-10-14-14-X-Y-Z-12-11-28

in the 133-haplotype set. There are 28 identical haplotypes as shown above with as many as 302 mutations in the whole set. It gives ln(133/28)/0.025 = 62 generations, which contradicts to 302/133/0.025 = 91 generations (without a correction for back mutations). This discrepancy shows that haplotypes in the set are derived from multiple (at least two) common ancestors. Examples of calculations in these situations are presented in the subsequent paper in this issue.

I2b1

There are only 15 haplotypes of this haplogroup in the 383-haplotype Swedish series, four of them identical (base) haplotypes:

14-23-15-10-15-15-X-Y-Z-14-12-32

The whole 15-haplotype set contains 28 mutations from the above base haplotype. Again, as in the case of I1 haplotypes, there is a significant discrepancy between a number of generations obtained from a fraction of base haplotypes, ln(15/4)/0.025 = 53 generations, and that from mutations, 28/15/0.025 = 75 generations (without a correction for back mutations). Typically, such a discrepancy implies that the whole set contains a subset of haplotypes from a rather recent ancestor, hence, a superposition of two subsets with different common ancestors. In those situations calculations of a timespan to a "common ancestor" across the mix of subsets results in wrong estimates. Subsets should be separated before any calculations of this kind. The literature, unfortunately, is full of these wrong miscalculations of "TMRCA"s.

It is of interest that a difference between the base haplotypes for Swedish I1 and I2b1 equals to as many as 11 mutations per 9-marker haplotype. This difference corresponds to 1.22 mutations per marker in the 9-marker haplotypes, and leads to 25 thousand years of difference between the respective ancestral haplotypes. With all uncertainties in estimates of time spans to their common ancestors (around 2-3 thousand years BP for both of them), the difference means that THEIR common ancestor lived about 14-15 thousand years BP. This is apparently as good as it gets down to the most ancient common ancestors of the nowadays bearers of haplogroups I1 and I2.

N3

72 haplotypes in the Swedish series contain 18 identical (base) haplotypes

14-24-14-11-11-13-X-Y-Z-14-14-30

which are the same as the base N3a Baltic Sea region haplotype show above:

14-24-14-11-11-13-11-12-10-14-14-30

These 72 haplotypes contain 149 mutations, that give 83 generations to a common ancestor. However, we have again a discrepancy with a number of generations obtained from a fraction of the base haplotype, which is ln(72/18)/0.025 = 55 generations. The set contains haplotypes from different common ancestors, and the respective subsets should be treated separately. This will be done in a study published in one of the following issues of the Proceedings. The Baltic Sea region N3a haplotypes shown above gave a clean set of haplotypes with a common ancestor who lived 45±5 generations BP, that is about 1,125±125 years BP, 8-10th century AD.

G

Seven Swedish haplotypes of G haplogroup are derived from an old ancestral haplotype

14-22-15-10-14(15)-15-X-Y-Z-12-11-29

This haplotype was deduced from those seven, since all of them are mutated. They contain 22 mutations, which corresponds to about 162 generations, or about 4,000 years to a common ancestor. It is likely that this ancestor was not an old inhabitant of nowadays Sweden, but it was a common ancestor elsewhere. His descendants came to Scandinavia either in different times, or there were a number

of descendants of the same common ancestor who lived elsewhere 4,000 years BP.

J

The situation with the Swedish haplotypes of haplogroup J mirrors that with those of haplogroup G. There are eleven of them in the Swedish series. All of them are heavily mutated, and there is no "base" haplotype among them. An ancestral (base) haplotype deduced by minimization of their mutations is as follows:

12-24-14-10-13-17-X-Y-Z-13-11-30

It is just one mutation away (marked in bold) from the most ancient Jewish base haplotype of haplogroup J1

12-24-14-10-**12-**17-11-15-13-13-11-30

which is described in the subsequent article in this issue, and whose ancestor lived 9,200±1,200 years BP

All 11 Swedish haplotypes contain 47 mutations, that correspond to 234 generations, or about 5,900 years to a common ancestor. Again, it is likely that descendants of this common ancestor came to Scandinavia either in different times, or there were a number of them, and in this case they could have come to Sweden any time in the past. Their common ancestor as well a common ancestor of their descendants would be still the same ancient ancestor who lived elsewhere.

R1b in the British Isles

In (Campbell, 2007) a set of 1242 R1b haplotypes was presented. Unfortunately, they were not all haplotypes from the complete set, which amounted 1625 haplotypes. In fact, only 50 of 10-marker haplotypes were shown out of 291 haplotype in the complete set. In other words, 383 haplotypes were missing in the presented set, 241 of them were "other" haplotypes, less frequent.

Let us see what a non-complete set would give us in terms of a calculated time span to a common ancestor.

These 1242 haplotypes contained 1708 mutations compared to the base (ancestral) haplotype, which is present in the set in the amount of 262 repeats

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

and which in fact is the Atlantic Modal Haplotype in its 12-marker format

13-24-14-11-11-14-12-12-12-13-13-29

Considering the number of mutations, we will calculate certainly a distorted number of generations to a common ancestor, since we are missing 383 haplotypes with many mutations in them. We can surmise that we are going to get a largely reduced number of generations, which is 1708/1242/0.026 = 53 generations to a common ancestor, that is merely 1325 years BP, the 7th century AD.

Now, taking the approach based on a fraction of base haplotypes, we get ln(1625/262)/0.026 = 70 generations, or 1,750 years to a common ancestor. It gives still a rather recent time when a common ancestor of the AMH lived, that is the 3rd century AD.

It all says that the AMH is derived from a rather recent ancestor, hence, there are so many of the AMHs among the current bearers of R1b haplogroup in Europe. However, the figure of 1,750 years to a common ancestor is certainly low, and apparently reflects a heavy admixture of descendants of recent ancestors in the 1,625 haplotype set.

R1b in the Flemish population

A set of 64 of 12-marker haplotypes was presented in (Mertens, 2007). However, the haplotypes did not include DYS##388 and 426, and had two "scattered" markers from the 25- and 37-marker panels instead. In order to stay within the calibrated range (though it would have not been difficult to re-calibrate the set) I used ten well-established markers, and did not consider DYS##437 and 438.

The 10-marker haplotypes had the following deduced base haplotype

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

which is the classical Atlantic Modal Haplotype in its 12-marker format

13-24-14-11-11-14-12-12-12-13-13-29

Incidentally, two markers, DYS##437 and 438, removed from the calculation, had the "mutation-minimized" values of their alleles 15 and 12, respectively, that also belong to the classical AMH.

There are only two of the base (ancestral) haplotypes in the whole 64-haplotype set. This signals that there are no admixture of a "recent" AMH ancestor, otherwise there would be many identical (base) AMH in the set. Hence, we are likely going to obtain a number of generations not distorted by subsets of recent ancestors.

There are 199 mutations in this 64-haplotype set, that gives 199/64/0.026 = 120 generations from a common ancestor. Based on two residual ancestral haplotypes we obtain ln(64/2)/0.026 = 133±15 generations. The last value is within the error margin with 120 generations, because just a routine inclusion of one more base haplotype into the set would give ln(65/3)/0.026 = 118 generations. Obviously, too few base haplotypes in a haplotype set makes the resulting number of generations very unstable and subject to a high error margin. At any rate, the data supports a conclusion that it is very likely that all 64 haplotypes in the set are descended from a single ancestor. Indeed, a haplotype tree (will be discussed in a future publication concerning R1b haplogroup common ancestors) supports this conclusion. With a correction for back mutations we obtain, that a common ancestor of all 64 individuals haplotypes of which composed the set, lived 136±8 generations ago, that is 3,400±200 years BP. This figure is close to 3,100 years BP obtained earlier for a common ancestor for R1b in Europe.

I1 (former I1a) in the Flemish population

A set of 22 of 12-marker haplotypes was presented in (Mertens, 2007). As it is described in the preceding section, I have employed 10-marker haplotypes for a timespan estimate to a common ancestor of those 22 individuals. Three of those haplotypes were identical, hence, present the following base (ancestral) haplotype:

13-22-14-10-13-14-X-Y-12-12-11-28

One can see that it differs by two mutations from the I1 Swedish base haplotype (see above), apparently derived from a recent common ancestor. However, it exactly matches the I1 base haplotype in the Jews (see the subsequent paper in this issue), which, as it is explained there, is not the Jewish at all, and was "borrowed" from the surrounding European I1 community:

13-22-14-10-13-14-11-14-12-12-11-28

In the Jews this haplotype is continuing for 4,100±1,900 years (ibid.). Let us see a Flemish side of the story.

The three base haplotypes out of 22 total give ln(22/3)/0.026 = 77 generations to a common ancestor. Since all the 22 Flemish haplotypes have 43 mutations from the above base 10-marker haplotype, this gives 43/22/0.026 = 75 generations to a common ancestor. This is a practically perfect match, which witnesses that all the 22 individuals had a single common ancestor, and that the haplotype set is "clean", it reflects only one lineage. When a correction for back mutations is introduced, we obtain that a common ancestor of this lineage lived 83 generations, that is about 2,100 years BP. This is on a lower side of the error margin in the Jewish series. It might be that it is the same common ancestor.

Actually, this figure sits well between 62 and 91 generations for the Swedish I1 common ancestor (see above). 62 generations is certainly inclined to a more recent ancestor of I1 haplotypes in Sweden, hence, the figure of 91 generations can be distorted either way. However, one can see that all these figures sit more or less in the same time frame. More detailed analysis of these haplotypes will be done elsewhere. The main purpose of this particular study is to illustrate the logic and principles of calculations and criteria which shape the approach.

Basque R1b* haplotypes

Basque DNA Project (see Ref. at the end) lists 44 of twelve-marker R1b* haplotypes, 17 of 25-marker haplotypes and a lesser amount of 37- and 67-marker haplotypes. There are no base haplotypes in the 12-marker set, and the set clearly tends to split into two subsets, with deduced base haplotypes, which differ from each other by two mutations (shown in bold)

13-24-14-11-11-14-12-12-12-**13**-13-**29**

13-24-14-11-11-14-12-12-12-**14**-13-**30**

The upper base haplotype is a classical Atlantic Modal Haplotype in its 12-marker format.

The whole 12-marker haplotype set contains 140 mutations from the AMH, which corresponds to 153 generations, or about 3,800 years from its common ancestor. However, it is clear that such a calculation would be flawed, since the

set contains at least two common ancestors. Indeed, a 12-marker haplotype tree (not shown here, since it will be considered in future publications) confirms such an observation. The tree splits into two major branches with base haplotypes shown above. A 25-haplotype tree rearranges the branches and makes a separation between them even more distinct. A larger, older branch has the following base haplotype

13-24-14-11-11-14-12-12-12-13-13-29-18-9-10-11-11-25-15-19-28-15-15-17-17

which gives 156 and 157 generations to a common ancestor from its 12-marker and 25-marker panels, respectively, or about 3,900±100 years to a common ancestor, who had the AMH as its ancestral haplotype. The second, younger branch, provides 1,600 and 2,000 years to its common ancestor based on its 12- and 25-marker panels, or about 1,800±200 years to its common ancestor, who lived at the beginning of AD. Its base haplotype

13-24-14-11-11-14-12-12-12-14-13-30-17-9-10-11-11-25-14-18-29-15-15-16-17

differed by 7 mutations per 25-marker haplotype, which translates to about 180 generations, or 4,500 years difference between these two base haplotypes. Hence, they cannot be derived directly from each other. Their common ancestor lived about 5,100±300 years BP.

References

Athey, W. Mutation rates – who's got the right values? J. Genetic Genealogy 3, i – iii, 2007

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

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

Basque DNA Project – http://www.familitreedna.com/public/BasqueDNA

Behar, D.M., Thomas, M.G., Skorecki, K., Hammer, M.F., Bulygina, E., Rosengarten, D., Jones, A.L., Held, K., Moses, V., Goldstein, D., Bradman, N and Weale, M.E. Multiple origins of Ashkenazi Levites: Y chromosome evidence

for both Near Eastern and European ancestries. Am. J. Hum. Genet. 73, 768-779 (2003).

Behar, D.M., Garrigan, D., Kaplan, M.E., Mobasher, Z., Rosengarten, D., Karafet, T.M., Quintana-Murci, L., Oster, H., Skorecki, K. and Hammer, M.F. Contrasting patterns of Y chromosome variation in Ashkenazi Jewish and host non-Jewish European populations. Hum. Genet. 114, 354-365 (2004).

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., Hill, K., Hurtado, A.M., Castro-de-Guerra, D., Torres, M.M., Groot, H., Michalski, R., Nymadawa, P., Bedoya, G., Bradman, N., Labuda, D. and Ruiz-Linares, A. Y-chromosome evidence for differeing ancient demographic histories in the Americas. Am. J. Hum. Genet. 73, 524-539 (2003)

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

Campbell, K.D. Geographic patterns of haplogroup R1b in the British Isles. J. Gen. Genealogy 3, 1-13 (2007)

Chandler, J.F. Estimating per-locus mutation rates. J. Genetic Genealogy 2, 27-33, 2006

Cinnioglu, C., King, R., Kivisild, T., Kalfoglu, E., Atasoy, S., Cavalleri, G., Lillie, A.S., Roseman, C.C., Lin, A.A., Prince, K., Oefner, P.J., Shen, P., Semino, O., Cavalli-Sforza, L.L. and Underhill, P,A. Excavating Y-chromosome haplotype strata in Anatolia. Hum. Genet. 114, 127-148 (2004)

Cordaux, R., Bentley, G., Aunger, R., Sirajuddin, S.M. and Stoneking, M. Y-STR haplotypes from eight South Indian groups based on five loci. J. Forensic Sci. 49, 1-2 (2004)

Derenko, M., Malyarchuk, B., Denisova, G., Wozniak, M., Grzybowski, T., Dambueva, I. and Zakharov, I. Y-chromosome haplogroup N dispersals from south Siberia to Europe. J. Hum. Genet. 52, 763-770 (2007)

DNA-Project.Clan-Donald, http://dna-project.clan-donald-usa.org/tables.htm

Felsenstein, J. Phylip, the Phylogeny Inference Package. PHYLIP, version 3.6. Department of Genome Sciences, University of Washington, Seattle, 2005

The International HapMap Consortium. A second generation human haplotype map of over 3.1 million SNPs. Nature 449, 851-862 (2007).

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., Golge, M., Usanga, E., Papiha, S.S., Cinnioglu, C., King, R., Cavalli-Sforza, L., Underhill, P.A. and Villems, R. The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. Am. J. Hum. Genet. 72, 313-332 (2003)

Mertens, G. Y-Haplogroup frequencies in the Flemish population. J. Gen. Genealogy 3, 19-25 (2007)

Sahoo, S., Singh, A., Himabindu, G., Banerjee, J., Sitalaximi, T., Gaikwad, S., Trivedi, R., Endicott, P., Kivisild, T., Metspalu, M., Villems, R. and Kashyep, V.K. A prehistory of Indian Y chromosomes: evaluating demic diffusion scenarios. Proc. Natl. Acad. Sci. US, 103, 843-848 (2006)

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., Rani, M.V.U., Thakur, C.M., Cavalli-Sforza, L.L., Majumder, P.P., and Underhill, P.A. 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 (2006)

Weale, M.E., Yepiskoposyan, L., Jager, R.F., Hovhannisyan, N., Khudoyan, A., Burbage-Hall, O., Bradman, N. and Thomas, M. Armenian Y chromosome haplotypes reveal strong regional structure within a single ethno-national group. Hum. Genet. 109, 659-674 (2001)

Zhivotovsky, L.A., Underhill, P.A., Cinnoglu, C., Kayser, M., Morar, B., Kivisild, T., Scozzari, R., Cruciani, F., Destro-Bisol, G., Spedini, G., Chambers, G.K., Herrera, R.J., Yong, K.K., Gresham, D., Tournev, I., Feldman, M.W., and Kalaydjieva, L. The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. Am. J. Human Genet. 74, 50-61 (2004).

Origin of the Jews via DNA Genealogy

Anatole A. Klyosov

Newton, Massachusetts 02459, U.S.A.
http://aklyosov.home.comcast.net

Summary

Some fifteen tribes to which most of the Jews commonly belong have been considered in terms of DNA genealogy. More than forty of the Jewish lineages have been identified, some of them as ancient as 13-14 thousand years before present (BP).

A large fraction of lineages and haplotypes of the Jews have an ancient origin, and have appeared in the Middle East. This is mainly related to tribes/haplogroups which are designated with letters J1, J2, E3b, and embrace about two third of nowadays Jews. These tribes can be traced up to 500-600 generations in depth. Naturally, the tribes themselves could have appeared much earlier, however, in many cases their offspring and/or their descendants did not survive.

Another third of the Jews according to their Y-chromosome belong to other tribes/haplogroups which historically are not commonly associated with bearers of the Jewish traditions, customs, religion. These tribes are such as R1a1, R1b, R2, Q, G, K, I. In each case a story of a penetration of those tribes to a Jewish community presents a natural mystery, a secret of which was carried to a grave long ago. It could be a romantic story, it could be a story far from a romantic one. Some of the stories happened three-to-four thousand years ago, when Jewry was only being established. It was time of an early unification and consolidation of different peoples based on common beliefs, on common lifestyle, on common values. In some cases an entry of a new haplogroup and new haplotypes to a Jewry happened to be in 8-11th century AD; it well might be times of consolidation of some part of the Jewish Diaspora in the Khazar Kaganat. In some cases new entries of haplogroups and haplotypes happened only 300-400 years ago; these were times of shaping of Jewish communities in Eastern Europe, and there were sometimes accompanied either with demographic expansions among the Jewry, or with passing by them a population bottleneck. Some striking

cases explained in this study have been apparently related to the Black Plague in Europe in the 14th century.

The time spans directly related to origin of the Jews in various tribes are as follows:

J1 (overall)	17,500±1,000 years BP
J1 (and below, Jewish)	15,500±4,100
K, T	8,500±1,000
E1b1b	6,800±400
J2	6,000±1,000
R1b	5,000±200
I1	4,100±1,900
G2	4,000±200
G1	3,600±200
I2	1,400±600
R1a1	1,350±300
Q	675±125
R2	650±50
G2c	575±50
I21b	500±125

The paper shows how these estimates were obtained and interpreted.

Introduction

As it was discussed in the preceding paper in this issue, origin of peoples in the context of DNA genealogy is an assignment of each of them to a particular tribe or its branch initiated (in a genealogical sense) by a particular ancestor who had a base haplotype (an "ancestral" haplotype). This also includes an estimation of a time span between the common ancestor and its current descendants.

This study aims at origin of the Jews. Before that, I will present a few lines of a general description of some related tribes in terms of DNA genealogy. As it will be shown below, the Jews have their roots in eight principal tribes of a total current number of twenty major tribes, defined in DNA genealogy. The main tribes are designated by letters from A through T, and the eight main tribes are – in the alphabetical order – E, G, I, J, K, Q, R, and T. Each one of them had split into their derivative, downstream tribes, some as long as 30-40 thousand years ago, some only 10-12 to 20 thousand years ago. Each of the tribes presumably, or rather by definition, had a founder, a man, an ancestor of the tribe. An assignment

to a tribe is determined experimentally, using procedures developed – and still being developing – in the framework of DNA genealogy.

Considering derivative, downstream tribes, each one approximately of 10 to 40 thousand years of age, the Jews are related to about thirteen of their sub-groups E1b1b, J1, J2, R1b, R1a1, G2c, Q, I, T, R2, K, G2, G1, each of which also includes sub-groups. They are listed here in the order of estimated number of the Jews in each one of them. Some Jews are assigned to some other tribes, but these events are relatively infrequent and can be considered as almost "accidental". Indeed, as a joke/wisdom goes, one mistake – and you are a father. Those infrequent cases of Jews in rare (for them) haplogroups will not be discussed here.

The next important question – for how long the Jews of a particular tribe "belong" to it in terms of their continuing male lineage, counting from the founding ancestor? How deep, time-wise, those lineages are tracing back in time from the present-day Jews? Clearly, there is a difference, it is for the last 200-300 years, or for some ten or even twenty thousand years. Therefore, time-wise estimates are so much emphasized in this paper, namely, how long ago a common ancestor of present-day Jews in each particular tribe had lived. I will shutter the intrigue of the story and inform readers up-front that the deepest Jewish ancestors in each one of the mentioned ten tribes lived as follows (in years):

J1 (overall)	17,500±1,000 years BP
J1 (and below, Jewish)	15,500±4,100
K, T	8,500±1,000
E1b1b	6,800±400
J2	6,000±1,000
R1b	5,000±200
I1	4,100±1,900
G2	4,000±200
G1	3,600±200
I2	1,400±600
R1a1	1,350±300
Q	675±125
R2	650±50
G2c	575±50
I21b	500±125

Naturally, these figures do not describe how long ago these haplogroups were originated (except probably the top figure). First, they consider Jewish lineages only. Second, these figures refer either to genealogical bottlenecks, when a

population had shrank to almost an extinction of the tribe, and a survivor (or a few survivors) practically initiated a new lineage with a new base (ancestral) haplotype, whichever one he happened to bear. A number of Jewish lineages in many haplogroups have initiated from common ancestors who lived 600-700 years BP, in a middle of the 14th century, in the times of Black Death in Europe. Apparently, the Jews who escaped, fled to the East, and began new lineages which can be traced back to the Middle Ages.

A story will be told here on a bottleneck in the J1 Jewish population in the 7th century AD which led to about a 40% share of the "recent Cohen Modal Haplotype" among the J1 Jews today, compared to a 30% share of the "older Cohen Modal Haplotype" in the same haplogroup J1, though the "older" CMH is no more than 300-600 years older than the "recent" CMH. A common ancestor of the both CMH lineages lived about 4,000 years BP.

A similar in kind story can be told on the Jewish R1a1 lineage which was originated apparently among the Khazars in the 9th century AD and passed through a bottleneck in the 14th century, apparently during the same Black Plague times. The same story has apparently happened with the Jews of Q and R2 haplogroups.

Essentially, the figures above, along with their interpretation, represent a main conclusion of this paper. The rest is what does it all mean, how these numbers were obtained and based on what, and how reliable are these numbers.

Haplogroups in which the Jews are represented

If all and everyone male Jews had the same SNP ("snip") M267 which defines haplogroup J1, the answer to the title of this paper would have been obvious. J1 is the haplogroup of the Middle Eastern origin. This would have meant that direct ancestors of all living Jews lived in the Middle East. And all direct ancestors of all Jews would have been members of the same ancient tribe.

But it is not so.

Direct ancestors of the Jews, as follows from their snips and haplogroups, belonged to various tribes. These were, as it was said above, first of all J1, J2, E1b1b, R1b, R1a1, G2c, I, Q, K, T, R2, G2, G1.

The first one, J1, as it was mentioned, is a sign of the Middle Eastern tribe. It is believed to be originated in the Southern part of Mesopotamia, or in the Southern

part of the Fertile Crescent, and from there ancestors of the Jews had arrived by a rather roundabout way to Erez-Israel.

J2 was originated, as it is believed, in the Northern part of Mesopotamia. Some say that it was the Northern part of the Fertile Crescent. Migrants down the road had brought it to Asia Minor and further to the North, to Armenia, as well as to the West, across Bosphorus and Dardanelle straits (which were dry land in the past) to the Balkans and Adriatic cost, to Apennines and neighboring territories. There is nothing surprising that the Jews split between J1 and J2 haplogroups. They split between dozen of other haplogroups on top of it. Peoples do split between haplogroups, particularly peoples having a rich (and bumpy) history. Which haplogroup, J1 or J2 is older in the Jewish males? I will try to answer these questions below.

Snip, which leads to E1b1b (formerly E3b) haplogroup, was originated in the Middle East, though some argue that it has appeared in the North Africa, from where it was brought via migration to the Middle East and Mediterranean areas. It is frequent among Greeks, Albanians, inhabitants of Southern Italy.

The snip that defines R1a1 haplogroup can be arguably called "Slavic", but it rather has the Aryan origin, if to borrow the name from the Indian Vedas. It was assigned largely to an ancient tribe which had arrived apparently between 10 and 12 thousand years ago to settle in the Southern part of modern-day Russia and Ukraine. These people named themselves Aryans in the Vedas which they brought to India and developed there. The Aryans had carried R1a1 around to large distances from the Southern part of the Eastern Europe to the Atlantics in the West, to Ural, Siberia, Central Asia, India and China in the East, to Mesopotamia, Egypt and Gulf of Oman in the South, to Apennines, Greece, Crete, Mediterranean and Adriatic islands in the South-West. The highest density of R1a1 is in East Europe, particularly in Russia, Ukraine, Belorussia, Poland, Slovakia, as well as in some areas of Central Asia, such as Kyrgyzstan and Tadzhikistan. The origin of R1a1 haplogroup sometimes is referred to the territory of nowadays Ukraine, however, it has never been substantiated. The oldest R1a1 haplotypes have been found thus far in the Balkans, close to the Adriatic Sea.

An unresolved question is – how R1a1 had gotten to the Jews? When? From whom? From the Aryans thousands years ago? From the Slavs rather recently? From the Khazars in the 8-11th centuries? Some data and thoughts will be provided below.

Haplogroup R1b with its sub-groups R1b1 and R1b1c are considered to be the West European haplogroups, which, of course, is not exactly true. Much older R1b haplotypes (meaning, from a much earlier common ancestor) can be found in Asia and in Armenia, for example. Basically the same questions as the above remain unresolved – when the respective snip had gotten to the Jews? Thousand of years back? Rather recently, in this millennium? From the Europeans, before the Jews have establish themselves as a ethnic, cultural, religious community? Or later, due to an invasion, such as from the Romans?

Haplogroup G is the old one. It has appeared allegedly in the Iranian East, or in the Middle East, or maybe in the modern day Pakistan and India. This is a relatively rare haplogroup, and it is unknown how it got into the Jewish community or well before the Jewry establishment. Naturally, it is tempting to assign it to the Khazars, as well as practically all "non-conventional" Jewish haplogroups, as many often do, without any time-wise calculations.

Haplogroup I is a numerous one in the Balkans (as a sub-group I2, formerly I1b) and in the North of Europe (as I1, formerly I1a). Along with R1b, it is carried by the majority of Western Europeans and Central Europeans. Only a minor fraction of the Jews have this haplogroup.

Haplogroup Q is observed mainly in the Siberian peoples and in the Native American Indians, somewhat in Central Asia and in Southern EuroAsian steppes, among Turk-speaking peoples. The Khazars?

Haplogroup K and its (former) subgroup K2, now a new haplogroup T, which are observed among the Jews, are rather rare. These are so-called EuroAsian haplogroups, which can be met on Mediterranean coastlines, in North-East Africa and in the Middle East, as well as in the South of Spain and France. It is commonly mentioned that Thomas Jefferson had haplogroup K, but nobody knows how his ancestors had acquired this haplogroup.

Haplogroup R2 is mainly (but not only) an Indian one. In this study I have suggested how it has gotten into the Jews and when.

Since extraordinary conclusions require extraordinary justifications, I have considered the methodology of the work in detail in the preceding paper in this issue. It describes how reliable are the mutation rates employed in the paper, how they were calibrated and verified, and how each of the listed haplogroups was analyzed, particularly when there was more than one common ancestor in mixed sets of haplotypes.

Part I

The Jews in Haplogroups J1 and J2

A Brief Overview

The literature on the Jews of haplogroups J1 (M267) and J2 (M172) contains mainly short, 6-marker haplotypes (Nebel et al, 2000, 2001a, 2001b; Thomas et al, 2000; Bonne-Tamir et al, 2003; Behar et al, 2003; Shen et al, 2004; Levy-Coffman, 2005), and in one case 10-marker haplotypes (Behar et al, 2004). Main observations and conclusions from the published studies can be summarized as follows:

(a) A significant number of the Jews carry the so-called "Cohen Modal Haplotype" in a 6-marker notation

14-16-23-10-11-12

Or in a 12-marker notation (FTDNA), in this case with two undefined markers (Behar et al, 2004)

12-23-14-10-X-Y-11-16-12-13-11-30
12-23-14-10-X-Y-11-16-11-13-11-30

The missing markers here are DYS##385a and 385b, replaced with X and Y above. These two 10-marker haplotypes are related to the parent haplogroup J (M304), formerly 12f2b* (Behar et al, 2004), and haplogroup J2 (M172), respectively, and extracted from (Behar et al, 2004).

(b) The above 6-marker haplotype was estimated to be originated 2,650 years ago, assuming a generation time as 25 years, or 2,100-3,250 years ago with a 95% confidence, and assigned to the Cohen Y-chromosome genealogy (Thomas et al, 1998).
NOTE: It will be shown below that this estimate is generally not true.

(c) The Cohen Modal Haplotype is rather frequent in both haplogroups J1 and J2. This is a kind of a mystery, according to researchers in the field, since the Cohens (kohen, kohanim, or cohanim, as a special status in Judaism) thereby appear to belong to two different genealogical lineages, that contradict with their heritage and traditions.

(d) The Jews belong to many different haplogroups (Levy-Coffman, 2005), and this is often explained by their Khazar heritage. These and similar explanations are not based, as a rule, on any time-wise estimates to common ancestors, and are claimed just matter-of-factly.

(e) The highest presence of "foreign", non-Middle Eastern, haplogroups was observed largely among Ashkenazi Jews, and among them Levites stand out with a relatively high frequency of their R1a1 haplogroup.

(f) A number of the Jews which belong to haplogroups J1 and J2 is approximately the same, and vary in different sets of haplotypes between 17% and 19% in both J1 and J2 haplogroups.

In order to examine these observations and conclusions with more extended haplotypes and to calculate time spans to common ancestors I have employed the commercial database YSearch, which contains more than 40 thousand haplotypes (including many repeats, though, which dot the database) and would be referred below as a "commercial" database. Generally, many would consider this as a rather shaky ground, since scientists typically (and reasonably) refer to any nonrandom population segment as "subject to ascertainment bias", and in practice would be dubious about the study. Therefore, special evaluations and justifications are required in those situations. In this particular case to justify the usage of the database I have collected and analyzed evidences that "scientific" selections of haplotypes and the "commercial" selections give the same or similar results in terms of haplotype tree structures and calculated time-wise estimates to common ancestors. It is important, that unlike the "scientific" set of haplotypes, the commercial data base contains many 12-, 25-, 37- and 67-marker haplotypes, including those of the Jews.

In order to compose a set of J1 haplotypes of the Jews, I have extracted from a section "Haplogroup J1" of YSearch database as many haplotypes as possible, whose bearers had ancestors (typically from the 19th century) with first names appeared to be Jewish, such as Aaron, Aharon, Aron, Avraham, Avram, Avrom, Baruch, Berel, Chaim, Ehiliy, Elya, Elyia, Ephraim, Feivish, Isaac, Israel, Itzok, Jankel, Joshua, Leib, Lejba, Meir, Mevshe, Moisey, Mordecai, Mordukh, Moses, Moshe, Nathan, Nochum, Reuven, Salomon, Shepshel, Shimon, Shlomo, Shmuel, Simon, Solomon, Tsvi, Tzvi, Yaakov, Yakov, Yechiel, Yehuda, Yerachmiel, Yesheyahu, Yeudah, Yisroel, Yitzchak, Zadok, Zebediah, and others, provided that the last names also appeared to be Jewish. They actually were, as a rule.

A set, containing 90 from the total number of 294 haplotypes in the "Haplogroup J1" section in the database (31% of total) was assembled, the respective 6-marker haplotype trees were drawn for the "scientific" and "commercial" sets (Fig. 1 and 2), and a number of mutations in the two sets were compared. Data are given in the next section.

It should be noted that numbers of the Jews in "Haplogroup J1" and "Haplogroup J2" sections identified as described above, were rather similar, 90 and 75, respectively. It is very unlikely that the Jews of either haplogroup send their DNA samples for the testing more often than in the other. However, since the haplogroup J2 section in the database contained 23 more Jewish names in its sub-group sections (one more haplogroup in each of J2a1e (formerly J2d) and J2a1b2 (formerly J2f2) groups, nine more in J2a1b (formerly J2f) of a total number of 58 there, and twelve more in J2a1b1 (formerly J2f1) of a total number of 36 there), while there were no Jewish names found in sub-groups of J1, then the ratio between Jewish haplotypes (picked as described above) was 90/98 in haplogroups J1 and J2. They are almost equal to each other, indeed.

With respect to a fraction of the Jewish haplotypes in haplogroups J1 and J2 overall, it was 31% of them in J1 (90 out of 294) and 7% in J2 (75 out of 1035). In sub-groups, the higher fraction was in J2a1b1, 33% (12 out of 36), that approximately the same as that in haplogroup J1. Clearly, haplogroup J1 is more representative for Middle-Eastern Jewish haplotypes (and for any other haplogroup, as it will be shown below) by a relative content of Jewish names, compared to J2.

Before considering in detail the Jewish haplotype trees in haplogroups J and J1, here are the principal finding in this section.

Haplogroup J1, Jewish and non-Jewish, is about 17,500±1,000 years old. The age of lineages which trace to present day Jews, or, in other words, a time span to the most ancient Jewish ancestors in haplogroup J1 is 15,500±4,100 years. There is a number of ancestors of particular Jewish lineages, which can be traced down in millennia following patterns of mutations in present day Jewish haplotypes in haplogroups J and J1. We have to mention "J" since some studies did not differentiate haplogroups J and J1. Among common ancestors which have been identified in this study are those of haplogroups J* and J1, who lived 11,100±600 years BP, as well as 9,200±1,200, 6,200±400, 5,100±500, 3,500±200 years BP, and several more recent lineages, five of them were identified (see Table at the end of this paper).

A number of Jewish lineages successfully passed some severe bottlenecks, and actually began new lineages. Naturally, they continued preceding lineages, however, they have started new branches on a haplotype tree. In some cases some "external", not Jewish, lineages entered the Jewish community, and started new branches among the Jews. In those cases "common ancestors" can occur outside of the Jewish circle.

One way or another, through a bottleneck or an "invasion", a number of Jewish lineages have been initiated rather recently, in the Diaspora times. The most significant ones, in terms of a number of their current descendants, are five Jewish lineages:

- The "recent Cohen Modal Haplotype" (rCMH), whose common ancestor lived 1,350±100 years BP, around a middle of the 7th century. This ancestor might well have been Bustenai ben Hanina, aka Bostenai ha-David, direct descendant of King David, the Fist Exilarch of the Third Dynasty, and the last holder of the Resh Galusa ("Head of the Diaspora") noble title, who lived in 590-670 AD. Since he was not a cohen, his descendants were not cohens as well. This explains why so many not cohens are bearers of the "CMH" (about 60%-80% of them).
- The "older Cohen Modal Haplotype" (oCMH), which is identical to the rCMH in the 6-marker format (this is the main reason of the confusion within the CMH), was derived from a common ancestor who lived 1,650±350 years BP, around 350±350 years AD. This by all means was a Cohen, and this explains why so many cohens are among bearers of the "CMH", however, far from to be all of them (about 20-40% of them).
- A common ancestor of the both rCMH and the oCMH lived 4,300±500 years BP. Incidentally, according to the Torah scholars, these are Abraham times.
- Altogether more than half of the Jews in haplogroups J and J1 are the bearers of the "CMH" (including its mutated variants), and most of those have the "CMH" in its non-mutated 6-marker format due to its rather recent origin. Overall, about 40% of the Jews in haplogroups J and J1 are direct descendants of King David, and between 10% and 20% are descendants of the Cohen-the-survivor at the beginning of the first millennium AD.
- Three more rather recent common ancestors of the present day Jews in haplogroup J1 lived 1,800±200, 1,600±200, and 510±50 years BP. Their current descendants account for about 10%, 20%, and 5% of the population, respectively. The rest of the Jews descended directly from ancient lineages.

Now let us consider a justification of the above conclusions.

The Jews of Haplogroup J* and J1

Figs. 1 and 2 show a J* and J1 6-marker haplotype trees, respectively, constructed from the published "scientific" data (Behar et al, 2003) and collected from YSearch commercial database as described above. As far as I know, there are no "scientific" data available on J1 Jewish haplotypes, with a proven M267 snip. The "scientific" J* (an upstream haplogroup with respect to J1) data set contains 194 haplotypes, and the "commercial" J1 set contains 90 haplotypes.

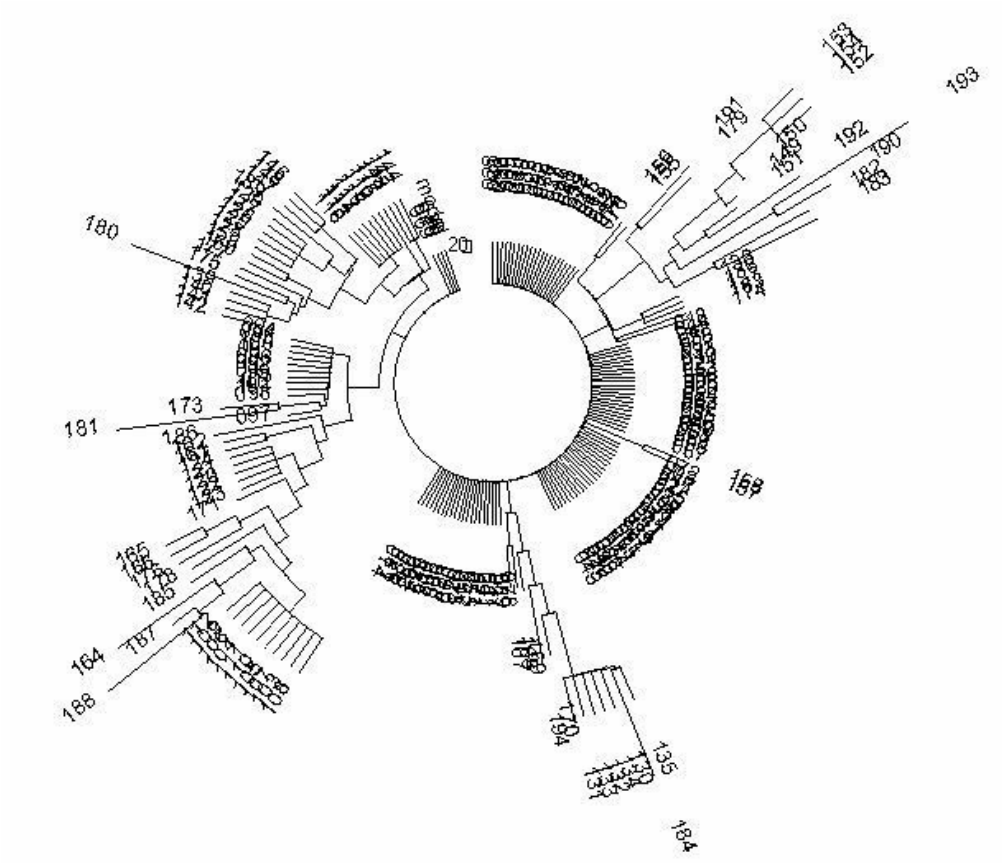

Figure 1. The 6-marker haplotype tree for 194 Jewish haplotypes of haplogroup J*. A "scientific" set (Behar et al, 2003)

There is a principal question – do the both data sets (J* and J1) reflect the same features in the haplotypes, or they are largely unrelated to each other?

The "scientific" set contains 91 base/modal haplotypes (14-16-23-10-11-12) out of 194, that is 47%. The "commercial" set contains 41 same base haplotypes (14-16-23-10-11-12) out of 90, that is 46%. The numbers are practically equal to each other. Both haplotypes are so-called Cohen Modal Haplotypes.

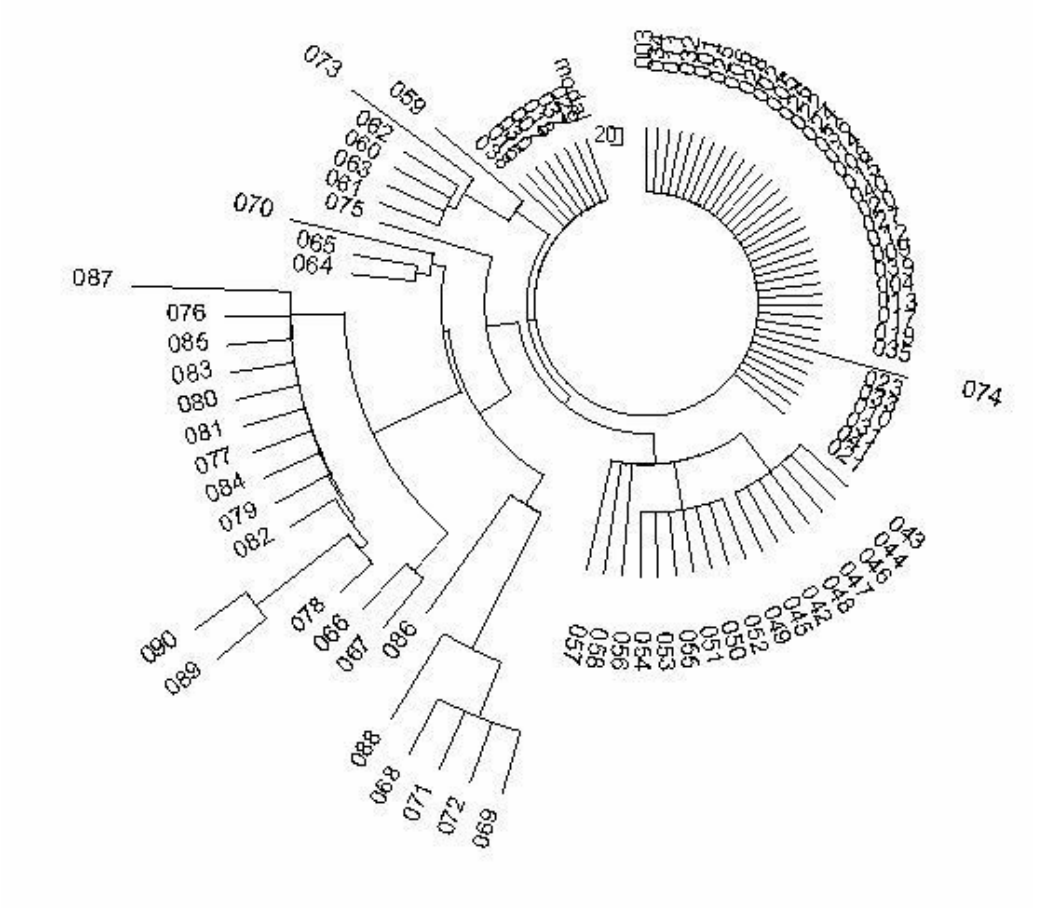

Figure 2. The 6-marker haplotype tree for 90 Jewish haplotypes of haplogroup J1. A "commercial" set, extracted from YSearch database

With respect to these base haplotypes, the "scientific" set contains 263 mutations in all 194 haplotypes, that is 1.36 mutations per haplotype. The "commercial" set contains 123 mutations in all 90 haplotypes, that is 1.37 mutations per haplotype. The numbers are again practically equal to each other.

In both cases the features are practically the same. Both data sets represent the same haplotypes, in terms of statistics. In is rather impressive, since the "scientific" set was collected at the end of 1990-s – beginning of the 2000-s with all precautions of a scientific choice of haplotype sources, while the "commercial" set was formed mainly in 2005-2008 from samples provided voluntarily and quite randomly.

The linear model based on a total number of mutations in its simplified (and generally incorrect, since there are typically more than one common ancestor for a set) way gives the following number of generations to a common ancestor: for the "scientific" set, 263/194/0.0096 = 141 generations, for the "commercial" set it gives the same figure: 123/90/0.0096 = 142 generations. However, a verification of the figures using our principal criterion, that is with a fraction of base haplotypes in the whole set, gives ln(194/91)/0.0096 = 79 generations ("scientific set") and ln (90/41)/0.0096 = 82 generations ("commercial set"). There is a large discrepancy between the both ways of calculations. In other words, one cannot calculate a time span to a common ancestor based just on a number of mutations in either J or J1 set of haplotypes, taken altogether, without their separation to discreet lineages. Therefore, a well-known dating of the coalescence of the Cohen haplotypes (or chromosomes) to the times between the Exodus and the destruction of the first Temple in 586 BC (Thomas et al, 1998), or to about 3300 years BP, appears to be poorly justified or even incorrect due to a heterogeneous population of Jewish haplotypes, which belong to different lineages, both ancient and rather recent ones.

That the population is a heterogeneous one is clearly seen from both Fig. 1 and 2. The both trees represent a combination of numerous "recent ancestral" haplotypes, sitting at the base of the trees, on the one hand, and ancient haplotypes, shooting away from the trees in a form of protuberances, on the other. In a case of a "smooth", "clean", single-ancestor haplotype tree, there should be either numerous base haplotypes and no protuberances, or in an opposite case, many "far away" haplotypes and very few (in any) base haplotypes.

After about 140 generations, as it follows from the number of mutations in J and J1 data sets, the amount of the base haplotypes should be not 46-47%, that is almost half of the total, but no more than one quarter of it. That is, if the set with 194 haplotypes was originated from one ancestor of 140 generations BP, it should have contained not 91, but only 50 base haplotypes. The set of 90 haplotypes should have contained not 41, but merely 23 base haplotypes.

Where the extra 41 base haplotypes (on top of 50) in the first data set, and 18 ones (on top of 23) in the second data set came from?

They came from a relatively recent ancestor(s), with the same "Cohen Modal Haplotype".

When the recent ancestor lived? Was there any earlier, ancient ancestor with the same "Cohen Modal Haplotype"?

Actually, there were two of them, and both lived in the Diaspora times. Since both of them belonged to the same J1 haplogroup, there obviously was a common ancestor for both of them. He lived 4,300±500 years BP, as it is shown later using more extended haplotypes.

Unfortunately, the "scientific" set of haplogroup J based on which the tree in Fig. 1 was constructed, was restricted by only 6-marker haplotypes (Behar et al, 2003). The literature contains only one set of 10-marker haplotypes of haplogroup J (Behar et al, 2004), however, it is smaller (85 haplotypes), and also shows a heterogeneous composition with respect to a common ancestor. It is a bi-modal one, with two modal haplotypes

12-23-14-10-X-Y-11-16-12-13-11-30
12-**25**-14-10-X-Y-11-16-**10**-13-**13-29**

in 14 and 13 repeats, respectively, out of those 85 haplotypes. The first one is the "Cohen Modal Haplotype" (CMH), and in a 6-marker format it looks like

14-16-23-10-11-12

The second is not the CMH at all, and in a 6-marker format it is

14-16-25-10-13-12

These two base haplotypes differ by seven mutations in the 10-marker format (shown above in bold), which approximately corresponds to 540 generations, or about 13,500 years (corrected for back mutations) between them. In fact, the difference is even greater, since in the 12-marker format their X-Y alleles are equal to 13-15 and 16-18, respectively, as it will be shown below, and they differ from each other by 13 mutations in their 12-marker haplotypes. Overall, in 12-, 25, and 37-marker haplotypes these two differ by about 0.9 mutations per marker, that correspond to 14,400 years between them. In other words, these two haplotypes have quite separate origins well before the historic times. Both of the above base haplotypes happen to come from rather recent and different ancestors

who lived in the first millennium AD, well into the Jewish Diaspora, as it will be shown below.

Let us consider both "scientific" and "commercial" sets of haplotypes and determine when common ancestors of the Jews of haplogroups J and J1 lived.

An ancient ancestor of about 11,100±600 years BP in haplogroup J

6-marker haplotypes (a "scientific" set)

The 6-marker haplotype tree (Fig. 1) shows four principal features: four obviously old branches, and 91 identical haplotypes around the "trunk" of the tree. These 91 "base" haplotypes are clearly recent in their origin, since there are too many of them to retain their original, ancestral, non-mutated form from old times. They very likely refer to a common ancestor who lived in the AD times. Also, as it will be shown below, this "base" 6-marker haplotype is a result of superposition of two rather recent haplotypes, in which by peculiarities of statistics these six markers are the same:

14-16-23-10-11-12

They collectively have been coined the "Cohen Modal Haplotype". Hence, the 6-haplotype tree is largely distorted, at least regarding the "base" haplotypes.

The most ancient branch, that is one that is coming from the oldest ancestor, is seen in the upper right-hand side in Fig. 1. The base (ancestral) haplotype deduced from the 20-haplotype "protuberance" is

14-16-24-10-11-12

The whole branch contains 59 mutations away from the base haplotype, and corresponds to 11,100±600 years to a common ancestor.

Despite a similarity between the most ancient 6-marker base haplotype and the CMH base haplotype (there is only one mutation difference between them), the 20 haplotypes of the "protuberance" branch differ by as many as 64 mutations with the 6-marker CMH. This corresponds to 500 generations of the mutational difference between present-day descendants of the ancient haplotype and the base CMH, that is about 12,500 years. This is indeed the sum of the time spans from the present time to the ancient ancestor (11,100±600 years BP), and to the recent CMH ancestor (1,350±100 years BP, see below). Hence, the CMH did not

descend from the above ancient haplotype. Otherwise a time span between them would be 9,750±700 years, not 12,500 years.

Indeed, as will be shown below, the most ancient Jewish ancestor in haplogroup J1 lived around 15,500±4,100 years BP.

Ancient Jewish ancestors between 3,500 and 10,000 years BP in haplogroup J

Other three old branches in Fig. 1 have the following base haplotypes:

15-15-23-10-11-12
14-15-24-10-11-12
15-16-24-9-11-12

The wide branch on the left-hand side in Fig. 1 contains two sub-branches, with base haplotypes first and second among the three above. One, in the lower left hand-side, contains 40 haplotypes and 76 mutations, that translates to 6,200±400 years to a common ancestor. Another, in the upper left corner, contains 29 haplotypes and 47 mutations, which translates to 5,100±500 years to a common ancestor. Hence, they both are derived from a common ancestor who lived 9,000±1,000 years BP. This common ancestor was identified directly from a haplotype tree, and it was estimated that he lived 9,200±1,200 years BP (see below and in the Table).

The narrow 12-haplotype branch on the lower right-hand side in Fig. 1 gives 3,500±200 years to a common ancestor. Its base haplotype is the third one above. Judging from differences in 3 alleles of this haplotype and each of the two others, which corresponds to 11,400 years of mutational evolution, they all derived from a common ancestor who lived 10,300±600 years BP. He might have been the same as 14-16-24-10-11-12 of 11,600±600 year old common ancestor, which has been identified above.

It should be noticed here that all estimates of this kind are very tentative with 6-marker haplotypes, and should be verified with more extended haplotypes, preferably with 25- and 37-marker haplotypes.

Overall, as it is shown below, the age of the Jewish lineages in haplogroup J1 to the present time is about 15,500±4,100 years.

Rather recent ancestors (in the Diaspora) in haplogroup J

6-marker haplotypes

Rather recent ancestors are recognized on haplotype trees as commonly flat branches, carrying series of identical haplotypes. Those haplotypes still preserve the structure of the ancestral haplotype. They are often surrounded on the branches with mutated forms of the same haplotype.

For example, the 6-marker haplotype tree in Fig. 1 reveals as many as nine of rather recent common ancestors. Their "base" haplotypes which are repeated at least three times are:

14-16-23-10-11-12	(91)
15-16-23-10-11-12	(11)
15-15-24-10-11-12	(10)
14-15-23-10-11-12	(7)
15-16-24-9-11-12	(6)
15-15-23-10-11-12	(5)
14-15-25-10-11-12	(5)
14-13-24-10-11-12	(4)
14-16-25-10-13-12	(3)

Typically, the more recent the common ancestor and more numerous his direct descendants nowadays, the higher the number in parentheses above. Those series of identical haplotypes typically sit on more extended branches along with their mutated forms, unless they are so numerous that form their own branches. For a number of those 6-marker haplotypes above, a time span to their common ancestors was determined using more extended haplotypes. For example, the CMH (91 identical haplotypes in Fig. 1) descended from two different ancestors, of 1,350±100 and 1,650±350 years BP; the 7-times repeat descended from a common ancestor of 1,800±200 years BP; the 3-times repeat descended from a common ancestor who lived 1,600±200 years BP, etc.

Principally the same base haplotypes for recent ancestors can be seen in another "scientific" set for Jewish haplotypes of haplogroup J (Fig. 3), which was published (Behar et al, 2004) a year later compared with data (Behar et al, 2003) in Fig. 1.

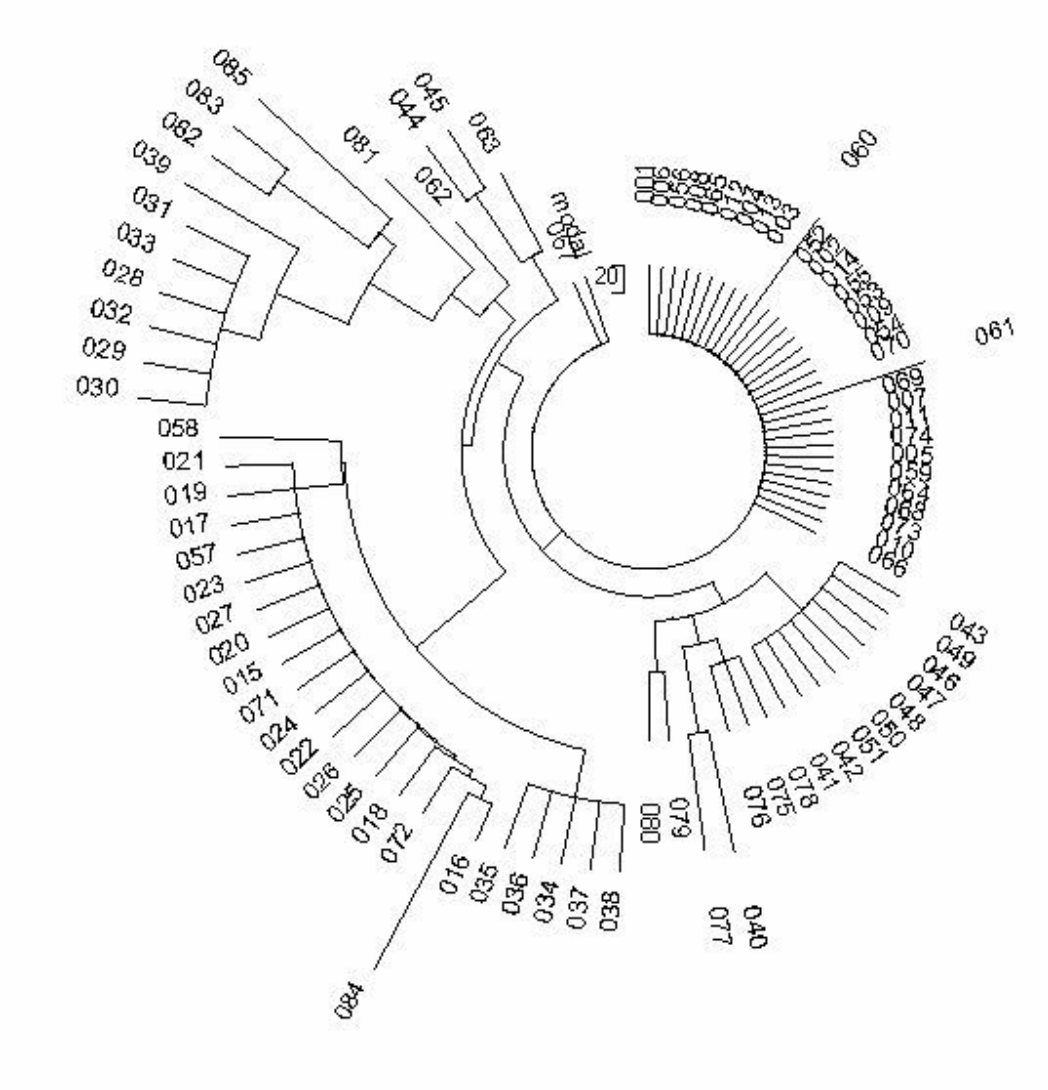

Figure 3. The 6-marker haplotype tree for 85 Jewish haplotypes of haplogroup J. A "scientific" set (Behar et al, 2004)

Five base haplotypes for rather recent ancestors in Fig. 3 are:

14-16-23-10-11-12 (29)
14-16-25-10-13-12 (16)
14-16-24-10-13-12 (5)
14-15-23-10-11-12 (9)
13-15-24-10-11-12 (6)

The first one is the CMH. The second is a wide branch, on the lower left-hand side in Fig. 3, derived from a common ancestor who lived 1,600±200 years BP, around the 5th century AD, as determined from 25- and 37-marker haplotypes (see below). The third haplotype is a part of the same branch. The fourth is a part of a branch with a common ancestor of 1,800±200 years BP, around the 3rd century AD. The fifth branch of haplotypes descended from a common ancestor who lived only 510±50 years BP; the branch is located on a tip of a branch (in 12- and 37-marker format) which is derived from a common ancestor who lived 9,200±1,200 years BP.

The tree in Fig. 3 does not contain ancient "protuberance" branches which would trace back to the ancestor of 11,100±600 years BP, who was present in another set of Jewish haplotypes published by the same author (Behar et al, 2003). However, the first (CMH) and the second haplotypes differ by four mutations in 6-marker haplotypes, which brings their respective ancestors 18,000 years apart, while the ancestors themselves lived only 1,350±100 and 1,600±200 years ago. This indicates that their common ancestor lived about 10,500±400 years BP. More detailed consideration of the same but extended haplotypes, in the 25- and 37-marker format, has shown that a distance between these two haplotypes is 14,400±2,600 years ("recent CMH") and 17,700±4,300 years ("older CMH"), and their common ancestor lived about 9,000±1,500 and 10,500±2,500 years BP, respectively.

These figures are rather close to those for the most ancient common ancestor of the Jewish lineages in haplogroup J1, who was estimated to live 15,500±4,100 years BP.

10-marker haplotypes

The 10-marker haplotypes made the recent ancestor branches even more pronounced (Fig. 4).

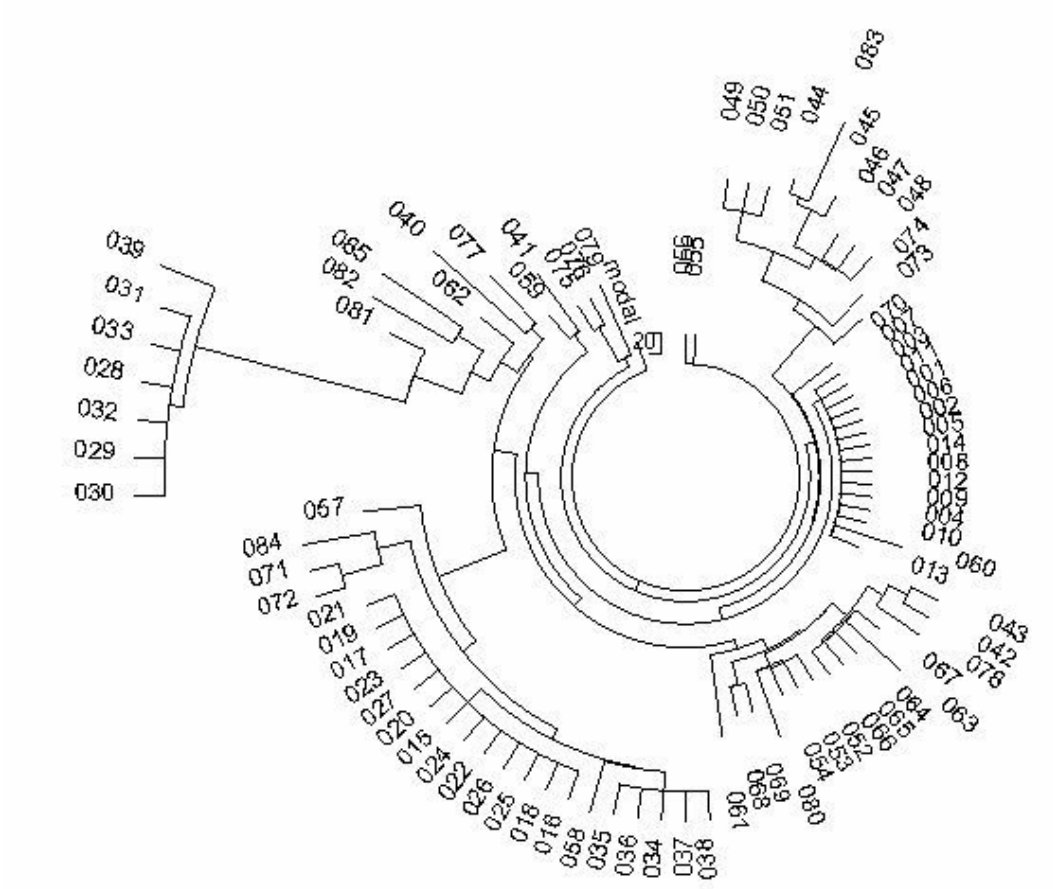

Figure 4. The 10-marker haplotype tree for 85 Jewish haplotypes of haplogroup J1. A "scientific" set (Behar et al, 2004)

The flat branch on the far left has only one mutation in all seven haplotypes. This formally corresponds to 6 generations to a common ancestors. A calculation using a share of the base haplotypes on the flat branch gives ln(7/6)/0.025 = 6 generations, that is exactly the same. The base haplotype in the 6- and 12-marker format is

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

where X-Y are DYS#385a,b missed in publication (Behar et al, 2004). They are equal to 12 and 19, as it is shown below, considering 12-marker haplotypes.

The most numerous haplotype in the whole set is again the Cohen Modal Haplotype, which is

12-23-14-10-X-Y-11-16-12-13-11-30
14-16-23-10-11-12

There are 14 identical CMH's in the haplotype set. It is the "recent Cohan Modal Haplotype".

The mutational difference (7 mutations) between the above two recent 10-marker ancestral haplotypes corresponds to is 9,700 years. This means that their common ancestor lived 5,600±200 years BP.

Another flat branch, in the lower left-hand side corner in Fig. 4, contains 13 identical haplotypes

12-25-14-10-X-Y-11-16-10-13-13-29
14-16-25-10-13-12

of 23 total on the branch, and contains 12 mutations. A common ancestor of those 23 individuals lived ln(23/13)/0.025 = 23 generations or 12/23/0.025 = 21 generations, that is 550±50 years BP. A practical fit between the two numbers of generations tells that the branch represents descendants from a single common ancestor.

One more branch in the lower right-hand side corner, containing 15 haplotypes, has a base haplotype

12-23-14-10-X-Y-11-16-11-13-11-31
14-16-23-10-11-12

This branch contains 17 mutations, that translates into 48±5 generations to a common ancestor. This is another CMH, the "older Cohen Modal Haplotype", with the same 6-marker haplotype as one shown above. The number of generations to the ancestor, refined with more extended haplotypes, turned out to be 66±14.

One can see that there are two "Cohen Modal Haplotypes" on the haplotype tree, with the same 6-marker haplotype, but with different 10-marker haplotypes:

12-23-14-10-X-Y-11-16-12-13-11-30
12-23-14-10-X-Y-11-16-11-13-11-31

Their X-Y alleles, missed in the publication (Behar et al, 2004), are equal to 13-15 and 13-17, respectively, as shown below using more extended haplotypes.

As one can see, the 6- and 10-marker haplotypes possess a rather low resolution capacity. Hence, we will examine more extended haplotype trees.

Ancient ancestor of 9,200±1,200 years BP in haplogroup J1

There is one clearly ancient ancestor branch which can be identified on the 12-, 25- and 37-haplotype trees. It is split between two extended branches on the right-hand side and bottom in Fig. 5 (12-marker haplotypes), it appears as the 8-haplotype branch in the lower left-hand side corner in Fig. 6 (25-marker haplotypes), and as the 4-hapotype branch in Fig. 7 (37-marker haplotypes). Its deduced base haplotype keeps varying when moves with extension of haplotypes, and looks as follows

14-16-24-10-11-12

12-24-14-10-12-17-11-16-13-13-11-30

12-24-14-10-12-17-11-16-13-13-11-30-18-8-9-11-11-26-14-20-29-14-15-16-16-10-9-20-22-15-13-17-18-32-36-12-10

A number of mutations in the first 12-marker panels and the 25-marker panel correspond to 402, 400, 392, 364 and 281 generations, that is on average 368±50 generations, or about 9,200±1,200 years to a common ancestor.

Figure 5. The 12-marker haplotype tree for 90 Jewish haplotypes of haplogroup J1. A "commercial" set, extracted from YSearch database

Other branches in the trees of Jewish haplotypes of haplogroup J1 belong to rather recent ancestors who lived in the Diaspora times, after the 70s years AD. Naturally, those ancestors had their direct ancestors, however, they established their own lineages after probably passing through genealogical bottleneck. That is why many Jewish lineages coalescent to rather recent common ancestors. However, "base" haplotypes of those "secondary" common ancestors can provide a lead to haplotypes of their ancient common ancestors.

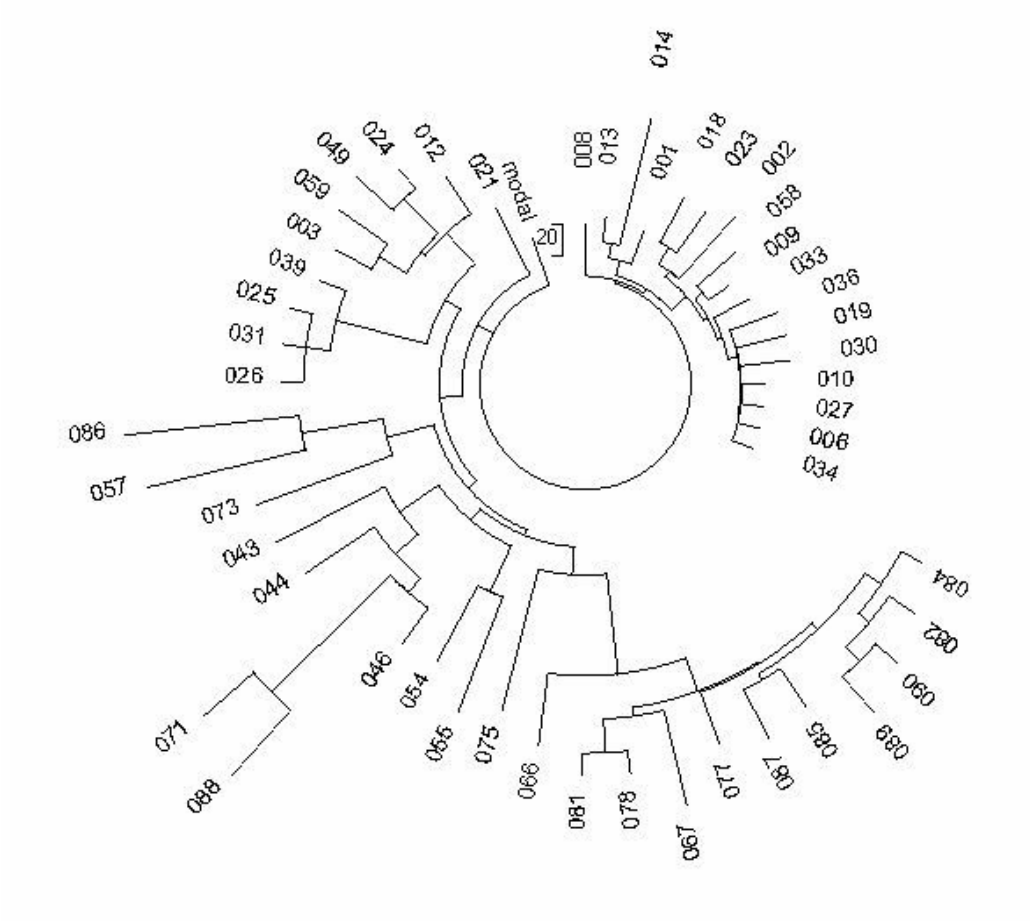

Figure 6. The 25-marker haplotype tree for 49 Jewish haplotypes of haplogroup J1. A "commercial" set, extracted from YSearch database

Rather recent ancestors (in the Diaspora) in haplogroup J1

Two "Cohen Modal Haplotype" ancestors, 1350±100 and 1650±350 years BP

6-marker haplotype trees in Figs. 1, 2 and 3 contain so-called "Cohen Modal Haplotype"

14-16-23-10-11-12

which sits on a "trunk" of the tree. It is typically the most abundant haplotype in Jewish J1 (and J*) haplotype sets. This very fact itself indicates that the CMH cannot be an ancient haplotype, otherwise it will be mutated and scattered along its haplotype branch.

This CMH starts splitting into its two principal forms, as Fig. 4 shows. The "recent CMH" haplotypes are located on the right-hand side of the haplotype tree, and the "older CMH" haplotypes moved to a lower right-hand side of the tree. It the 10-marker format the two CMHs differ by two mutations.

The two CMH's look quite different in their age (that is, their distance from the respective common ancestors) in Fig. 4. Indeed, the rCMH presents 14 identical haplotypes, while the oCMH present a rather mutated branch. However, it is deceiving impression, since many mutations in the rCMHs happened in their markers located further on along the extended haplotype, while in the oCMH mutations occurred in the 10-markers shown in Fig. 4. Hence, a "smooth" branch for the rCMH and a "wavy" branch for the oCMH in Fig. 4.

In their 12-marker format the two CMHs are separated by four mutations:

12-23-14-10-13-**15-**11-16-**12-**13-11-**30**
12-23-14-10-13-**17-**11-16-**11-**13-11-**31**

As it is calculated below, their common ancestors lived 1,350±100 and 1,650±350 years BP, respectively.

A difference in four mutations in a 12-marker format corresponds to 200 generations of a difference between the two haplotypes, that is about 5 thousand years. This figure will be detailed below using 25- and 37-marker haplotypes, and it turned out to be 4,300±500 years.

On the 12-marker haplotype tree (Fig. 5) the rCMH is still sitting around the trunk, on top of the tree. There are 17 identical, base rCMH. The 9-haplotype oCMH branch is located in the lower left-hand side of the tree. Clearly, the rCMH is more numerous in this population. A preliminary estimate shows that these 9 oCMH haplotypes contain 12 mutations, which translates to 59 generations, or about 1,475 years to a common ancestor. This is just a preliminary evaluation, since mutations in 25- and 37-marker haplotypes are not considered here. However, it turned out that it is a pretty good estimate since an average value calculated using all the panels in the extended haplotypes is 1,650±350 years to a common ancestor.

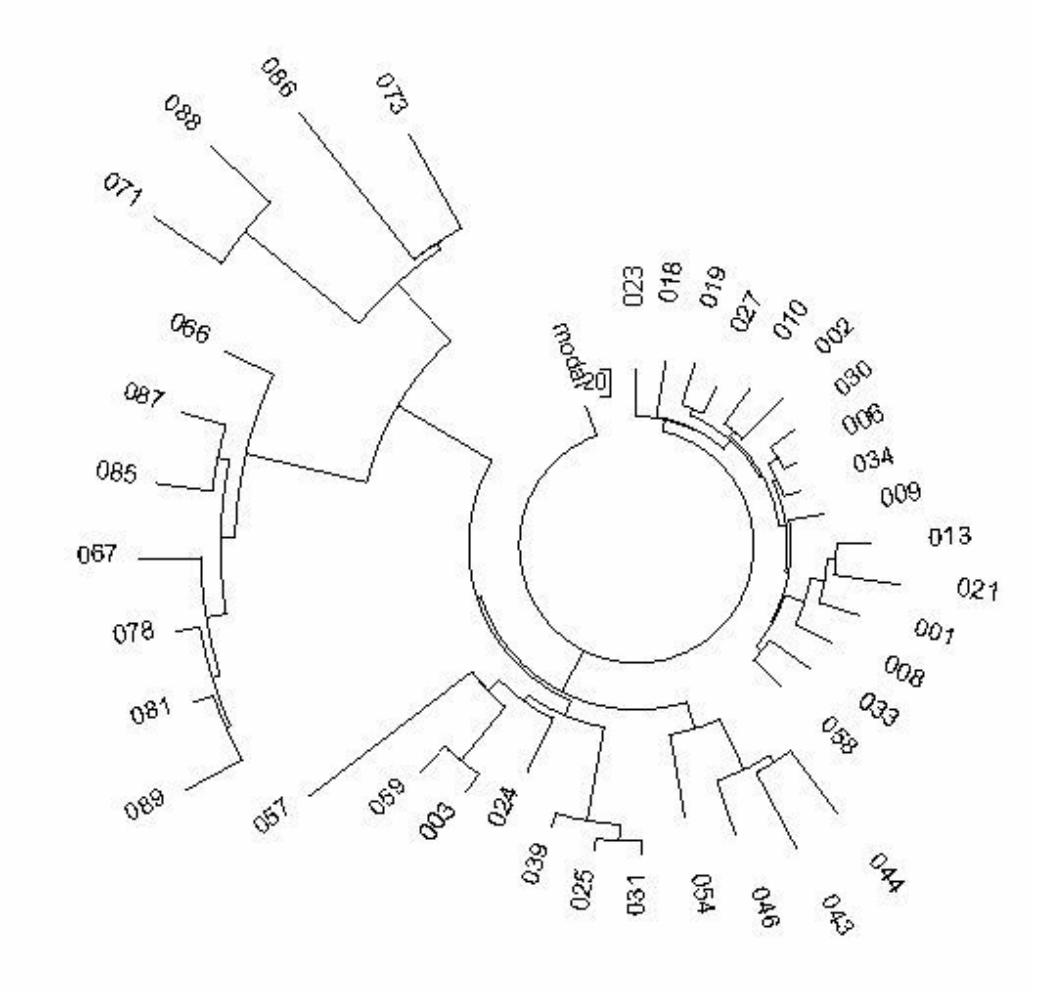

Figure 7. The 37-marker haplotype tree for 38 Jewish haplotypes of haplogroup J1. A "commercial" set, extracted from YSearch database

In the 25-marker haplotype tree (Fig. 6) the both CMH branches are located on both sides at the top of the tree, the "recent CMH" on the right, the "older CMH" on the left. As one can see, the effect of mutations on the two branches is rather similar, though oCMH (on the left) is certainly some older.

In the 37-marker haplotype tree the "recent CMH" 16-haplotype branch is on the right-hand side, and the 7-haplotype "older CMH" branch is located at the lower left-hand side. Now it is obvious that their time spans to respective common ancestors are rather similar, but the oCMH is a bit older.

A number of mutations in the rCMH branches in 25- and 37-marker haplotype trees gave 55, 55 and 52 generations, that is on average 54±2 generations, or 1,350±50 years to the common ancestor. If to take the figure literally (which is some stretch, of course), this gives 658±50 years AD, when a common ancestor for the "recent CMH" apparently lived. For the "older CMH" the time span to a common ancestor is 1,650±350 years BP, that is around 360±350 years AD. Both ancestors were, of course, different individuals. Some conjectures regarding who were they are given below.

The 37-marker haplotypes for the "recent" and the "older" Cohen haplotypes are listed in the Table. A difference between their haplotypes is 4 mutations in the 12-marker base haplotypes, 9 mutations in the 25-marker haplotypes, and 17 mutations in the 37-marker haplotypes. This corresponds to 200, 243 or 233 generations of mutational difference between the "recent" and "older" Cohen haplotypes, on average 225±22 generations, or 5,625±550 years. Taking into account the respective time spans to their common ancestor, it means that a common ancestor of the both "CMH"s lived 4,300±500 years BP.

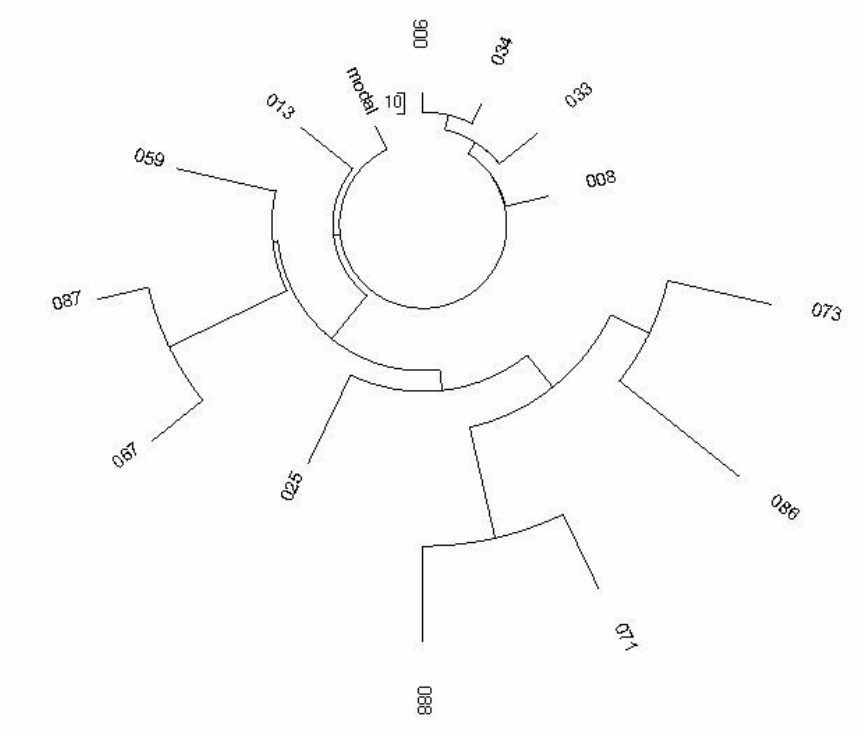

Figure 8. The 67-marker haplotype tree for 13 Jewish haplotypes of haplogroup J1. A "commercial" set, extracted from YSearch database

The 67-marker haplotype tree is shown in Fig. 8. It contains only 13 haplotypes, down from the initial 90 of 12-marker haplotypes. The tree hardly can be used for calculations due to insufficient statistics, however, it clearly shows four "recent Cohen haplotypes" in the upper part of the tree. Other eight haplotypes are descendant from old ancestors, including two "older" Cohen haplotypes (025 and 059),

Other rather recent ancestors (1,600±200; 1,800±200; and 510±50 years BP)

There are at least three more branches, patterns of mutations in which lead to rather recent ancestors. In the 12-marker format their base haplotypes are as follows

12-25-14-10-16-18-11-16-10-13-13-29

12-23-14-10-13-19-11-15-13-14-11-31

12-24-13-10-12-19-10-15-13-12-11-29

The first one is deduced from the 15-haplotype branch at the bottom of Fig. 5, as the 10-haplotype branch in the lower right-hand side area in Fig. 6, and as a 7-haplotype 37-marker flat branch on the left in Fig. 7. Patterns of mutations in all those extended haplotypes give 65, 60, 73, 77, 63, 66 and 50 generations, that is on average 65±9 generations, or about 1,600±200 years to a common ancestor.

The second one is obtained from the 9-haplotype branch located on the upper right-hand side in Fig. 5, moves in a truncated form to the lower left-hand side in Fig. 6, and forms a four-haplotype branch on the lower right-hand side in Fig. 7. Different panels of the extended haplotypes give for this branch 75, 73, 60 or 80 generations, that is on average 72±8 generations, or 1,800±200 years to a common ancestor.

The last base haplotype belongs to a flat 5-haplotype branch which sits on a top of a branch on the right-hand side in Fig. 5, two remaining haplotypes at the lower left-hand side in Fig. 6, and the same two haplotypes in the upper left-hand side in Fig. 7. Its common ancestor, who had a unique mutation in DYS#426, lived only 510±50 years BP, as will be described in detail below.

The earliest common ancestors of the Jewish lineages in haplogroup J1

Some base haplotypes of rather recent ancestors very significantly differ from each other, sometimes by as much as one mutation per marker on average. It means that a common ancestor of these base haplotypes lived some 10 thousand years BP or earlier, and estimation of these time spans can give us an age of the haplogroup itself, at least with respect to lineages of the present day descendants. The rather recent ancestors may have passed a population bottleneck, however, two or more of them still continue a series of lineages from their common ancestor.

The largest differences between ancestral (base) haplotypes in the Jewish lineages are observed between the "recent CMH" (a common ancestor of 54±4 generations BP) and a base haplotype with a unique mutation in the slow marker DYS#426 (see the next section, common ancestor of 20±2 generations BP). The mutational differences in their 12-, 25- and 37- marker haplotypes are 12, 26 and 42 mutations, respectively. These mutational differences are translated to 1160±320 generations between their base haplotypes, and to 15,500±4,100 years to their common ancestor. This apparently refers to times when the oldest Jewish lineage has appeared in haplogroup J1.

A very slow marker DYS#426 and a Jewish lineage associated with it. To the depth of haplogroup J1

Mutations in the marker DYS#426 are so infrequent that they are almost irreversible. In haplogroups of earlier origin, including C, D, E, F, G, H, I, J, K, L, M, N, O – a great majority of people have allele = 11 in DYS#426. Only in younger haplogroups, Q and R, a great majority of people have there an allele = 12. It mutates very seldom. For example, among 343 haplotypes in "haplogroup J1" section in YSearch database only 23 haplotypes contain a mutation from 11 to either 10 or to 12. Most of the individuals who bear that mutation are descendants of just one common ancestor, whose marker mutated a long time ago, - or recently, on that matter. Essentially, there are just two such common ancestors in haplogroup J1 who bear that mutation: one with an allele = 10, another with an allele = 12.

It turned out that an individual with DYS#426=10 has penetrated the Jewish community just 510±50 years ago, and the mutation was carried over among his direct descendants among the Jews in haplogroup J1, as will be shown in this section.

Regarding haplogroup J2, there are only 14 individuals among total of 1,302 with a mutated DYS#426. None of them is Jewish.

In general, a number of carriers of mutated marker DYS#426 (as well as of any marker, on that matter) in any given haplogroup depends just on how long ago the mutation has occurred, and how many male descendants have survived since then and gave male offspring. Hence, the number by itself can be any number. It does not carry any specifics except that the mutation has occurred some time ago.

A haplotype tree of all bearers of mutated DYS#426 in haplogroup J1 is shown in Fig. 9.

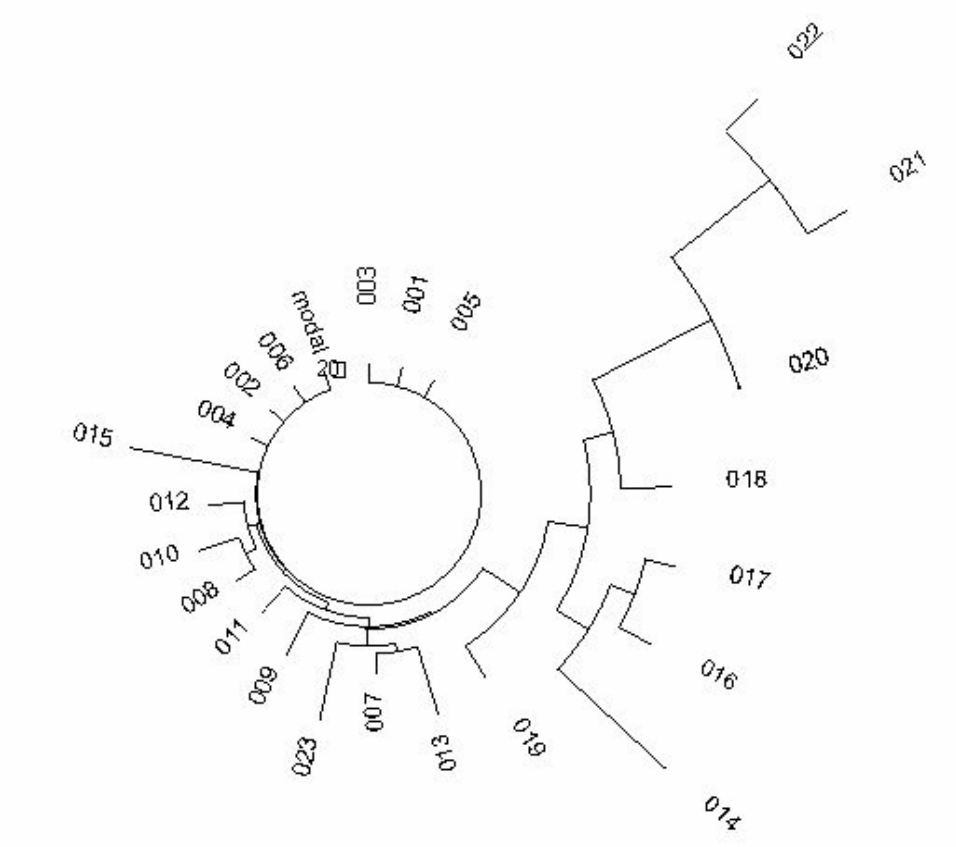

Figure 9. The 12-marker haplotype tree for 23 haplotypes (Jewish and non-Jewish) of haplogroup J1, carrying a mutation in DYS#426. A "commercial" set, extracted from YSearch database. The Jewish haplotypes No. 002, 003, 004, 005, 006, 008, 010, 011.

There are 8 Jewish individuals among those 23 whose haplotypes have a mutation in DYS#426. All of the Jewish haplotypes are located around the "trunk" of the tree, all have the allele = 10. All are clearly descended from a rather recent ancestor.

There are three common ancestors for haplotypes in Fig. 9. One is a common ancestor for all the eight Jewish individuals. A time span to his lifetime can be calculated by two different ways, as it was suggested in this article. One is to identify his base haplotype

12-24-13-10-12-19-10-15-13-12-11-29

which presents as five identical haplotypes among the eight total of his descendants, and then calculate ln(8/5)/0.024 = 20 generations to the common ancestor. Another way is to determine a total number of mutations in those 8 haplotypes with respect to the base haplotype. Since there are four mutations only, then 4/8/0.024 = 21 generations to the common ancestor. Since these figures are practically equal to each other, all the eight Jewish individuals indeed have a single ancestor, who lived 510±50 years BP, around the very end of the 15^{th} century.

Other 8 non-Jewish individuals (including one adoptee) in the set, who have the allele = 10 in DYS#426 and expectedly are descendants of another single ancestor, have the following base haplotype which differ by only one mutation (shown in bold), however, it is much older compared to the Jewish one:

12-24-13-10-12-19-10-15-**12**-12-11-29

All the eight haplotypes are mutated, and the base (ancestral) haplotype was deduced by minimizing on mutations. The eight haplotypes contain 24 mutations with respect to the base haplotype, which gives 3,600 years to their common ancestor.

Haplotypes of the remaining seven individuals which have the allele = 12 in DYS#426 form a distinct branch on the right-hand side in Fig. 9. Most of them have European names, along with a Palestinian individual among them (haplotype 014). Their deduced base haplotype

12-24-14-10-12-14-12-13-12-13-11-29

reveals as many as 46 mutations with respect to all seven haplotypes in the branch, which translates to 9,400 years to their common ancestor.

This ancient common ancestor of the non-Jewish individuals with alleles = 12 in DYS#426 was obviously not a direct ancestor of the Jewish individuals. First, the alleles 10 and 12 clearly require another, more ancient common ancestor, apparently with the allele = 11. Second, the difference between the base haplotypes with alleles 10 and 12 above is tremendous, and requires much longer time span then the 9,400 years. There are 12 mutations on 12 markers in their base haplotypes. This accounts for about 25,000 years of the mutations difference between two common ancestors. In turn, it gives an estimate of 17,500±1,000 years to the "total" common ancestor of all the 23 individuals, Jewish and non-Jewish, in haplogroup J1.

This is apparently close to the ultimate "depth" of haplogroup J1.

The "recent Cohen Haplotype" in haplogroup J1 in the Jewish population

Who was that individual who lived around the 7th century, successfully passed a genealogical bottleneck, and had initiated a "demographic explosion" of the Jews of haplogroup J1, carrying the "recent Cohen Modal Haplotype"? It is very likely that he had a prominent position in the Jewish society in his time, and he must be still well known nowadays. It seems plausible that he was Bustenai ben Hanina, aka Bostanai ha-David, the First Exilarch of the Third Dynasty, who lived (estimate) in 590-670 AD. His father, Haninai ha-David, was a direct descendant of King David (1037-967 BC), the second king of the united Kingdom of Israel. Bustenai was a leader of the Jewish community in Babylonia and the last holder of the *Resh Galusa* noble title, which means "Head of the Diaspora" in Aramaic and held by direct descendants of the House of King David. He had belonged to a noble line that was always preserved in Jewish history. Bustenai used his influence to strengthen Jewish life throughout Babylonia.

With two wives, Izdundad, Princess of Persia, and Princess Adoa ha-David, Bustenai had three sons, Hisdai ha-David, Haninai ben Bustanai, and Shahrijar, who had followed with three grandchildren of Bustenai. The number of offspring then quickly multiplied, and two centuries later there were dukes, counts and barons in France, Italy, Spain. A number of his descendants went on to become *rashei galusa* themselves (Eisen, 2004). In only three-to-four generations after Bustenai, some of his grandchildren and grand-grandchildren had arrived to Spain and Portugal, after the Muslim expansion in the 8th century, and brought there the Bustenai haplotype that later was coined the "Cohen Modal Haplotype". Soon,

less than a century after it, by the end of the 8th century, the Jews of Spain and Portugal have becoming known as Sephardim.

The above hypothesis might explain an outburst of the Bustenai haplotype aka rCMH (though Bustenai was not a Cohen, he was from the tribe of Judah), starting from the 7th century AD, and why Sephardim and Ashkenazim have this haplotype in about equal proportions, despite many centuries of division of these Jewish populations. According to (Behar et al, 2003), 163 Jews from 399 named themselves Sephardim in the survey, and 69 identified themselves as Cohens. This is 17% of the total, which is close to 20-40% estimate of the "Older CMH" in the Jews of haplogroup J1 (see above). 36 of the Cohens, that is 52%, had the 6-marker CMH. In Ashkenazim, 47% of the Cohens had this haplotype (Behar et al, 2003). If to consider not only the exact CMH but also its mutated variants (within 5 mutations in the 6-marker CMH), then 92% of Ashkenazim and 75% of Sephardim bear the mutated CMH. As one can see, Ashkenazim and Sephardim split the CMH almost equally. Furthermore, around 10% of lay Israelites in both Ashkenazim and Sephardim groups carry the exact CMH; many of them apparently are descendants of King David. 37% and 38% of them respectively bear mutated variants of the CMH [data are calculated from (Behar et al, 2003].

It seems that since descendants of Bustenai have gotten into the both main groups of Jewish people, Ashkenazim and Sephardim, his haplotype proliferated equally well throughout the Jewish community even after their separation after the 8th-9th century.

Due to a superposition of the "recent CMH" and the "older CMH", three quarters of the Jews in haplogroups J and J1 have the 6-marker "CMH", which is indistinguishable from each other in its non-mutated short format.

A distribution of descendants of the "recent" and "older" CMHs in a current Jewish population in haplogroup J1 can be also illustrated as follows. In a search for matches of the "older CMH" within the entire the YSearch database, only two exact 12-marker haplotypes in haplogroup J1 were found. For the "recent CMH", as many as 36 exact matches were found in haplogroup J1.

It is rather striking that almost half of the Jews in haplogroup J1 are apparently direct descendants of King David (allegedly through their recent ancestor Bustenai who lived about 1350 years ago) and bear the "classic" 6-marker CMH, and many of them have intact 12-marker CMH in its base form. From about 90 individuals carrying Jewish names (themselves and/or their not so distant great-grandfathers) in haplogroup J1, 17 have the exact 12-marker (allegedly) Bustenai haplotype, and 19 more have it is a slightly mutated form (one or two mutations

per the 12-marker haplotype). In the same 90-haplotype set, six individuals carry the "older Cohen haplotype" in its exact form, and nine more have it is a slightly mutated variant (one or two mutations per the 12-marker haplotype).

A possible origin of the "older CMH" will be discussed below in a section "The Cohen Haplotypes" in haplogroups J1 and J2 in the Jewish population". Briefly, the mystery of the "cohen" part in the CMH can be resolved as follows. Clearly, if the "recent CMH" was indeed proliferated initially by Bustenai, a descendant of King David himself, a bearer of the "12 tribes haplotype" and presumably belonged to the Judah tribe, if to follow the biblical story, then his descendants were not the cohens. However, it cannot be denied that the cohens form a significant (or noticeable) part of bearers of the CMH. That simply means that the ancestor of the "older CMH" was a cohen himself, and belonged to the Levy tribe.

A common ancestor of both rCMH and oCMH lived 4,300±500 years BP, in the Abraham times, if to follow the biblical story. According to the Torah and its interpretations by the scholars, the patriarchs of the twelve tribes of Israel were born between 3575 and 3560 years BP, and Abraham himself lived three generations earlier.

The Jews of Haplogroup J2

Jewish haplotypes of haplogroup J2 are noticeably younger (in terms of time spans to their common ancestors) compared to those of haplogroup J1. This can be estimated by different means, all of them are rather approximate. However, they give rather consistent results.

It was shown above that ancient ancestors in haplogroup J1 could be traced down to 17,500±1,000 years BP, and those of the Jews in haplogroup J1 to 15,500±4,100 years BP. However, the oldest Jewish ancestor in haplogroup J2 was found to live only at 6,000±1,000 years BP (see below).

At the same time haplogroup J2 itself, which includes both Jewish and non-Jewish haplotypes, is certainly much older than six thousand years old. All 1302 twelve-marker haplogroups in "Haplogroup J2" section of YSearch database (including subgroups) contain 7452 mutations with respect to the "mutation minimized haplotype"

that is on average 48 mutations per 100 markers. Among 12-, 25- and 37-marker haplotypes this average figure is equal to 55±6 mutations per 100 markers, which corresponds to 360±50 generations, that is 9,000±1,300 years. This is the lowest limit of the figure, since many of those haplotypes belong to descendants of rather recent ancestors. Therefore, it is very likely that earliest Jewish ancestors in haplogroup J2 had appeared in the haplogroup relatively late compared with age of the haplogroup itself.

Before considering in detail the Jewish haplotypes of haplogroup J2, let us briefly outline the principal findings of this section.

- The oldest lineage of the present day Jews in haplogroup J2 was initiated by a common ancestor who lived 6,000±1,000 years BP. This is much more recent compared with 15,500±4,100 years BP for a common ancestor of the oldest Jewish lineage in haplogroup J1.
- Another ancient common ancestor of the Jews in haplogroup J2 lived 3,100±200 years BP, and he was a direct descendant of the most ancient common ancestor who lived 6,000±1,000 years BP.
- The so-called "Cohen Modal Haplotype" in haplogroup J2 is 1,200±200 year old, and its common ancestor lived around the 9th century AD. The two CMH's in haplogroup J1, that is the "recent" one and the "older" one, are separated from the CMH in haplogroup J2 by 12,600 and 9,100 years , respectively, and have nothing in common, except that THEIR common ancestor belonged to an upstream haplogroup J, and their six alleles (in "scientific" notation of the loci) by sheer chance happened to be the same.
- Besides the CMH, there were at least two more recent common ancestors of the Jews in haplogroup J2, both of them lived in times of the Diaspora, between 29 and 35 generations BP, or between 725 and 875 years BP. A quarter of the Jews of haplogroup J2, living today, are direct descendants of their common ancestor who lived 725±50 years BP. Their 6-marker haplotype is 14-15-22-10-11-12.
- The Jewish haplotypes of haplogroups J1 and J2 do not mix with non-Jewish haplotypes from the same haplogroups on the haplotype tree. In other words, the Jewish haplotypes have their own niche. This is a very distinct feature, compared with – typically – most (if not all) of other haplogroups, in which the Jewish haplotypes were acquired whom "outside", from the non-Jewish environment. Naturally, in those other cases Jewish haplotypes were not unique, and readily mixed with non-Jewish haplotypes on the trees.

- The Lemba haplotypes have nothing to do with ancient Jewish lineages, leaving alone that their haplogroups have not been even determined in the first place. The allegedly "CMH" haplotypes in Lemba are derived from ancestor(s) who lived just a few centuries ago. They might be of Arabic origin. Jewish "trace" cannot be excluded, however, in this case it should have penetrated the Lemba community some 300-400 years BP. There are no mutations in those haplotypes, which points out to their quite recent origin.

We will consider 6-marker haplotypes first. They typically are not good for calculations of time spans to common ancestors, however, they provide a good overall view on a simplified haplotype tree, particularly regarding rather recent ancestors. For recent ancestors, 6-marker haplotype trees commonly produce wide flat branches of identical (or almost identical) haplotypes, regardless how far from the tree "trunk" these flat branches are located. A far-away location in those cases (for flat branches) just means that the "recent base haplotype" is as many as several mutations away from the base haplotype of the tree, that is there is a large time gap (typically, thousands of years) between their respective common ancestors.

Commonly, true "ancient" haplotypes form a "ladder" (not flat branches) directed away from the "trunk". After a preliminary sorting out which haplotypes are produced directly from ancient ancestors, and which ones are from rather recent ancestors, I have used 12, 25-, and 37-marker haplotypes to more reliably calculate time spans from those ancestors.

6-marker haplotypes (a "scientific" and a "commercial" sets)

Again, here we have one more confirmation that "scientific" sets of haplotypes, published in the literature, on the one hand, and "commercial" set, extracted from YSearch database, on the other, produce basically the same data. As it was described earlier in this paper, 194 of 6-marker Jewish haplotypes of an upstream haplogroup J (Behar et al, 2003) contained 263 mutations, that is 1.36 mutations per haplotype, or 22.7 mutations per 100 markers. A set of 90 of 6-marker Jewish haplotypes of haplogroup J1 (YSearch) contains 123 mutations, that is 1.37 mutations per haplotype, or 22.8 mutations per 100 markers.

In the same vein, a set of 88 haplotypes of haplogroup J2, published in (Behar et al, 2004), contains 160 mutations, that is 1.82 mutations per haplotype (30.3 mutations per 100 markers), and 75 haplotypes of haplogroup J2 (YSearch) contains 135 mutations, that is 1.80 mutations per haplotype (30.0 mutations per

100 markers). The respective haplotype trees (Figs. 10 and 11) are pretty much similar. Both contain branches of short 6-allele sequences

14-16-23-10-11-12

at the top of the tree, so-called the "Cohen Modal Haplotypes". Besides the J2 CMH, both trees contain at least three more of obviously recent series of haplotypes (flat branches of identical haplotypes) on the right- and left-hand sides and at the bottom of the tree.

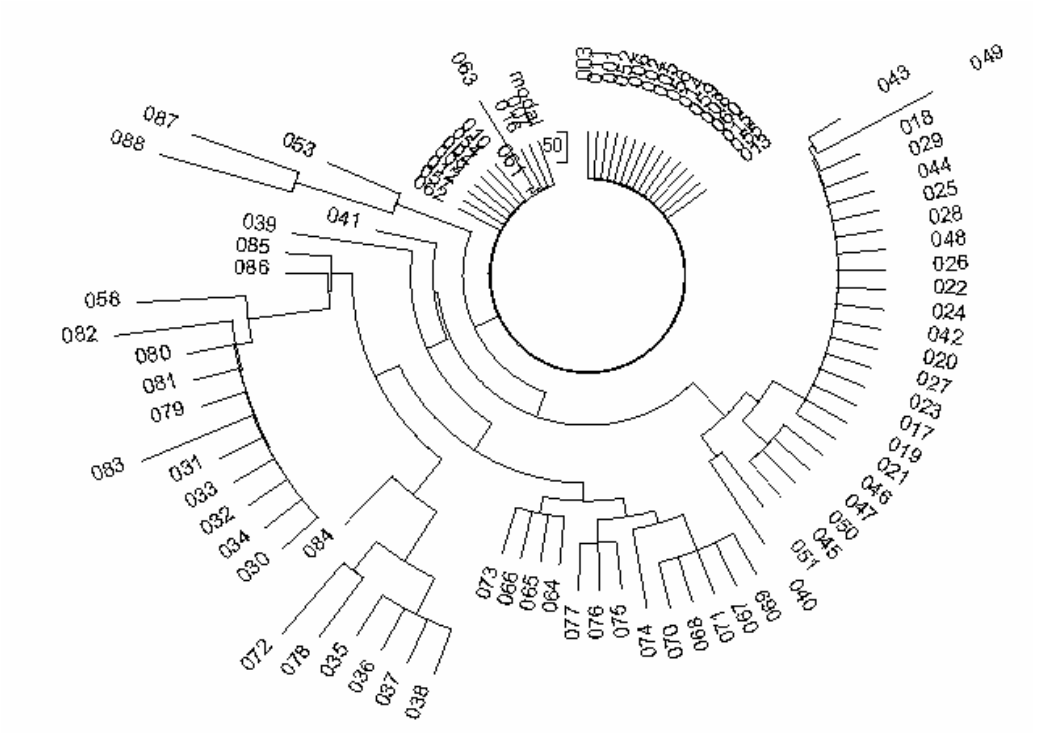

Figure 10. The 6-marker haplotype tree for 88 Jewish haplotypes of haplogroup J2. A "scientific" set (Behar et al, 2004)

Fig. 10 (a "scientific" set of 88 haplotypes) shows 25 identical "Cohen modal haplotypes" on the top, that is 28% of the total. In J1 haplogroup these haplotypes were a mix of the "recent" and "older" 6-marker CMH, hence, there were more of them on the tree (41 from the total number of 90, that is 46%). In J2 haplogroup they present in a smaller amount, though the age of the recent J1 and J2 CMH is

very close to each other, 54 and 48 generations, 1350±100 and 1200±200 years BP.

A largely flat 24-haplotype branch on the right-hand side of the tree (Fig. 10) contains 17 identical haplotypes

14-15-22-10-11-12

which are clearly descended from a recent ancestor, otherwise they would have been in a lesser number due to mutations. These 24 haplotypes contain only 12 mutations, that is 8.3 mutations per 100 markers. This corresponds to 55 generations to a common ancestor. Regardless of a number of mutations, 17 identical (base for this branch) haplotypes out of 24 on the branch point out at ln(24/17)/0.0096 = 36 generations to a common ancestor. A better accuracy is hard to expect from 6-marker haplotypes. As it will be shown below, an estimate with 37-marker haplotypes gave 29±2 generations to a common ancestor of this branch. This recent common ancestor for the branch lived around 725±50 years BP.

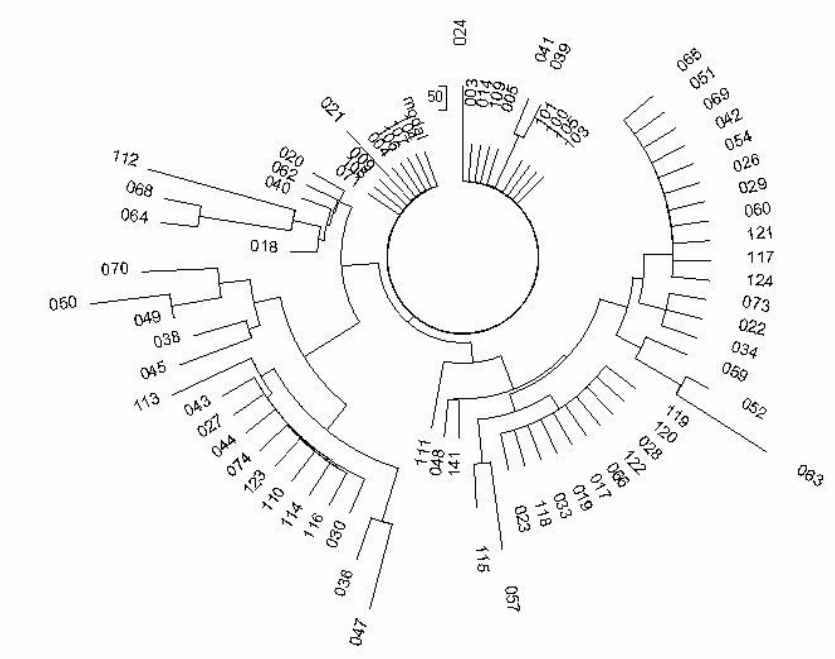

Figure 11. The 6-marker haplotype tree for 75 Jewish haplotypes of haplogroup J2. A "commercial" set, extracted from YSearch database

The “commercial” (extracted from YSearch database) set of 75 haplotypes in Fig. 11 contains practically the same branch on the right-hand side with the same base haplotype

14-15-22-10-11-12

One more ancestor produced a series of 8 identical (base) haplotypes

15-16-23-9-11-12

in the 11-haplotype branch on the left-hand side in Fig. 10. These 11 haplotypes contain only 4 mutations. It turned out that the same but extended haplotype gave many more mutations in the 12-, 25, and 37-marker panels, and it was shown that the common ancestor lived 3,100±200 years BP. This is exactly a case when 6-marker haplotypes are not reliable for estimations of time spans to a common ancestor.

Recent ancestors of these two base haplotypes, the CMH (on the top of the both trees) and that one on the right-hand side in Fig. 10 (14-15-22-10-11-12) lived about 500 generations apart, as follows from comparison of their 12-marker haplotypes (8 mutation difference between the two, see the Table). This difference corresponds to about 12,500 years between them. That is why the two recent haplotypes are located on the haplotype tree so far from each other. They are practically unrelated within a timeframe of the Jewish history, though eventually they lead to a common ancestor within haplogroup J2, about 6,000 years BP, as it will be shown later in this section.

Yet another recent ancestor of the Jews in haplogroup J2 can be identified from a flat branch, with a base haplotype

15-15-23-10-11-12

This branch (at the bottom on the right-hand side of the tree, Fig. 11) is more developed in the “commercial” set (10 identical, base haplotypes of 12 in the branch) compared with the “scientific” set (four identical haplotypes, at the bottom, Fig. 10, as part of a more extended branch). The twelve haplotypes contain only three mutations, that is 4.2 mutations per 100 markers, and correspond to only 27 generations to a common ancestor. The 37-marker haplotypes showed 35±5 generations to the common ancestor (Table).

To sum it up, there are at least three recent common ancestors of the Jews in haplogroup J2, all of them lived in times of the Diaspora, all between 29 and 49 generations BP, or between about 700 and 1200 years ago:

14-16-23-10-11-12
14-15-22-10-11-12
15-15-23-10-11-12

In the "scientific" 6-haplotype set they account for 64 of 88 haplotypes, that is about three quarters of all the Jews in haplogroup J2. In fact, at least 73 of 88 haplotypes, that is more than 80% of Jews in haplogroup J2, are descendants of ancestors who lived in the Diaspora times. Clearly, Jews in haplogroup J2 are much "younger" in terms of a large share of their recent common ancestors compared to the Jews in haplogroup J1.

The remaining 15 haplotypes on the tree (Fig. 10) are derived from a deduced, ancient base haplotype

15-15(16)-23-10-11-12

from which they differ by an average value of 37 mutations per 100 markers. This points out at several thousand years to a common ancestor. More accurate calculations are given with the help of 37-marker haplotypes (the next section). 14 haplotypes in the "commercial" set (Fig. 11) allegedly descended from an ancient common ancestor. They are derived from the same deduced base haplotype

15-16-23-10-11-12

from which they differ by an average value of 44 mutations per 100 markers. This also points at several thousand years from a common ancestor. As it will be shown in the next section, refined (more extended) haplotypes give the value as 6,000±1,000 years BP.

10-marker haplotypes (a "scientific" set)

A set of 10-marker haplotypes of haplogroup J2 was published in (Behar et al, 2004), and the respective tree was constructed in Fig. 12. Principally, the 10-marker tree is close to the 6-marker haplotype tree (Figs. 10 and 11), and it contains 16 identical "Cohen Modal Haplotypes"

12-23-14-10-X-Y-11-16-11-13-11-30

at the top of the tree. X and Y are markers which have not been determined in (Behar et al, 2004), however, it is clear from the established 12-marker haplotypes (see below) that they are 13-17. One can see that it is the CMH of haplogroup J2 indeed. It was determined from 12- and 37-marker haplotypes as

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

in a 12-marker format, and

14-16-23-10-11-12

in a 6-marker format. The same "CMH" was identified below, using 12- and 37-marker haplotypes. This shows again that a "scientific" and "commercial" sets of haplotypes give principally the same results.

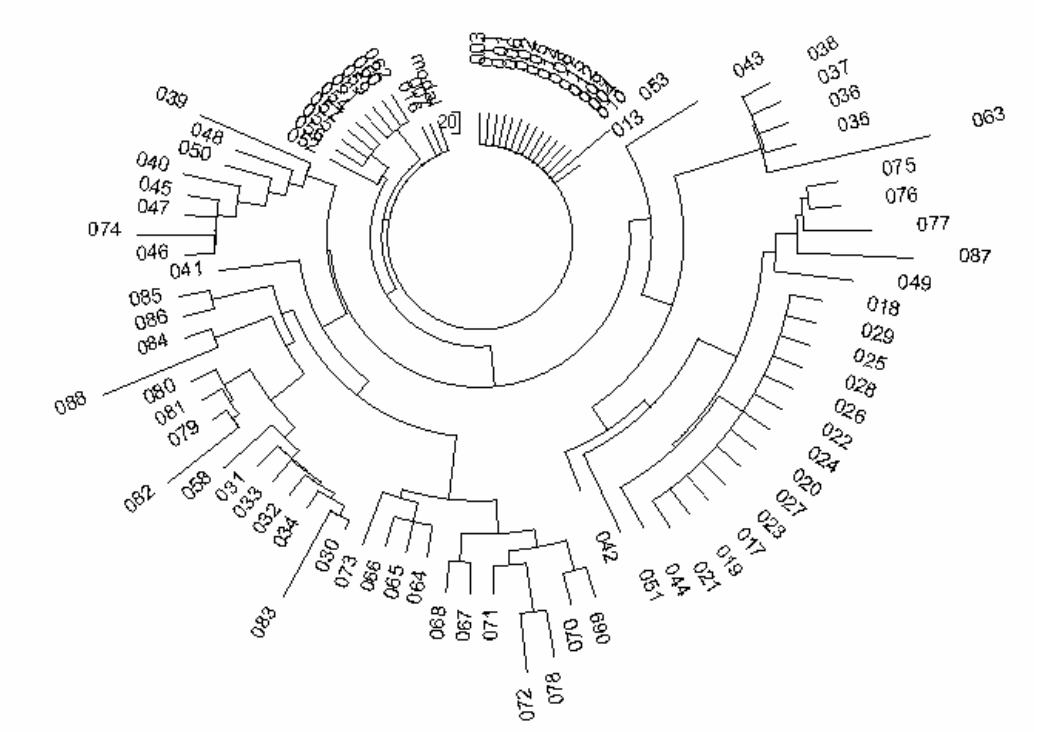

Figure 12. The 10-marker haplotype tree for 88 Jewish haplotypes of haplogroup J2. A "scientific" set (Behar et al, 2004)

In the same manner results obtained with the "commercial" sets of haplotypes help to reconstruct incomplete haplotypes of the "scientific" set. The haplotype, listed in a row of other haplotypes in (Behar et al, 2004)

12-22-14-10-X-Y-11-15-12-14-11-31

is in fact a base haplotype from a rather recent ancestor, who lived 900±50 years ago (Table). The above haplotype has the X-Y pair equal to 14-15, and its 6-marker haplotype is as follows:

14-15-22-10-11-12

The above 10-marker haplotype is a base for a 19-haplotype branch on the left in Fig. 12. This branch contains 20 mutations with respect to this base haplotype. This gives 10.5 mutations per 100 markers, which is close to 7.5, 10 and 11.9 mutations per 100 markers in 12-, 25- and 37-marker panels for the recent "Cohen modal" haplotype (see below), and gives 62 generations to a common ancestor. This figure will be refined below.

Also, a haplotype

12-23-15-10-X-Y-11-15-12-13-11-29

which was incomplete in (Behar et al, 2004), is in fact the ancient haplotype of 6,000 years old. The table lists it in a 37-marker format. It has the alleles X-Y equal to 14-16.

This 10-marker haplotype is positioned on the left-hand side in Fig. 12 and is a base for 35 haplotypes, which have 120 mutations from this base. This corresponds to 34 mutations on average per 100 markers. It shows a good fit with 38±6 mutations per 100 markers in haplotypes (an average from 12-, 25- and 37-marker panels) for the ancient branch identified below from "commercial" haplotype sets. It is the same branch as derived from a common ancestor who lived 235 generations, or 5,875 years BP, to be absurdly and unrealistically precise. These figures will be round up below to 6,000 years BP.

12-marker haplotypes (YSearch database)

The J2 "Cohen" haplotypes are located as a compact branch on the haplotype tree (Fig. 13) in the lower right-hand side area. On the 37-marker tree these haplotypes will move as a separate 10-haplotype branch at the upper right-hand side tree (Fig. 14).

The branch in Fig. 13 contains 18 haplotypes, between those marked 103 and 108, and has eight out of 10 of the CMH haplotypes on the 37-marker tree. This 18-haplotype 12-marker CMH branch contains 6 identical (base) haplotypes

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

which are exactly as those found in the 10-marker "scientific" set with X-Y equal to 13-17, and as the "CMH" on the 37-marker tree (Fig. 14). These 18 haplotypes contain 20 mutations from the above base CMH, that is 9.26 mutations per 100 markers, or 48 generations to a common ancestor. Six base haplotypes out of 18 on the tree give
ln (18/6)/0.024 = 46 generations to a common ancestor. This practically exact fit proves that the set of CMH haplotypes is "clean" in terms of an absence of other lineages, and indeed leads to a single (in terms of DNA genealogy) common ancestor. It nicely fits to 49±9 generations obtained for the same haplotypes on the 37-marker haplotype tree, as it is shown below, and it is not far away from the 48 generations obtained from 6-marker haplotypes, and from 62 generations obtained from the 10-marker haplotypes. It should be noted that the last two estimates were obtained from the same publication (Behar et al, 2004).

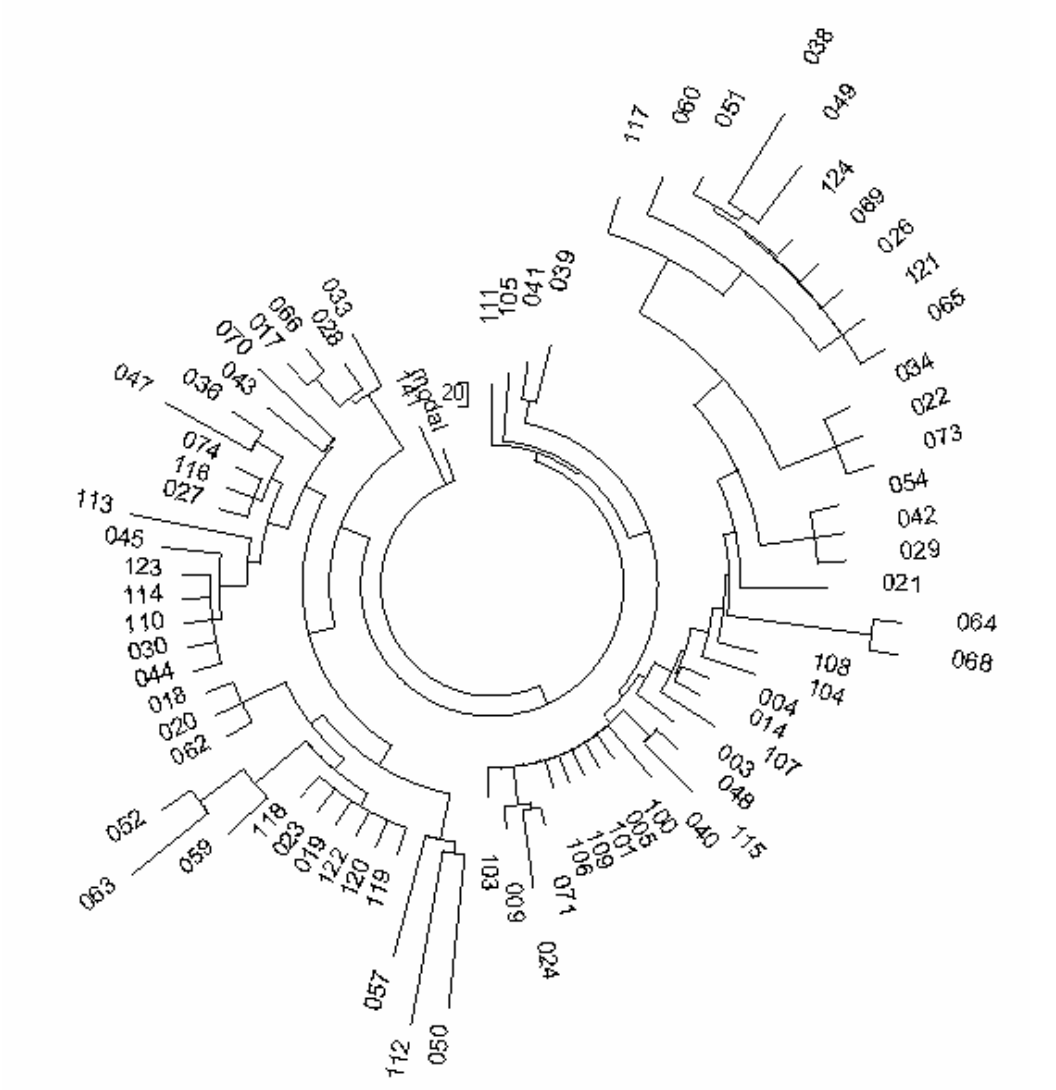

Figure 13. The 12-marker haplotype tree for 75 Jewish haplotypes of haplogroup J2. A "commercial" set, extracted from YSearch database

If to add to this set a small 4-haplotype branch at the top of the tree (Fig. 13), which contains two CMH haplotypes from the 37-marker tree (see below), there will be 6 base haplotypes from 22, and 27 mutations in those 22 CMH haplotypes. This in turn would give ln (22/6)/0.024 = 54 generations, and 27/22/0.024 = 51 generations from a common ancestor. It is again within the

same time span that is identified above to a common ancestor of the "CMH" in haplogroup J2.

This all point at the "Cohen modal haplotype" as a rather recent haplotype in J2 haplogroup, which has initiated a branch of descendants in 7-9th century AD.

All other 53 haplotypes except the branch of CMH haplotypes have the following haplotypes as a base

12-23-15-10-14-16-11-15-12-13-11-29

It is exactly the same ancient base haplotype as identified below from the 37-marker haplotype tree. In the 12-marker format it contains 251 mutations, that give 39 mutations per 100 markers, which translates to 246 generations, or 6,150 years to a common ancestor. This is in the same time range as 5,875 years time span, identified above with 10-marker "scientific" set of haplotypes.

A recent 10-haplotype branch at the upper right-hand side of the tree (Fig. 13) has the following base haplotype

12-22-14-10-14-15-11-15-12-14-11-31
14-15-22-10-11-12

which is exactly the same as 12-marker base haplotypes for a 37-marker three-haplotype branch in the upper left-hand side in Fig. 14, and in the 6-marker haplotype tree in the long flat branch on the right-hand side in Fig. 10. Its mutations point at a common ancestor who lived 29±2 generation, or 725±50 years BP.

It is of interest that even 6-marker haplotypes with their relatively low informative capacity give reasonably close figures to those obtained with 12- and 37-marker haplotypes (see below).

37-marker haplotypes (YSearch database)

37-marker haplotypes with their refined capacity are better suited for calculations time spans to common ancestors, since branches in them are much better refined. A 25-haplotype 37-marker tree of Jewish haplotypes of haplogroup J2 (Fig. 14) contains two distinct branches. One branch in the upper right part of the tree, containing 10 haplotypes, is obviously young, since it sits next to the "trunk" of

the tree. Besides, 8 out of these 10 were the "Cohen Modal Haplotypes" sitting as a comb on top of the tree in Fig. 11.

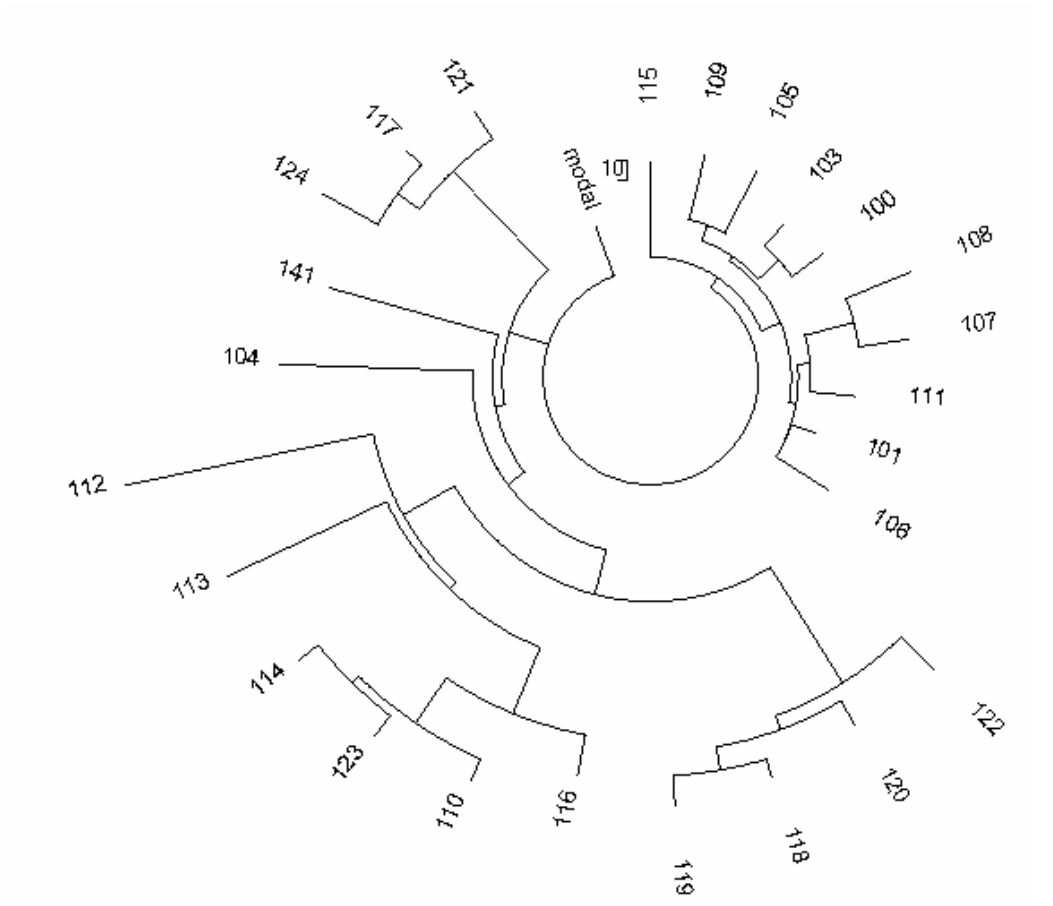

Figure 14. The 37-marker haplotype tree for 25 Jewish haplotypes of haplogroup J2. A "commercial" set, extracted from YSearch database

Another, wide branch, consists of a series of rather "young" branches, identified above with 6-marker haplotypes, all leading to a common ancient ancestor.

A "J2 Cohen modal haplotype" branch, 1,200±200 years old (~9±2 century AD)

It turned out that the younger branch on the upper right side in Fig. 3-5, represents the "Cohen Modal" haplotype with just a few mutations, namely 9, 25 and 44 mutations in the 12-, 25- and 37-marker panels, which gives 39, 56 or 51

generations, on average 49±9 generations to a common ancestor, approximately only 1,200±200 years ago. In a 6-marker and 12-marker format the "J2 Cohen Modal" haplotype is, respectively

14-16-23-10-11-12

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

One can see that in the 6-marker format the haplotype is the same as the CMH in haplogroup J1. In the 12-market format the J2-CMH has four mutations away from the J1-CMH (mutated markers are shown in bold):

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

Surprisingly, 49±9 generations is practically the same time span as that for the "recent CMH" in haplogroup J1 (54±4 generations). It is, of course, a pure coincidence, since their 37- marker haplotypes differ by as many as 30 mutations (!), which is about 503 generations, or 12,600 years apart. Besides, they belong to two different haplotypes, J1 and J2.

After five more Jewish haplotypes of haplogroup J2a1b (formerly J2f) were added to the tree (all available 37-marker haplotypes from J2f haplogroup in YSearch), it was noticed, that all five went to the CMH branch, increasing the number of haplotypes there from 10 to 15. It turned out that a 37-marker base (modal) haplotype for these Jewish J2a1b haplotypes is exactly the same as that for the CMH J2 haplotypes (Table). After the combining the branches and mutations in them, the number of mutations became 11, 36 and 65 in the 12-, 25-, and 37-marker panels, that leads to 46±13 generations, or 1,200±300 years to a common ancestor. This is within an error margin with the above figures for J1-CMH and J2-CMH.

Ancient branches (6,000±1,000 and 3,100±200 years BP)

The rest of the 37-marker J2 Jewish tree (Fig. 14) represents one wide 15-haplotype branch consisting mainly of descendants of rather recent ancestors, identified above on their 6-marker base (for each branch) haplotypes

14-15-22-10-11-12
15-16-23-9-11-12
15-15-23-10-11-12

The respective 37-marker haplotypes of their common ancestors are listed in the Table. The respective 12-marker haplotypes are as follows:

12-22-14-10-14-15-11-15-12-14-11-31
12-23-15-9-14-16-11-16-12-13-11-29
12-23-15-10-14-17-11-15-12-13-11-29

All of them are positioned very far from the J2-CMH on a time scale, between approximately 12 thousand years for the first one to 6 thousand years for the last one. Hence, they are practically unrelated to the J2-CMH within the time frame of the Jewish DNA genealogy.

The deduced 37-marker haplotype for a common ancient ancestor of all haplotypes on the separate branch (other than J2-CMH) of the J2 Jewish tree, including haplotypes descended from the three rather recent ancestors, identified above, is as follows

12-23-15-10-14-16-11-15-12-13-11-29-15-8-9-11-11-26-15-20-30-13-13-15-16-10-10-19-22-16-14-17-18-36-37-12-9

Compared to the J2-CMH branch, containing only 9 mutations per 100 markers on average, the 15-haplotype ancient branch contains as many as 38±6 mutations per hundred markers in haplotypes (an average from 12-, 25- and 37-marker panels). It corresponds to 234±40 generations to a common ancestor, that is approximately 5,800±1,000 years ago. All sub-branches of the ancient branch are of the same age with respect to their common ancestor, as can be seen in a frequency of their mutations. For example, a four-marker sub-branch on bottom left has 41±8 mutations per 100 markers, the three-marker sub-branch on the upper right has 42±7 mutations per 100 markers, and so forth. This all fits to 38±6 mutations calculated for the whole wide branch. Considering that, 10-marker haplotypes gave 5,875 years to a common ancestor, and 12-marker haplotypes gave 6,150 years, this on average gives 6,000±1,000 years to a common ancestor.

Another ancient ancestor was identified from the 37-marker haplotype tree in Fig. 14. Six haplotypes in a lower left-hand side area contain 15, 32, and 60 mutations in their 12-, 25, and 37-marker panels, which translates to 116, 131, or 125 generations, respectively, to a common ancestor. On average it gave 124±8 generations, or 3,100±200 years to a common ancestor.

These two base haplotypes of the ancient common ancestors are rather similar to each other, and might refer to a direct lineage to which both of them belong.

Indeed, they differ from each other by only 2, 6 and 9 mutations in their 12-, 25-, and 37-marker haplotypes, which translated to 2,900±700 years between them. In fact, this is a difference between 6,000 and 3,100 years, determined from mutations in haplotypes of their present day descendants.

Rather recent common ancestors, who lived 875±125 and 725±50 years BP

Two flat branches are located on different sides of the 37-marker haplotype three. They clearly derived from quite remote, however, rather recent ancestors. Indeed, a 4-haplotype branch in the lower right-hand side area contained only 12 mutations in all their 37 markers, that is in 148 markers total. This gives 35±4 generations , or 875±100 years to a common ancestor. The 3-haplotype branch in the upper left-hand side area contained only 4 and 7 mutations in their 25- and 37-marker panels. It gives 30 or 27 generations, that is 725±50 years to their common ancestor.

These two base haplotypes (Table) differ from each other by 7, 10 and 12 mutations in their 1-12, 13-25, and 26-37 marker panels, which corresponds to 12,400±2,300 year difference between the two common ancestors for each branch. Taking into account that they lived 875±100 and 725±50 years BP, we obtain a time span to their common ancestor, which is 7,000±1,200 years BP. It is within error margin with a time span of 6,000±1,000 to a common ancestor obtained earlier. He is likely the common ancestor for all Jews in haplogroup J2.

To summarize, all 37-, 12-, 10- and 6-marker haplotypes give rather consistent data regarding base (ancestral) haplotypes and time spans to common ancestors for branches on tree of haplotypes. About a quarter of all haplotypes in haplogroup J2 are the CMH in their "classical" 6-marker format, and about 40% of Jews in haplogroup J2 have the CMH in more or less mutated form. They are derived from a rather recent common ancestor who lived in 9±3 century AD. Almost half of all haplotypes descended directly from a common ancestor who lived 6,000±1,000 years ago. Approximately one quarter of all haplotypes are those from a rather recent common ancestor who lived 725±50 years ago.

The "Cohen Haplotype" in haplogroups J1 and J2 in the Jewish population

An origin of the "Cohen haplotype" has attracted a great attention since the end of 1990-s, when it was described (Skorecki et al, 1997; Thomas et al, 1998). It

appeared confusing for many that the "Cohen Haplotype" exists in both haplogroups J1 and J2, and that its origin was also contradictory. Also, it appeared rather strange that despite while J1 and J2 allegedly split some 12 thousand years ago, "Cohen haplotypes" in haplogroups J1 and J2 have rather similar structures; they are the same in a 6-marker format, and differ by only three mutations in a 12-marker format. Last but not least was a question how come that the cohens, high priests, who maintain their lineage for the last three thousand years or so, happened to be in two different haplogroups, J1 and J2.

This work adds more intriguing aspects to the issue, but largely – seemingly – resolves the mystery. It turned out that there are two "Cohen" lineages in haplogroup J1: the "recent" one (rCMH) originated 1,350±100 years BP, in the ~ 7th century AD, the "older" one (oCMH) appeared 1,650±350 years BP, around 350±350 AD. Besides, the "Cohen" lineage in haplogroup J2 also originated 1,200±200 years BP, in the ~ 9th century AD.

These issues can be resolved as follows. The rCMH was apparently proliferated initially by Bustenai, a descendant of King David himself, as it was explained in Part II of this study. He was a bearer of the "12 tribes haplotype" and presumably belonged to the Judah tribe, if to follow the biblical story. Understandably, his descendants were not the cohens, hence, so many of them in the "CMH" present day bearers. The ancestor of the oCMH was a cohen himself, hence, so many cohens among the bearers of the "CMH" nowadays. A common ancestor of both rCMH and oCMH lived 4,300±500 years BP, in the Abraham times, if to follow the biblical story again.

The "CMH" in haplogroup J2 is by many thousands years remote from the both CMHs in haplogroup J1, despite it was originated only 1,200±200 years BP. It has nothing to do with King David, "12 tribes haplotype" and with the cohens, unless some of them became the cohens on unrelated reasons other than the true lineage. Cohens as high priests could have gotten into J2 haplogroup on a number of different reasons – spousal infidelity, child adoption, change at birth – accidental or a deliberate one, and different origins of the Cohens themselves.

Now, how much the "Cohen haplotypes" in haplogroups J1 and J2 are distant from each other? Indeed, in the 6-marker format there is no difference between them:

14-16-23-10-11-12

In the 12-marker format the "Cohen haplotypes" in haplogroup J1

12-23-14-10-13-**17**-11-16-**11**-13-11-**31** (“older”, 1,650±350 years BP)

12-23-14-10-13-**15**-11-16-**12**-13-11-**30** (“recent”, 1,350±100 years BP)

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

and in haplogroup J2

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

it differs from the both J1-CMHs by one and three mutations, which is very little considering their assignments to two different haplogroups. However, with an extension of the haplotypes the difference grows progressively, and, for instance, a difference between the J2-CMH and the rCMH (haplogroup J1) is 12 mutations in the 25-marker haplotypes, and 30 mutations (!) in the 37-marker haplotypes. This corresponds to 503 generations, that is 12,600 years between them. With the CMH-J2 and oCMH-J1 a difference in their ancestral haplotypes is of 9,100 years. Clearly, “Cohen” haplotypes in haplogroups J1 and J2 have mutated quite independently from each other.

In fact, the Jewish haplotypes in haplogroups J1 and J2 are quite distant from each other indeed. This is illustrated in Fig. 15, which shows a combined tree of 67-marker haplotypes of haplogroups J, J1 and J2, to which Jewish haplotypes of haplogroup J2a1b (former J2f) were added. 67-marker haplotypes were chosen for their maximum available resolution.

Figure 15. A combined tree of 67-marker Jewish haplotypes of haplogroups J (haplotypes J01 and J02), J1, J2, and J2a1b (former J2f, haplotypes 130, 131, 134, 138, and 140). The 30 haplotypes were extracted from a "commercial", YSearch database. J1 haplotypes have numbers below 100 (between 006 and 088), J2 above 100 (between 101 and 122)

All J1-haplotypes sit on a separate branch on the left, all J2-haplotypes sit on a separate branch on the right. There is only one haplotype J2 which joined the J1

branch (number 112, on the very right). Both J haplotypes joined the J1 branch. All five J2a1b haplotypes joined the J2 branch randomly, without showing any clustering; in fact, interesting enough, all five joined the "recent Cohen Modal Haplotype" sub-branch, only 54 generations old.

Actually, Fig. 15 shows that haplotypes J1 and J2 have evolved separately for a very long time, both forming a mosaic of alleles rather similar in each of the haplogroup but quite distinct between them.

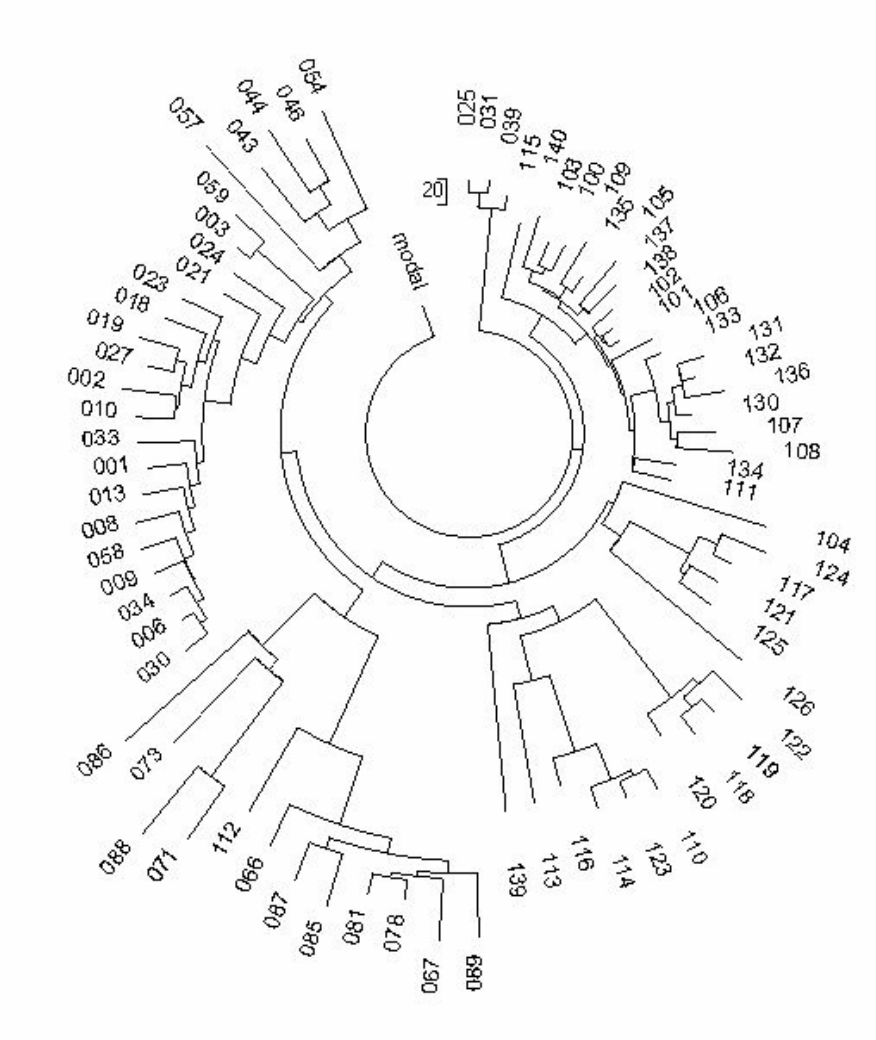

Figure 16. A combined tree of 37-marker Jewish haplotypes of haplogroups J1 and J2. The 75 haplotypes were extracted from a "commercial", YSearch database. J1 haplotypes have numbers between 001 and 089, J2 between 100 and 140.

Furthermore, the same in kind separation is observed on a tree of 37-marker haplotypes (Fig. 16), with a less resolution. The separation of all 75 haplotypes J1 and J2 was sharp, except only one J2 haplotype (the same number 112) which joined the J1 branch on the left, and three J1 haplotypes (on a mini-branch, numbers 025, 031 and 039) which joined the J2 branch on the right.

Jewish J2 haplotypes do not mix with non-Jewish J2 haplotypes

Jews have two "modal" haplotypes in haplogroup J2, depending on a branch. In a 6-marker format it is

14-16-23-10-11-12,

and besides it, one more can be considered as modal one

15-15-23-10-11-12

It is known, however, that in an overall haplogroup J2, which is 93% non-Jewish, the modal haplotype is

14-15-23-10-11-12

This haplotype is rather rare among Jewish J2 haplotypes, and presents only in 1-3% of total. It points out that the Jewish haplotypes in haplogroup J2 could be distant from others, forming their own niche. Indeed, it is proven to be correct. Fig. 17 shows a combined tree of 37-marker haplotypes of Jews J1 (38 haplotypes), Jews J2 (45 haplotypes), and non-Jews J2 (104 haplotypes). Jewish haplotypes were extracted from the database as fully as possible (as it was described above), while non-Jewish haplotypes were taken randomly from more than 1000 haplotypes, provided that the names of the current carrier of the haplotype and his most distant known ancestor were non-Jewish to the best of my judgment.

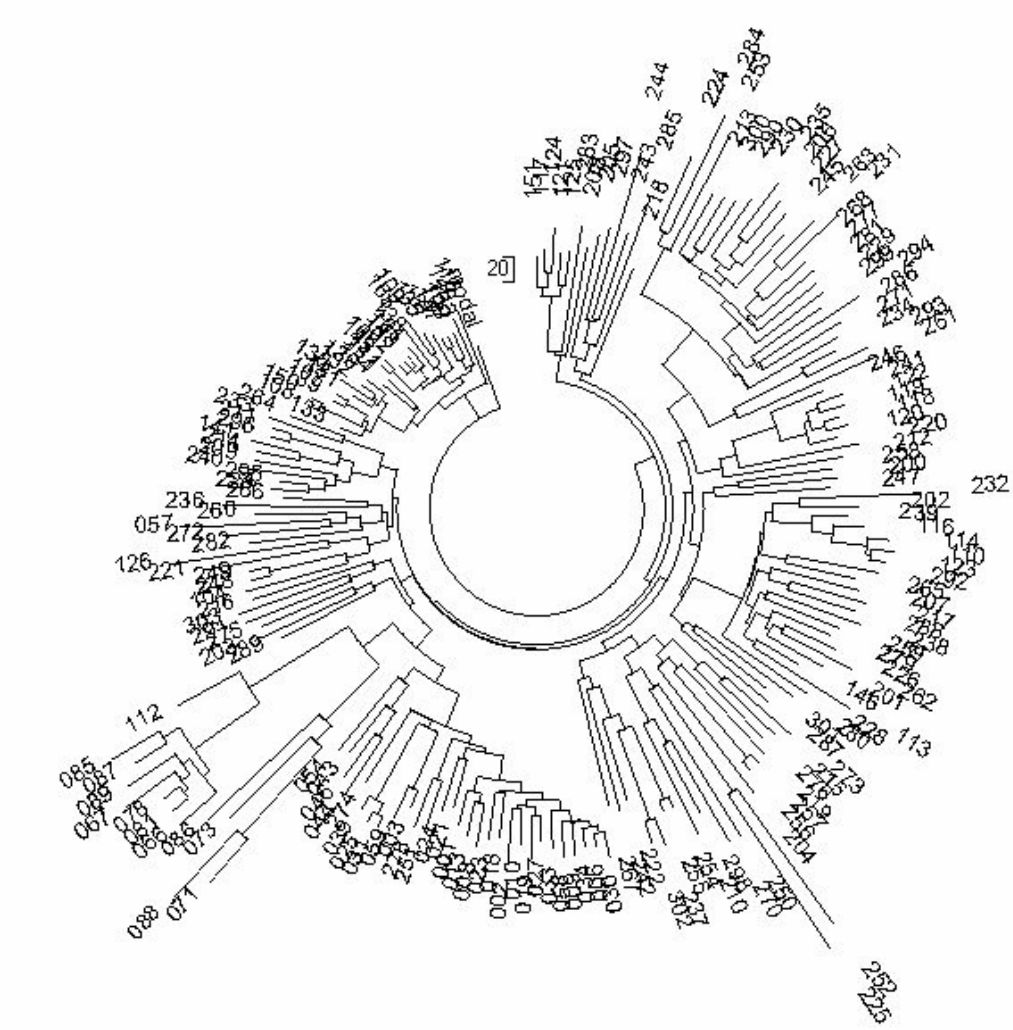

Figure 17. A combined tree of 37-marker Jewish (haplogroups J1 and J2) and non-Jewish (selected from haplogroup J2) haplotypes. The 187 haplotypes were extracted from a "commercial", YSearch database. J1 Jewish haplotypes have numbers between 001 and 089, J2 Jewish haplotypes between 100 and 151, J2 non-Jewish haplotypes have numbers between 200 and 303. 38 Jewish haplotypes of J1, 45 Jewish haplotypes of J2, and 104 non-Jewish haplotypes of J2 are shown.

The results turned out to be striking ones. Non-Jewish J2-haplotypes have taken the whole right-hand side, plus a few scattered inclusions into the left-hand side. One the right, among about 100 non-Jewish haplotypes, there less than ten Jewish haplotypes. Almost all Jewish haplotypes (J1 and J2 separately) have taken the

left-hand side of the tree. It also can be seen that non-Jewish J2 haplotypes are older (more distant from the tree base) than Jewish J2 haplotypes, but have a similar age with those of J1 haplotypes (15,500±4,100 years, as it was calculated in the preceding part of the study). Indeed, as it was estimated above, the oldest Jewish J2 haplotypes are of 6,000±1,000 years old.

The Lemba haplotypes

A list of 136 Lemba haplotypes was published in (Thomas et al, 2000). 41 of them are typical Bantu haplotypes, with a base haplotype

15-12-21-10-11-13

(the same as that in Bantu people) and have a common ancestor between 250 and 309 generations BP, that is 6,300 to 7,700 years BP.

Another 23 Lemba haplotypes belong to apparently haplogroup Q, with a base haplotype

14-12-23-10-15-14

which has a rather recent ancestor who lived between 1500 and 1800 years BP, that is in the 3^{rd}-6^{th} century AD.

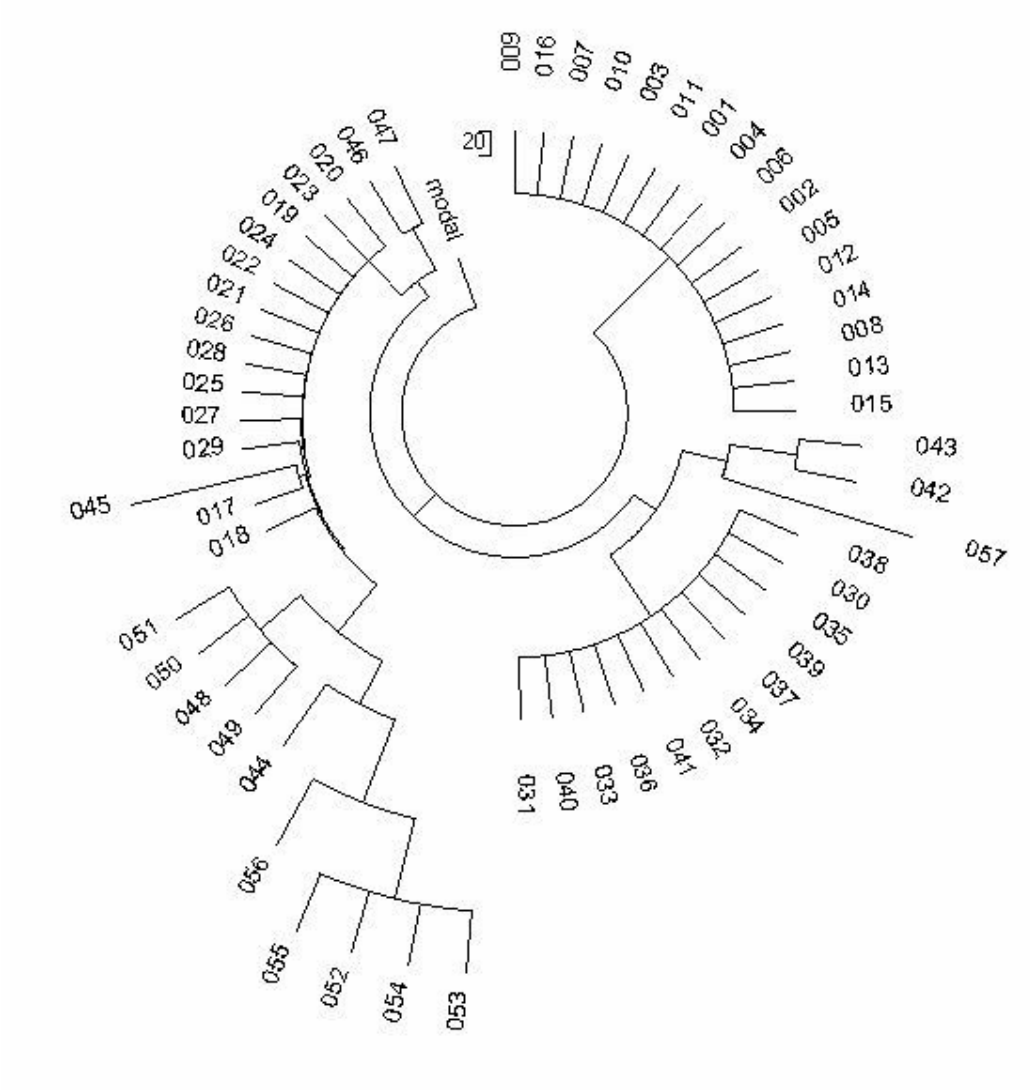

Figure 18. The Lemba 6-marker 57-haplotype tree, apparently of haplogroup J. A "scientific" set (Thomas et al, 2000)

There are a few scattered Lemba haplotypes, apparently from different haplogroups, and finally there are 57 haplotypes of apparently haplogroup J, which in turn split into three different branches. It is shown in Fig. 18. The three base haplotypes, one for each branch, are shown in their descending order:

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, which obviously came from a very recent common ancestor. As one can see from the haplotype tree (Fig. 18), the branch is in the upper right-hand side, with none of these haplotypes are mutated as yet. Its common ancestor should have lived no more than a few centuries ago.

The second one, being a base haplotype for a 26-haplotype branch on the left, points out at a common ancestor, who lived between 2,000 and 2,300 years BP, that is at the beginning of AD. It is rather common haplotype in the Arabic world. Its amount/fraction of the base haplotypes gives 78 generations to a common ancestor, and 21 mutations in all 26 haplotypes give 92 generations to a common ancestor. In any case, they are not the "Cohen haplotypes" in their origin. Their common ancestor had a haplotype with two mutations away from the CMH.

The third base haplotype supports a branch of 15 haplotypes on the lower right-hand side, 12 of them, which resemble the CMH, are identical to each other, and occupy a flat branch. Again, there are no mutations in them, and they must have come from a very recent ancestor, again a few centuries ago. From a fraction of the base haplotype, their common ancestor lived only ln (15/12)/0.0096 = 23 generations ago, that is about 600 years BP.

The three mutated haplotypes in this series are quite different, and clearly do not belong to the same group. 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 too much of a stretch. They could have been Jewish (but originated just a few centuries ago), they could have been Arabic. The 6-marker format cannot clarify the dilemma. Also, it is obvious that they are derived from quite a recent ancestor.

In conclusion, the so-called "Cohen Modal Haplotype" in the "Black Jews of Southern Africa" has nothing to do with an ancient history of the Lemba. It is a rather recent acquirement.

Part II

The Jews in Haplogroups R1b* and E1b1b*

The Jews of Haplogroup R1b and its subgroups R1b1, R1b1c

R1b (M343) is considered to be mainly a Western European haplogroup, and – according to some views – it is believed to appear some 30-35 thousand years BP in Europe, apparently when Neanderthals were still on the continent. It seems that this estimation regarding R1b* is grossly exaggerated , and a common ancestor of the present day R1b* bearers in Europe lived no earlier than some 5,000 years BP. An Asian common ancestor of R1b* lived much earlier, some 13,700 years BP. In fact, we do not know which haplogroup of those who had mingled with Neanderthals belonged to.

The above data follow from a mutational difference between various base (ancestral) European R1b* haplotypes, which is maximum 10 mutations per 25-marker haplotypes (about 7,000 years between the respective common ancestors), and 3,900 years BP to the oldest common ancestor of R1b* in Europe), on the one hand, and a difference between the oldest European base R1b* haplotype and the Asian R1b* base haplotype, which is 21 mutations per a 25-marker haplotype (about 21,600 years between the respective common ancestors, or 13,700 years to a common ancestor of the European and the Asian R1b* haplotypes), on the other. Some data and thought regarding the origin of R1b haplogroup in Europe and Asia along with respective time span estimates are given in the preceding paper in this issue and will be discussed in more detail elsewhere.

The most widespread haplotypes in haplogroup R1b and its subgroups R1b1 (P25) and R1b1c (M269) belong to the so-called Atlantic Modal Haplotype, in a 6-marker format

14-12-24-11-13-13

and in a 12-marker format

13-24-14-11-11-14-12-12-12-13-13-29

This haplotype is also called Benelux Modal haplotype, Ancient Celtic haplotype, Tribes of Wales haplotype, East Anglia Modal haplotype, and some other names.

At present the YSearch database contains 16 thousand haplotypes R1b with subgroups, out of a total number of 40 thousand haplotypes in the database. This means, however, no more than that the Western Europeans are the most eager to learn their DNA-related roots and can afford it.

The question is – where the Western European haplotypes have gotten to the Jewish community from, and when – still in ancient times, before the Jews had even shaped as peoples, or more recently, maybe during the Diaspora?

There is a small set of 47 ten-marker haplotypes of Jews of haplogroup R1b, published in the literature (Behar et al, 2004). Besides, a more extended set of 109 haplotypes was sent to me by a "Jewish R1b Project", which contained 6-, 12-, 37- and 67-marker haplotypes.

Before we go into details, here is a brief summary of findings.

- The Jews of R1b (and subgroups) haplogroup are all descended from one common ancestor, who lived 5000±200 years BP. His haplotypes in 6- and 12-marker formats are as follows (the 37-marker haplotype is shown in the Table):

  14-12-24-11-13-13

  13-24-14-11-11-14-12-12-12-13-13-29

  This is the classical Atlantic Modal haplotype, as shown above.

  There was clearly an invasion of this haplotype into people who some 1,500 years later formed – jointly with bearers of other haplogroups, described in this paper – the Jewish community. This haplotype was probably carried by people during the Exodus, along with bearers of haplogroups J1 and J2, among others. This haplotype might have been brought from Europe with slaves, or with warriors, merchants, or otherwise, by people who had settled in a community which much later became Jewish. It could have brought from Armenia or from Asia, where common ancestors of R1b haplotypes lived more than 10 thousand years BP (see the preceding paper in this issue).

- There are at least three branches of haplotypes derived from this base (ancestral) haplotype, which themselves became ancestral for three rather recent extended families of Jews R1b (with subgroups). Each branch contains groups R1b, R1b1 and R1b1c, and R1 in some branches. Such a

mix of subgroups in each branch implies that either the resolution was not enough (even in 67-marker branches) to resolve the subgroups, or subgroup typing was not accurate, or definitions of the subgroups is principally flawed, or all the above. It might be, for example, that all of them are R1b1c, however, some of them were typed for R1b or R1b1 only.

- One group/branch of R1b includes descendants of a rather recent ancestor, who lived 1,100±250 years BP, around the 10th century AD, and has the following base haplotype

  14-12-24-10-14-12

  12-24-14-10-11-14-11-12-12-13-14-29

  About a quarter of all present-day R1b Jews are descendant from this ancestor.

- Another group/branch of R1b includes descendants of a rather recent ancestor, who lived 950±200 years BP, around the 11th century AD, and has the following base haplotype

  14-12-24-11-13-13

  13-24-14-11-11-14-12-12-11-13-13-28

  Another quarter of all preset-day R1b Jews are descendant from this recent ancestor.

- Yet another group/branch of R1b includes descendants of one more rather recent ancestor, who lived 775±150 years BP, around 13th century AD, and had the following haplotype

  14-12-24-10-14-12

  12-24-14-10-11-15-12-12-12-13-14-28

  About one-eights of all R1b (with subgroups) Jews descended from this rather recent ancestor.

Approximately one-third of the Jews in haplogroup R1b* descended directly from the ancient ancestor who lived 5,000±200 years BP. It is an earlier time

compared with the West-European R1b*, a common ancestor of the nowadays bearers of R1b* lived around 3,500±400 years BP.

Let us consider Jewish R1b* haplotypes, branches and haplotype trees in more detail.

6-marker haplotypes (a "scientific and a "club" sets)

Both 6-marker sets are shown in Figs. 19 and 20.

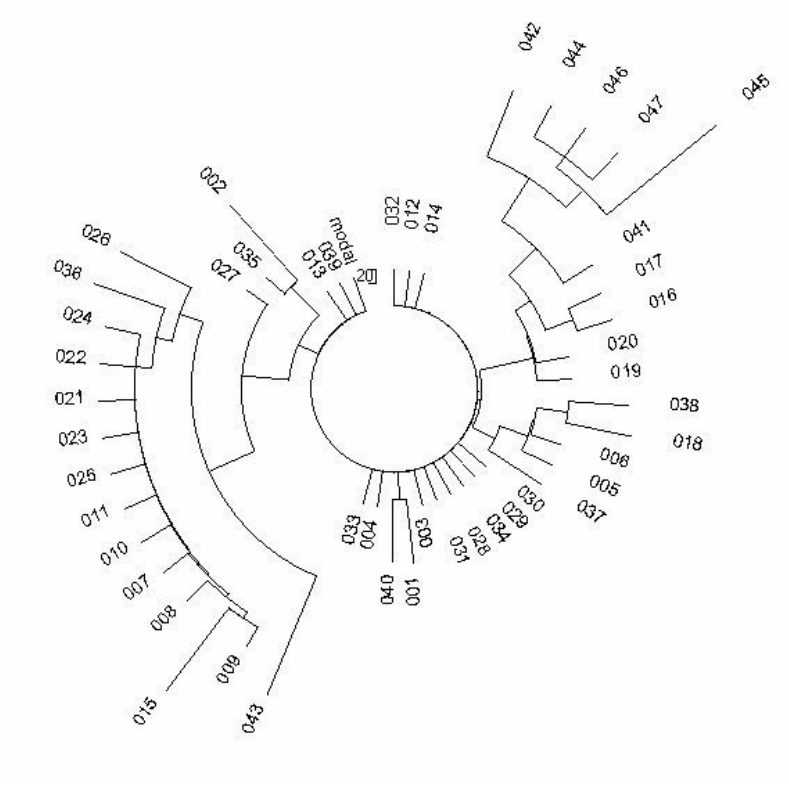

Figure 19. The 6-marker haplotype tree for 47 Jewish haplotypes of haplogroup R1b* (R1b, R1b1, R1b1c). A "scientific" set (Behar et al, 2004)

The "club set" of 109 haplotypes includes 17 haplotypes R1b (M343), 48 haplotypes R1b1 (P25), 42 haplotypes of R1b1c (M269). Besides, one haplotype was of R1 (M173) haplogroup and one more was unassigned.

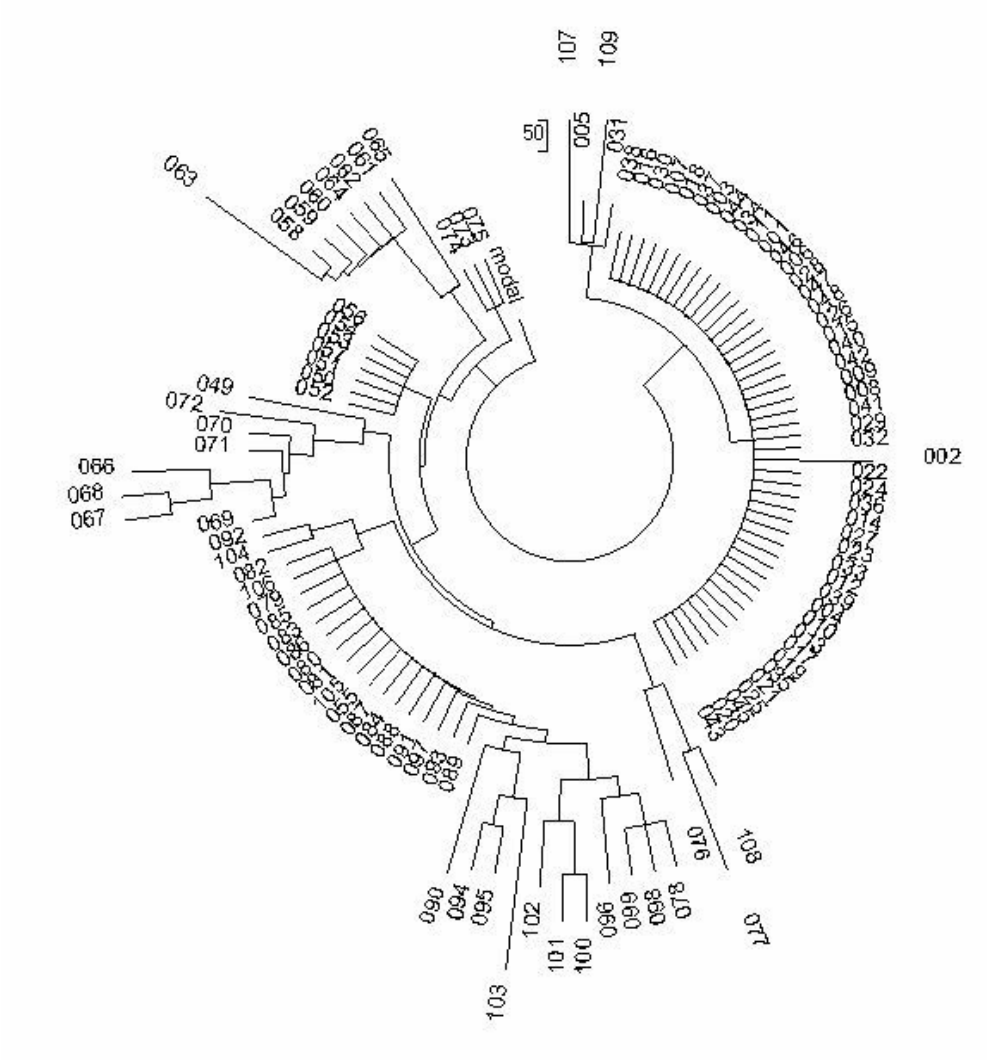

Figure 20. The 6-marker haplotype tree for 109 Jewish haplotypes of haplogroup R1b*. A "club" set (the Jewish R1b Project)

The "scientific" set (Behar et al, 2004) consists of 47 haplotypes of R1b haplogroup (SNP P25). Despite some slight differences, both trees are composed the same way, with the "scientific" tree is much smaller. In the "scientific" set (Fig. 19) the base haplotype of the tree

14-12-24-11-13-13

is the same as the Atlantic Modal Haplotype (AMH) shown above. There are 13 base haplotypes on the tree (sitting around the "trunk" of the tree at its base).

The 13-haplotype flat branch on the left contains 10 identical haplotypes

14-12-24-10-14-12

The "club set" (Fig. 20) gives the same two base haplotypes, as shown in the "scientific" set: the AMH on the left-hand side

14-12-24-11-13-13

accompanied with old protuberances (the most mutated haplotypes), and the flat branch on the right, with 45 identical haplotypes

14-12-24-10-14-12

out of 50 haplotypes on the branch. It might appear that the branch is derived from a rather recent ancestor, however, more extended haplotypes, such as 12- and 37-marker haplotypes, reveal many more mutations in them.

Let us take a look at the 15-haplotype ancient branch in the "scientific" set of 6-marker haplotypes. It is shown on the right-hand side in Fig. 19, contains 27 mutations, and is derived from the base ancestral haplotype

14-12-23-10-13-13

The number of mutations translates to 231 generations, or 5,800 years to a common ancestor. This figure can be considered only as an estimate, however, it shows that the ancestor is ancient indeed.

Clearly, as many as 13 base 6-marker Atlantic Modal haplotypes which is the base haplotype of the tree, could not left intact from the ancient ancestor who lived more than 5,000 years ago, unless a tree contained almost 100 haplotypes. In fact, it contains only 47 haplotypes. Hence, many of those base AMH are derived from a recent ancestor. It produced about one-third of current Jewish descendant haplotypes of haplogroup R1b in a 6-marker format. Another one-third is derived from one more recent ancestor, having a 6-marker haplotype

14-12-24-10-14-12

Yet another third of the Jews in haplogroup R1b forms a lineage from an ancient ancestor who lived more than 5,000 years ago, and had a base haplotype

14-12-23-10-13-13

Thus, both "scientific" and "club" sets of 6-marker haplotypes reveal the same ancient ancestors and two-to-three rather recent ancestors with the same base haplotypes in the both sets.

12-marker haplotypes (the "club" set)

A 12-marker haplotype tree did refine the above data regarding branches of the Jewish R1b tree. The 48-haplotype wide branch on the right (Fig. 21) includes a series of 16 identical haplotypes deriving from the recent ancestor with the 6-marker haplotype, identified above

14-12-24-10-14-12

and in its 12-marker format

12-24-14-10-11-14-11-12-12-13-14-29

This haplotype was analyzed in its 37-market format and was found to belong to a rather recent ancestor who lived 1,100±250 years BP, around the 10th century AD. Its haplotype significantly (4 to 5 mutations in a 12-marker format) differed from a series of known "Atlantic Modal" haplotypes (see below), and is characteristic only for the Jewish lineage in R1b haplogroup.

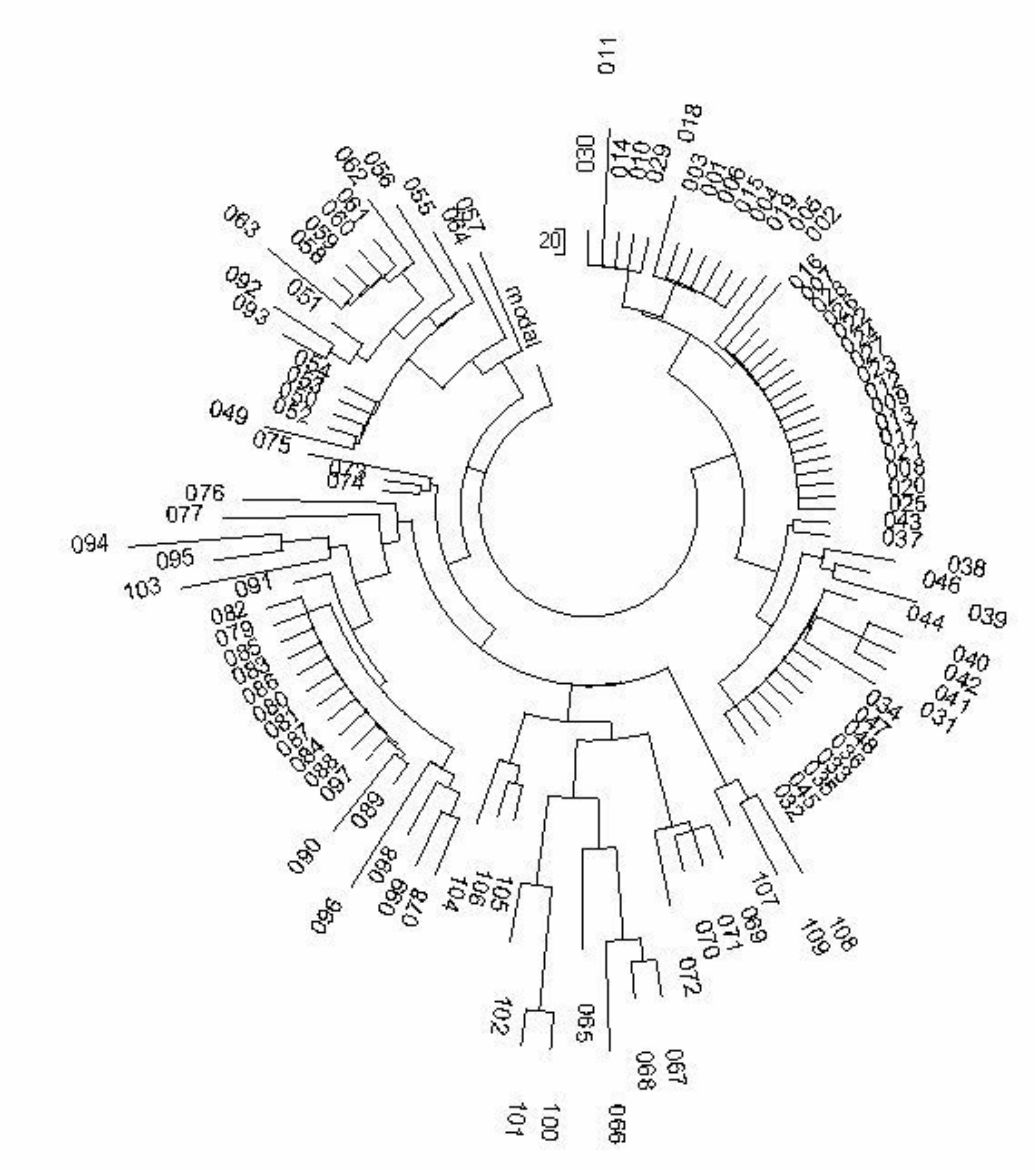

Figure 21. The 12-marker haplotype tree for 109 Jewish haplotypes of haplogroup R1b*. A "club" set (the Jewish R1b Project)

8 identical haplotypes on the right-hand side represent another recent ancestor haplotype, which had the same 6-marker haplotype as shown above

14-12-24-10-14-12

however, it is quite a different one (three mutations away) in a 12-marker format

12-24-14-10-11-15-12-12-12-13-14-28

from the 12-marker haplotype shown above. That is why 6-marker haplotypes sometimes can be deceiving in terms of an overall number of mutations. A common ancestor having the last haplotype lived 775±150 years BP, around the 13th century AD, as it is shown below using a 37-marker haplotype.

One more recent ancestor haplotype shown up on the left hand side as a flat branch having 12 identical haplotypes, which in a 6-marker format

14-12-24-11-13-13

looks as the classical Atlantic Modal haplotype.

However, in a 12-marker format

13-24-14-11-11-14-12-12-11-13-13-28

it significantly, by two to five mutations, deviates from different AMHs, described below. As it was established using 37-marker haplotypes, these haplotypes descended from an ancestor who lived about 950±200 years BP.

The whole 61-haplotype left-hand side branch in Fig. 21 is much older, has 230 mutations, and is derived from an ancestor who lived 190 generations ago, that is close to 5,000 years BP. Its 6-marker haplotype was

14-12-24-11-13-13

which is again the "classical" Atlantic Modal haplotype, as well as its 12-marker haplotype

13-24-14-11-11-14-12-12-12-13-13-29

Finally, the left-hand part and bottom of the tree (Fig. 21) contains clearly the oldest branch. It consists of 17 haplotypes, having as many as 86 mutations, that is 0.42 mutations per marker. It corresponds to 266 generations to a common ancestor for the branch. This apparent ancient common ancestor had the following 6 marker haplotype

14-12-23-10-13-13

and it was the same one that was identified on the 6-marker haplotype tree as a haplotype of an ancient ancestor (see above). Its 12-marker haplotype is as follows:

13-23-14-10-12-14-12-12-12-13-13-29

There is no match for this base haplotype on a 37-marker haplotypes tree. Partly, because half of those 17 twelve-marker haplotypes are not available in a 37-marker format, and the other half was largely diluted by other haplotypes on a new, rearranged 37-marker branch (see below). As a result, the time span to the ancestor moved from the 266 generations down to 216±20 generations, that is to 5,400±500 years to a common ancestor (Table). It seems that it is the same ancient ancestor of all the Jews in haplogroup R1b*, who loved 5,000±200 years BP.

37-marker haplotypes (the "club" set)

The set confirms the above findings and further refines time spans to common ancestors of Jews of R1b haplogroup (with subgroups). The data showed that there are at least four principal ancestors for the current Jewish population in R1b haplogroup.

An ancient ancestor, 5,000±200 years old

All 56 haplotypes (Fig. 22) show the following base haplotype, which is the "classical" Atlantic Modal haplotype:

14-12-24-11-13-13

13-24-14-11-11-14-12-12-12-13-13-29

This haplotype in a 37-marker format is shown in the Table. Its first, 12-marker panel has 224 mutations, the first 25 markers have 438 mutations, and all 37 markers show 826 mutations in all 56 haplotypes with respect to the above base haplotype. This gives 200, 205 and 196 generations, respectively, to a common ancestor for the whole haplotype tree, with an average 200±5 generations, or 5,000±200 years BP.

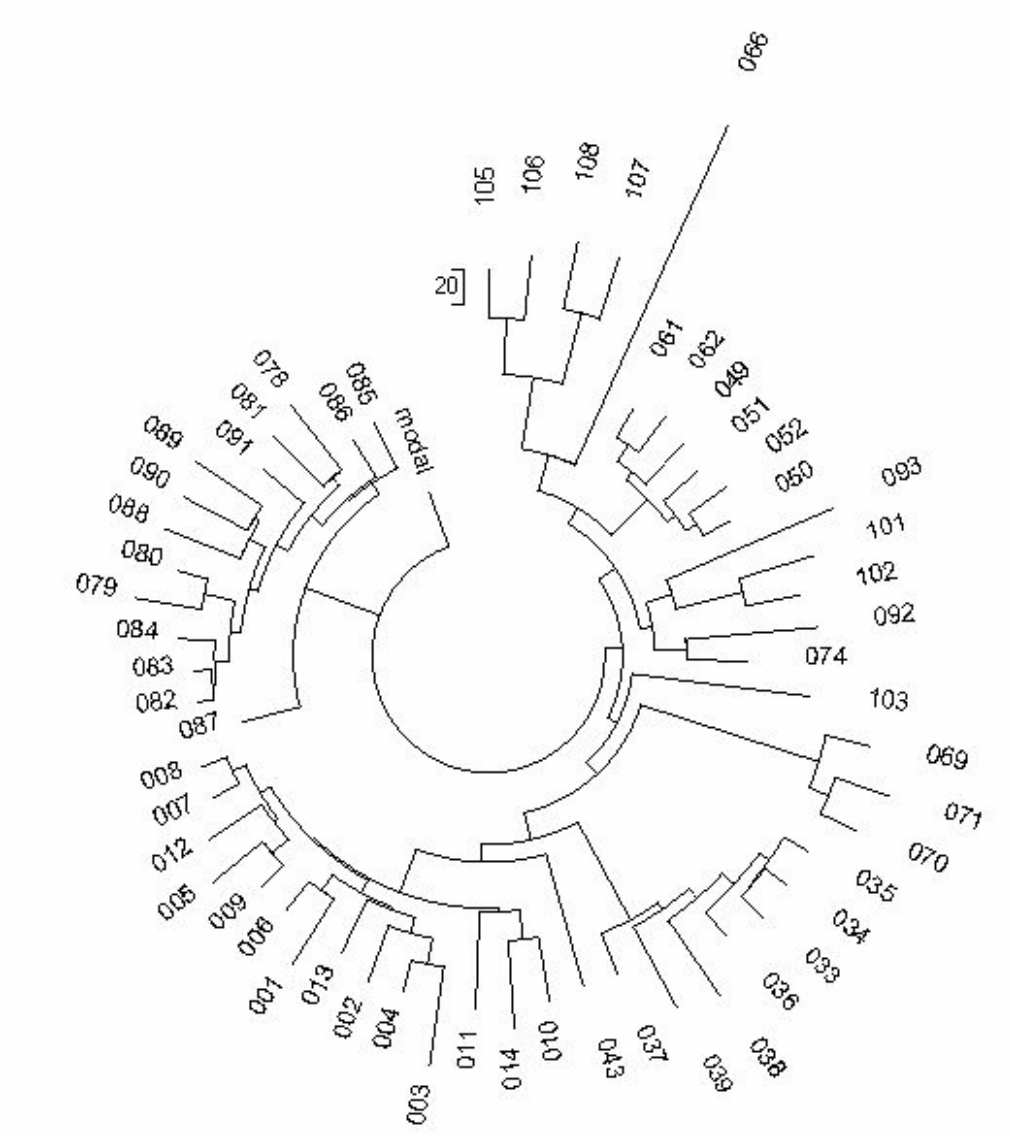

Figure 22. The 37-marker haplotype tree for 56 Jewish haplotypes of haplogroup R1b*. A "club" set (the Jewish R1b Project)

When the wide branch of 42 haplotypes is analyzed, which is the whole tree except the 14-haplotype recent branch on the left, the ancient haplotype has a different structure

14-12-24-10-14-12

12-24-14-10-11-14-12-12-12-13-14-29

which is 3 mutations away (on the 12-marker haplotype) from the ancient haplotype of 200 generations and 5,000 years old, a base for the whole tree. However, it gives practically the same 196±4 generations, that is 4,900±100 years to a common ancestor. It shows that the 14-haplotype recent branch is derived from the ancient ancestor.

The wide 20-haplotype branch on the right-hand side has 6- and 12-marker base haplotypes

14-12-24-11-13-13

13-24-14-11-11-14-12-12-12-13-13-29

It is the same 6-marker base AMH as one on the recent branch on the left. However, the 12-marker haplotype has four mutations compared with the recent AMH. Unlike the recent AMH, the right-hand branch was originated from an ancient ancestor. Based on mutations, the 20-haplotype branch has 231, 225 and 193 generations from a common ancestor for 12-, 25- and 37-haplotypes (90, 169 and 291 mutations, respectively), that is 216±20 generations from a common ancestor on average, or about 5,400±500 years BP.

Again, it is within the error margin at the 5,000 years time span to the ancient ancestor of the Jews of haplogroup R1b (with subgroups).

A recent common ancestor, 950±200 years old (~11th century AD)

A deduced base haplotype of a 14-haplotype branch at the upper left-hand side of the haplotype tree (Fig. 22) is listed in the Table. In a 6-marker format it is a "classical" AMH

14-12-24-11-13-13

In 12-marker format it is

13-24-14-11-11-14-12-12-11-13-13-28

It deviates from known European AMHs (shown below) in the 12-marker format by two to five mutations.

This branch contains 52 mutations in all 14 haplotypes, that gives 38±8 generations to a common ancestor, that is about 950±200 years BP, around 11th century AD.

Recent common ancestors, ~775±150 years old (~13th century AD) and 1100±250 years old (~10th century AD), and their common ancestor of 2,800±300 years old

The wide 22-haplotype branch at the bottom of Fig. 22 consists of two sub-branches. One has a 6-marker base haplotype

14-12-24-10-14-12

and a 12-marker base haplotype

12-24-14-10-11-15-12-12-12-13-14-28

(the full 37-marker haplotype is shown in the Table). These haplotypes were identified earlier on the 6- and 12-marker haplotype trees. The 37-marker haplotype sub-branch on its 12-, 25- and 37-marker panels gives 37, 26 and 31 generations respectively to their common ancestor, that is on average 31±6 generations, or 775±150 years BP.

The second sub-branch of 14 haplotypes has the same 6-marker base haplotype as that above, but a different, 3-mutation away 12-marker base haplotype

12-24-14-10-11-14-11-12-12-13-14-29

This haplotype was also identified in the preceding section, and is given in the 37-marker format in the Table. Its mutations point out at a common ancestor who lived 43±10 generations, that is 1100±250 years BP.

A common ancestor of the whole 22-haplotype branch had the following 6- and 12-marker haplotypes

14-12-24-10-14-12

12-24-14-10-11-14-11-12-12-13-14-29

and lived 112±12 generations, or 2,800±300 years BP. He had only two mutations in the 37-marker format compared with the preceding base haplotype (the Table).

Jewish vs. non-Jewish haplotypes of haplogroup R1b with subgroups

The groups and subgroups R1b, R1b1 and R1b1c did not affect a distribution of haplotypes and their branches on the tree. In fact, they are practically randomly mixed on the tree. Therefore there was no need to indicate which subgroup was more represented on which branch. Not one was.

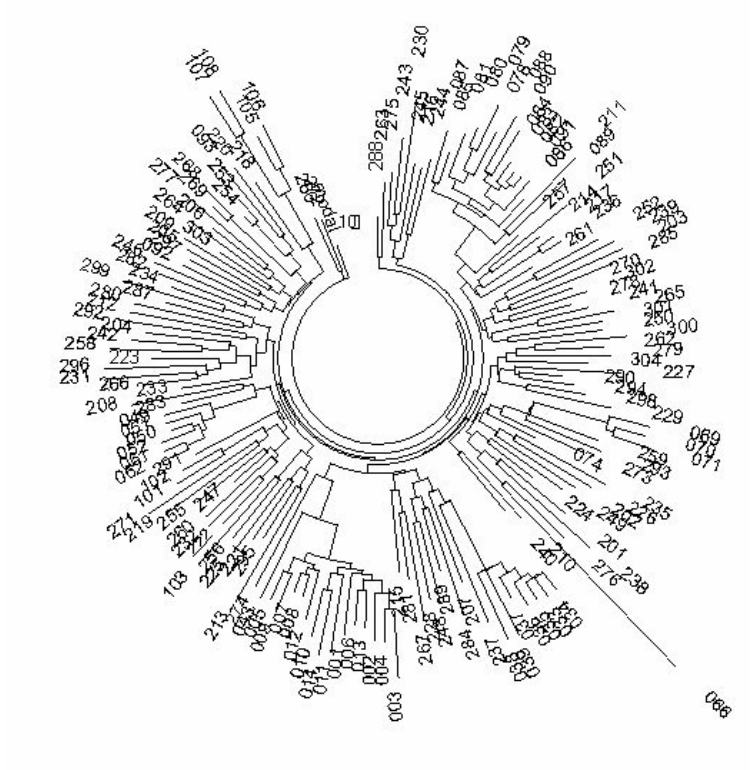

Figure 23. A combined tree of 37-marker Jewish (haplogroups R1b, R1b1 and R1b1c) and non-Jewish (selected from haplogroup R1b) haplotypes. The 160 haplotypes were extracted from a "commercial", YSearch database. R1b* Jewish haplotypes have numbers between 001 and 109, R1b non-Jewish haplotypes have numbers between 201 and 304.

Finally, 105 non-Jewish haplotypes of haplogroup R1b were added to the 37-marker Jewish 56-haplotype tree (Fig. 23). It turned out that the Jewish and non-Jewish 37-marker haplotypes were completely segregated. The 160 haplotypes split into 14 different branches. If a Jewish branch is marked as J, and a non-Jewish as N, that is how the tree looks like: N-J-N-J-N-N-J-N-J-N-J-J-J-N

Jewish R1b haplotypes and Atlantic Modal haplotypes

In haplogroup R1b (with subgroups R1b1 and R1b1c) Jewish haplotypes occupy their own niche. Their roots were originated some 5,000 years ago, and a common ancestor had a "classical" West European R1b modal 6-marker haplotype

14-12-24-11-13-13

In a 12-marker format the Jewish ancient R1b haplotype

13-24-14-11-11-14-12-12-12-13-13-29

exactly matches known West-European modal haplotypes, such as Benelux Modal haplotype, Ancient Celtic haplotype, Tribes of Wales haplotype, East Anglia Modal haplotype. These are possible areas from where this haplotype could have reached ancient pre-Jewish people.

Other common ancestral haplotypes of Jews of haplogroup R1b (with subgroups) do not match other AMH haplotypes and do mot mix with Western-European haplotypes of a haplotype tree. This indicates that present-day Jewish R1b haplotypes have evolved into separate lineages, not common in the West European pool of R1b haplotypes.

Main base (common ancestor) Jewish R1b haplotypes are as follows:

Ancient haplotypes (ancestors 3,000 and more year old)

13-24-14-11-11-14-12-12-12-13-13-29
13-23-14-10-12-14-12-12-12-13-13-29
12-24-14-10-11-14-12-12-12-13-14-29
12-24-14-10-11-14-11-12-12-13-14-29

Rather recent haplotypes (10^{th} – 13^{th} centuries AD):

12-24-14-10-11-14-12-12-12-13-14-29
13-24-14-11-11-14-12-12-11-13-13-28
12-24-14-10-11-15-12-12-12-13-14-28

Only the common ancestral haplotype for the whole tree (the top one above) matches one of the AMHs (the top one below). Other AMHs, including North Sea and others "modal" R1b haplotypes, have no matches with Jewish R1b base haplotypes:

13-24-14-11-11-14-12-12-12-13-13-29
13-24-14-10-11-16-12-12-12-13-13-29
13-24-14-10-11-14-13-12-12-13-13-28
13-23-14-11-11-14-12-12-12-13-13-29

It all shows that the Jewish haplotypes of R1b haplogroup are probably represent the most versatile, complicated pattern compared with any of the Jewish ten haplogroups described in this paper.

The Jews of Haplogroup E1b1b (former E3b) and some subgroups

Haplogroup E1b1b (SNP M35) has allegedly appeared in North Africa or at the Middle East approximately 25 thousand years BP, and derived from an older North African haplogroup E (SNP M96). Haplogroup E1b1b is a relatively wide-spread among Jews and observed in about 15-20% of Ashkenazim and up to 30% Sephardim.

Three sources of E1b1b (and some subgroups) haplotypes are considered here: (a) the publication of 2004 (Behar et al, 2004), (b) a data set from YSearch database (2008) and (c) a collection of haplotypes of the Jews who belong to haplogroup E1b1b and its variants – E (SNP M40, M96), E1b1b1 (former E3b1, SNP M35), E1b1b1a (former E3b1a, SNP M78), E1b1b1a1 (former E3b1a1, SNP V12), E1b1b1a2 (former E3b1a2, SNP V13, V36), E1b1b1b2 (former E3b1b2, SNP M165, M183), and E1b1b1c1 (former E3b1c1, SNP M34). It was taken from the "Jewish E3b Project" (see References).

The most widespread haplotypes in haplogroup E* in 6- and 12- marker formats are

13-12-24-10-11-13
13-24-13-10-16-18-11-12-12-13-11-30

In haplogroups E1b1* (former E3*) and E1b1b* (former E3b*) they are exactly the same. These are referred not to Jewish haplotypes, but for all available haplotypes from said haplogroups, total number of haplotypes 1,900, 1,850 and 1,265, respectively.

As it was mentioned earlier, there are three available sets of Jewish E1b1b (and some other subgroups) which are considered in this paper. One of them, consisting of 71 haplotypes of E1b1b haplogroup (in a 10-marker format), and referred to as the "scientific" set, was published in (Behar et al, 2004), another – 140 haplotypes of E1b1b – was collected from YSearch database as it was described above, the third one was obtained from the "Jewish E3b Project", and referred to as the "Club set". It contained 117 haplotypes of haplogroups E, E1b1b and subgroups, listed above. The "scientific" set (1) included only 10-marker haplotypes, the two other included also 12-, 37- and 66-marker haplotypes, which has allowed to perform a more refined analysis of haplotype trees.

Before we go into details, here is a brief summary of findings.

- The Jews of E1b1b haplogroup (and subgroups, which are not segregated even on a 66-marker tree, as it is shown below) all descended from one common ancestor, who lived 6,800±400 years BP. His haplotypes in 6- and 12-marker formats are as follows (the 37-and 66-marker haplotypes are shown in Table 2):

  13-12-24-10-11-13
  13-24-13-10-16-18-11-12-12-13-11-30

  It is exactly the same haplotype as shown above as the most frequent haplotype in all three haplogroups, E*, E1b1* and E1b1b*. Hence, there is nothing distinctly "Jewish" in this haplotype. Based on an overall number of mutations in the whole YSearch database (haplogroup E and subgroups), a common ancestor of haplogroup E lived 10,700±800 years BP, haplogroup E1b1 10,300±600 years BP, and haplogroup E1b1b 6,700±300 years BP. It should be considered, though, that the above figures are based on superposition of haplotypes from both ancient and rather recent ancestors, and are given here as an estimate only. On the contrary, the Jewish ancestor of 6,400±400 years BP is the most ancient one among the Jews of haplogroup E1b1b* and its age is determined from

the "oldest branch" in a haplotype tree. It is very likely that there was an invasion of the above haplotype more than 6 thousand years ago into people who much later became Jewish. This haplotype was very likely carried by people during the Exodus, along with bearers of other Middle Eastern haplogroups, J1 and J2 in particular. About two-thirds of all present-day Jews of haplogroup E1b1b descended from this common ancestor.

- There are at least five branches of haplotypes derived from this ancient base (ancestral) haplotype, which themselves became ancestral for at least three rather recent extended families of Jews of E1b1b* haplogroup. Some branches contains several subgroups. Such a mix of subgroups in a number of branches implies that either the resolution was not enough (even in 66-marker branches) to resolve the subgroups, or subgroup typing was not accurate, or definitions of the subgroups is principally flawed, or not all of them were typed for different subgroup level, or all the above.

- Besides the ancient ancestor of 6,400±400 years BP, from whom, as it was said above, about two-thirds of all present-day Jews of haplogroup E1b1b* are descended, three other rather recent common ancestors, who lived 1,000±250, 975 ±100, and 750±200 years BP, were identified. It appears that there were a series of genealogical bottlenecks among the Jews around the 11-13th centuries AD, related probably to Khazars, or settling of the Jews in Eastern and Central Europe, or expulsion of Jews from England in 1290. It is know, that on July 18, 1290 every professing Jew in England was ordered out of the Realm, for ever, by King Edward I. Between sixteen and seventeen thousand Jews had to flee, and none returned until some 400 years later.

- Said three rather recent common ancestors and their descendants have resulted to one-quarter-to-more-than-one-third, one-sixth, and one-fifth, respectively, of all present-day Jews of E1b1b* haplogroup. Their haplogroups in a "scientific" (6-marker) and 12-marker formats are as follows, respectively (37- and 66-marker haplotypes are shown in Table 1):

13-12-25-9-11-14
14-12-24-10-11-13
13-12-24-10-11-13

14-25-13-9-17-18-11-12-12-13-11-30

13-24-14-10-16-17-11-12-13-14-11-32
13-24-13-10-17-18-11-12-12-13-11-30

The last 6-marker haplotype is the "youngest" one (that is derived from the most recent ancestor), and it is identical to its parent, ancient haplotype. Hence, the abundance of this particular haplotype on a 6-marker haplotype tree (Figs. 24 and 25, both on the right-hand side). Another abundant 6-marker haplotype (Figs. 24 and 25, both on the left-hand side) is the first one among the three haplotypes above, since its share is the highest one among the Jews, that is between 25% and 38% of all in haplogroup E1b1b*.

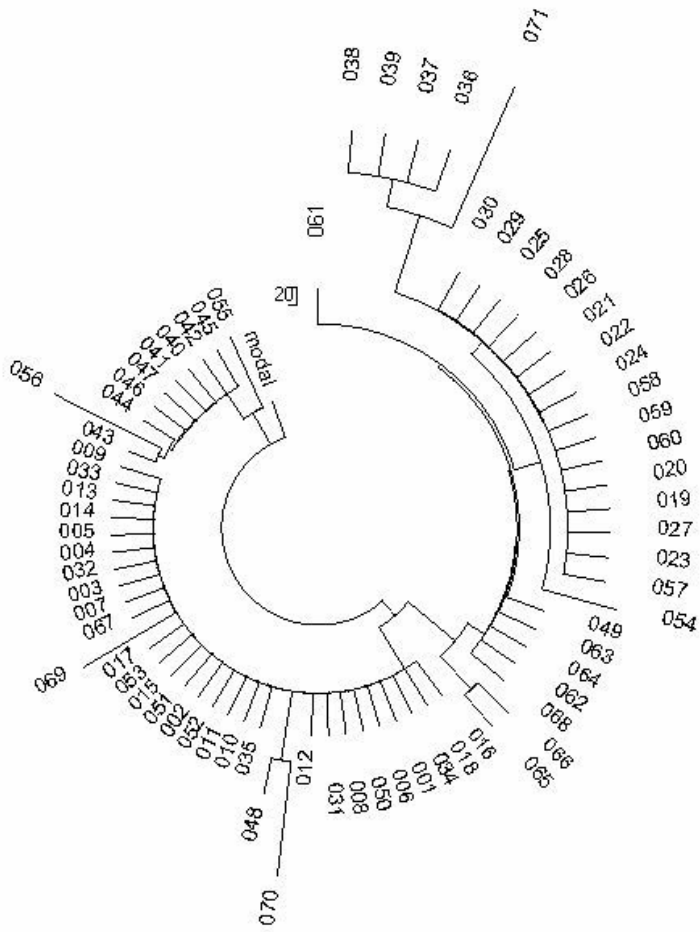

Figure 24. The 6-marker haplotype tree for 71 Jewish haplotypes of haplogroup E1b1b (former E3b). A "scientific" set (Behar et al, 2004)

Let us consider these haplotypes, branches and haplotype trees in more detail.

6-marker haplotypes ("scientific", "commercial", and a "club" sets)

Figs. 24-26 show three sets of 6-marker haplotypes arranged as the respective trees. All of them are similar in kind. Each one contains several "flat" branches, in which haplotypes are identical to each other. These haplotypes are of rather recent origins. An old branch would not have contained so many identical haplotypes; they would form a series of mutated variants. However, in some cases those flat branches represent a mix of haplotypes deriving from both ancient and rather recent ancestors. Those flat branches split into separate branches on transition to more extended haplotypes.

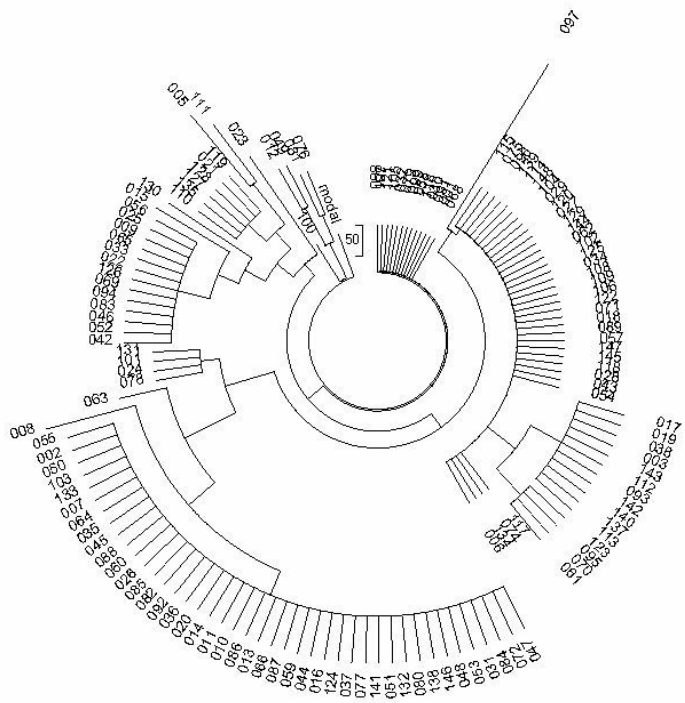

Figure 25. The 6-marker haplotype tree for 140 Jewish haplotypes of haplogroup E1b1b. A "commercial" set, extracted from YSearch database

There are five "flat" branches in each of the three 6-marker haplotype set (Figs. 24-26):

13-12-24-10-11-13
14-12-24-10-11-13
13-12-24-10-11-14
13-12-25-9-11-14
13-12-25-10-11-13

The first one is a superposition of haplotypes derived from both ancient and rather recent ancestors, who lived 6,800±400 and 750±200 years BP. This flat branch is located on the right in the "scientific" and "commercial" sets (Figs. 24 and 25) in the amount of 23% and 24% respectively of all haplotypes in the set, and on the top (as a base haplotype to the set) in the "club set" (Fig. 26), in the amount of 21% of the total amount. These data show again how similar in kind are sets of haplotypes taken from quite different sources.

Figure 26. The 6-marker haplotype tree for 117 Jewish haplotypes of haplogroup E1b1b*. Haplotypes haplogroups E (haplotypes 001 and 002), E1b1b (80 haplotypes), and sub-groups E1b1b1a (031 and 103), E1b1b1a1 (037), E1b1b1a2 (038), E1b1b1a3 (044, 045, 046), E1b1b1b2 (047, 048, 049), E1b1b1c1 (25 haplotypes) are shown. A "club" set (the Jewish E3b Project).

The second haplotype belonged to another rather recent ancestor, who lived 975±100 years BP. In each of the three sets (Figs. 24-26) the flat branch is remote from the tree (the upper right-hand area in Fig. 24, on the left in Fig. 25, and on the right in Fig. 26). The "scientific" and "commercial" sets both contain less than 10% of these haplotypes, the "club" set contains 16% of them.

The third ancestral haplotype is approximately 1,300 years old. That is why there are only relatively few of them in all three sets – after 1,300 years about 40% of 6-marker base haplotypes are mutated. The "scientific" and "commercial" sets each contains 10% of these haplotypes, and the "club" set contains 4% of them. They are located in the upper left corner (Fig. 24), on the top (Fig. 25), and in the lower right corner (Fig. 26). There are just few of them in 37-marker and more extended formats, and they have not been analyzed in detail. After all, they represent only a small branch of an ancient wide branch.

The next, fourth haplotype, forms the largest flat branch in each of the three sets (on the left and bottom in Fig. 24, and the bottom in Fig. 25, and on the left in Fig. 26). It belonged to a rather recent common ancestor, who lived 1,000±250 years BP. This 6-marker haplotype occupies between 30% and 40% of each of the trees.

The last ancestral haplotype belonged to a recent ancestor, who lived approximately 550 years BP. The haplotype is a rather minor one, and occupies only 7%, 8% and 7% in the "scientific", "commercial" and "club" trees (in the bottom right corner in Figs. 24 and 25, and in the bottom left corner in Fig. 26). Again, the data show how similar the sets of haplotypes are despite quite different sources they were taken from. There are only three of these haplotypes in a 37-marker format, and they have not been analyzed in detail due to lack of statistics.

10-marker haplotypes (the "scientific" set)

A 71-haplotype 10-marker set was published (Behar et al, 2004), and the respective tree constructed based on these data is shown in Fig. 27.

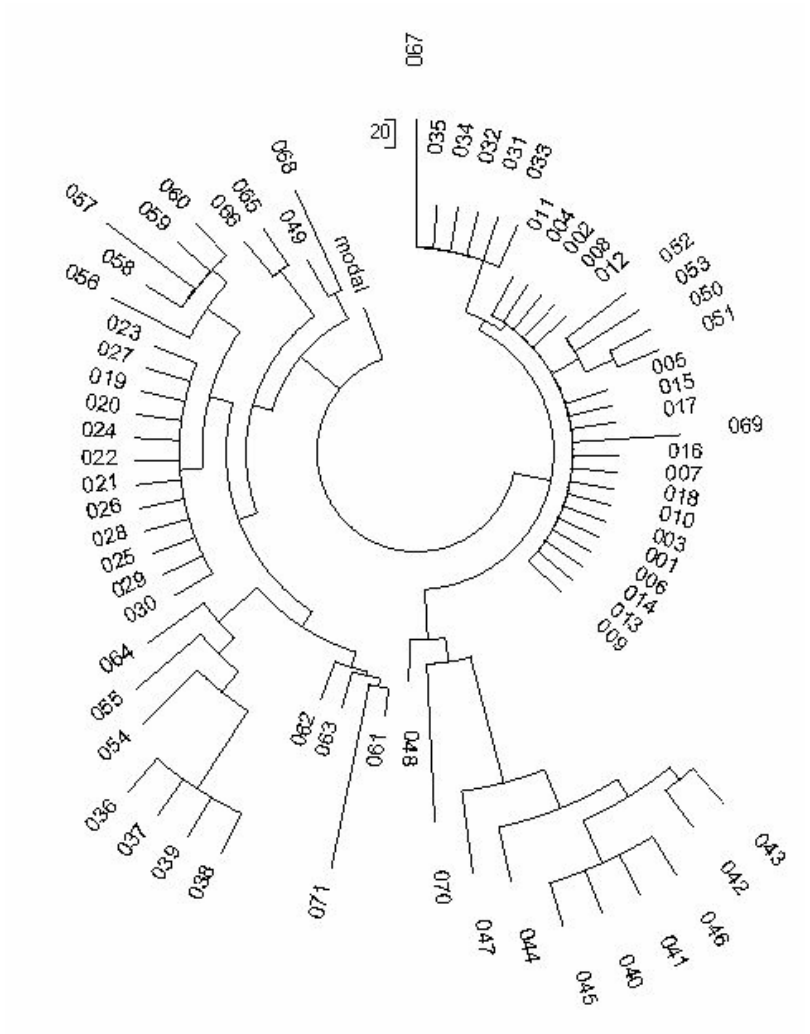

Figure 27. The 10-marker haplotype tree for 71 Jewish haplotypes of haplogroup E1b1b (former E3b). A "scientific" set (Behar et al, 2004)

Compared with the 6-marker tree (Fig. 24) the tree is more detailed, showed two ancient common ancestors (two "protuberances" at the bottom of the tree on the left- and right-hand sides), and also confirmed that haplotypes

13-12-24-10-11-13
13-24-13-10-X-Y-11-12-12-13-11-30

and

13-12-25-9-11-14

14-25-13-9-X-Y-11-12-12-13-11-30

in a 6- and 10-marker format, in which X and Y have not been identified in (Behar et al, 2004) indeed belonged to rather recent ancestors. Even in a 10-marker format they still form flat branches on the left- and right-hand side, respectively, in Fig. 27, that points at their rather recent origin. The first one, as it was determined from 37- and 66-marker haplotypes (see below) is still a superposition of haplotypes derived from ancestors who lived 6,800±400 and 750±200 years BP, and an ancestor of the second haplotype (in the 6- and 12-marker formats) lived 1,000±200 years BP.

A pair of alleles X-Y in the first haplotype is 16-18 in the ancient (6,800-year old) haplotype or 17-18 in the recent (750-year old) haplotype, as follows from our data shown in this article. A pair of X-Y in the second haplotype is 17-18 (see above).

Two old branches at the bottom of Fig. 27 have ancestral haplotypes

13-12-24-10-11-13
13-24-13-10-X-Y-11-12-13-13-11-31

and

13-12-24-10-11-14
14-24-13-10-X-Y-11-12-11-12-11-31

The first base haplotype (lower right-hand-side area in Fig. 27) came from an ancestor who lived 3,500 years BP. It matches 6-marker haplotypes from two old ancestors, 6,800±400 and 750±200 years BP, but has a unique 10-marker haplotype, which has two mutations away from both of them.

The second base haplotype (lower left-hand side area in Fig. 27) is derived from an ancestor who lived 2,100 years BP and also has a unique 10-marker haplotype. Clearly, 10-marker haplotype data can be considered only as rough estimates compared with 37- and 66-marker haplotypes.

12-marker haplotypes ("commercial" and "club" sets)

12-marker haplotype trees for a "commercial" set (YSearch database) and the "club" set of Jews of E1b1b* haplogroup are shown in Figs. 28 and 29.

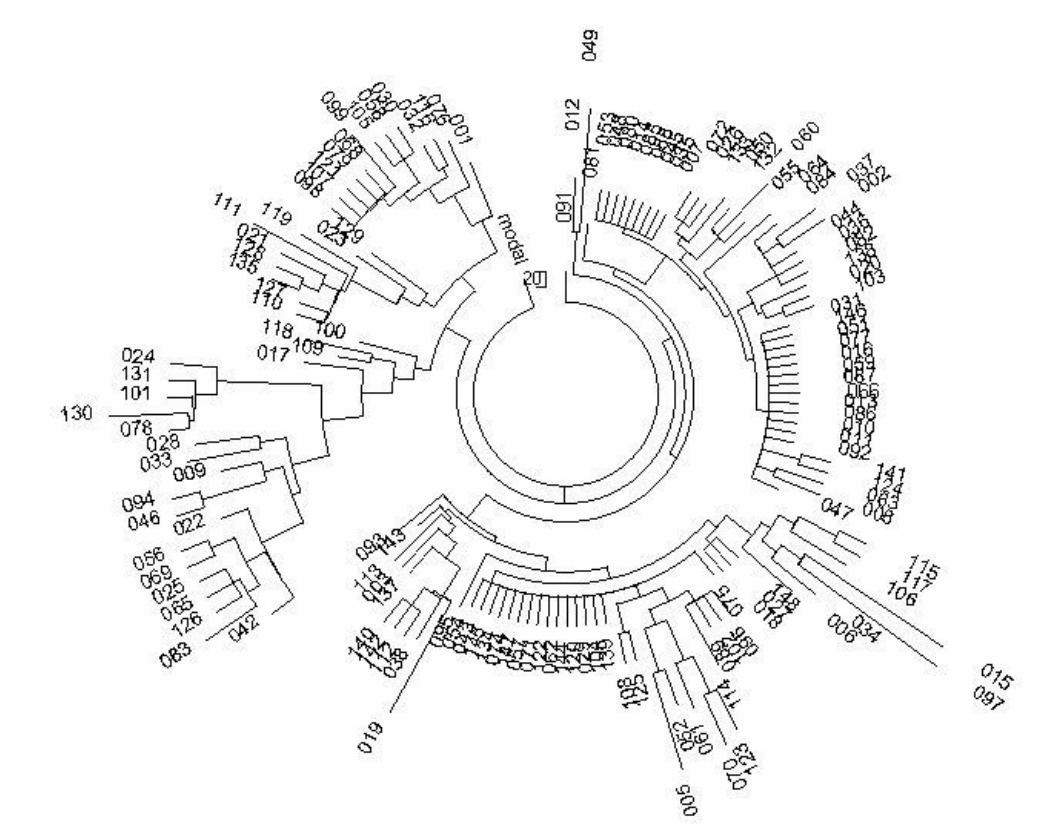

Figure 28. The 12-marker haplotype tree for 140 Jewish haplotypes of haplogroup E1b1b. A "commercial" set, extracted from YSearch database

One can see that some rather recent ancestors were so recent, that even in a 12-marker format their descended branches of haplotypes are still flat and consist of a series of identical haplotypes. One of these ancestral haplotypes (at the bottom of Fig. 28, the "commercial set") is:

13-12-24-10-11-13
13-24-13-10-17-18-11-12-12-13-11-30

This is a combined ancient haplotype of 6,800±400 years old and a "younger" haplotype of 750±200 years old. It explains why the DYS#385a allele is 17, and not 16 as in the ancient haplotype. There are few (if any) of the ancient 12-marker haplotypes left in the flat branch, and haplotypes from more recent ancestors prevail, with the allele of 17. This combined ancient and the more recent

haplotype is also present in the "club set" on the top of Fig. 29, as a base haplotype of the whole tree. Its content is 11% in the "commercial" tree and 9% in the "club" tree. Again, the trees are pretty much similar.

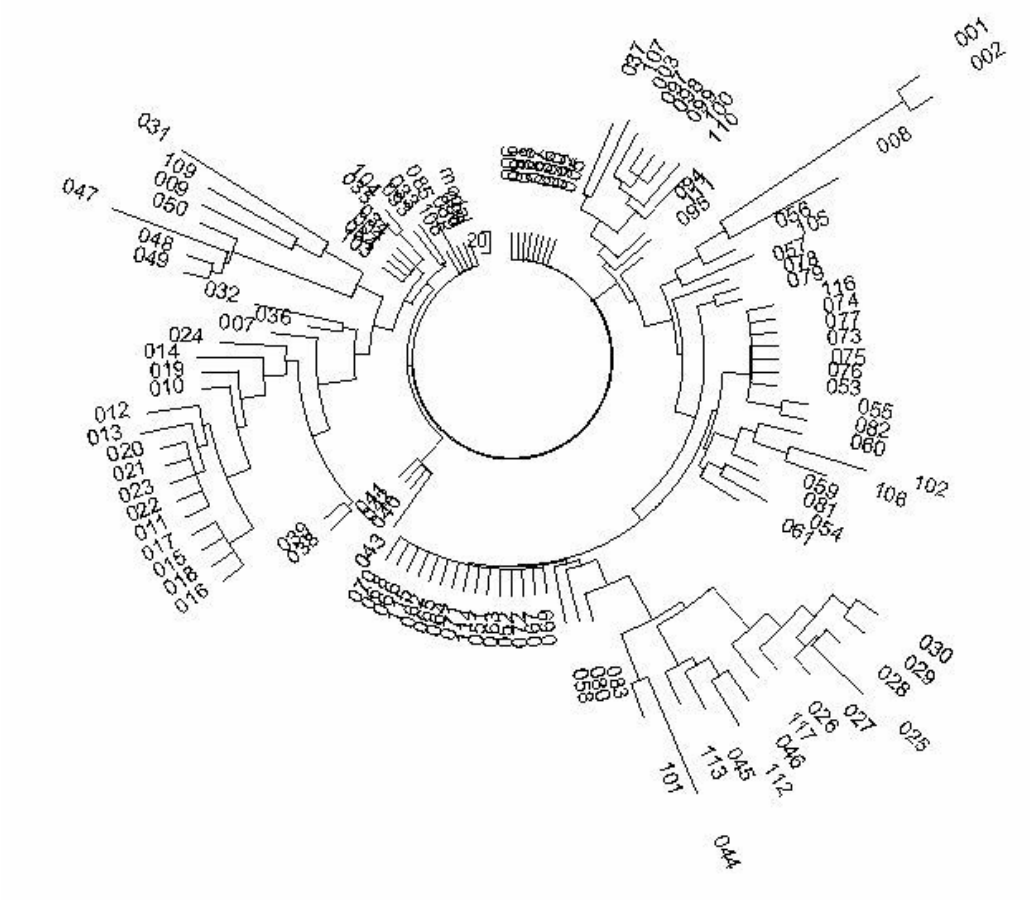

Figure 29. The 12-marker haplotype tree for 117 Jewish haplotypes of haplogroup E1b1b*. Haplotypes assignment to haplotypes is shown in Fig. 26. A "club" set (the Jewish E3b Project).

Another "flat branch" haplotype, which belonged to a rather recent ancestor who lived 1,000±250 years BP, is

13-12-25-9-11-14
14-25-13-9-17-18-11-12-12-13-11-30

It is located as 15 identical 12-marker haplotypes at the bottom of Fig. 29 (the "club set") and as 9 identical haplotypes at the top of Fig. 28 (the "commercial set"), around 10% of all haplotypes in each case.

A base haplotype with the same 6-marker sequence and just one mutation away from the above 12-marker haplotype

13-12-25-9-11-14
14-25-13-9-17-18-11-12-12-13-11-31

is located on the right-hand side in Figs. 28 and 29 (a "commercial" and "club" set) as 14 and 7 identical 12-marker haplotypes, that is 10% and 6% of the total amount of haplotypes, respectively. These are descendants of a rather recent ancestor, who lived about 1,000 years BP, around 11th century AD.

Finally, three distinct ancient branches are located in the lower left corner in Fig. 5 (the "commercial set"), and in the lower left and right corners in Fig. 29 (the "club set"), with their common ancestors in each case lived between 4,000 and 7,000 years BP. A more detailed consideration of these branches is done in next sections.

37-marker haplotypes

There are no 37-marker haplotypes published in peer-review journals, let alone Jewish 37-marker haplotypes. We will consider two sets available from databases, namely a "commercial" series of E1b1b (with subgroups) haplotypes (YSearch data base) and a "club" set of E1b1b (with subgroups) haplotypes. The respective trees composed of 57 and 77 of these haplotypes are shown in Figs. 30 and 31.

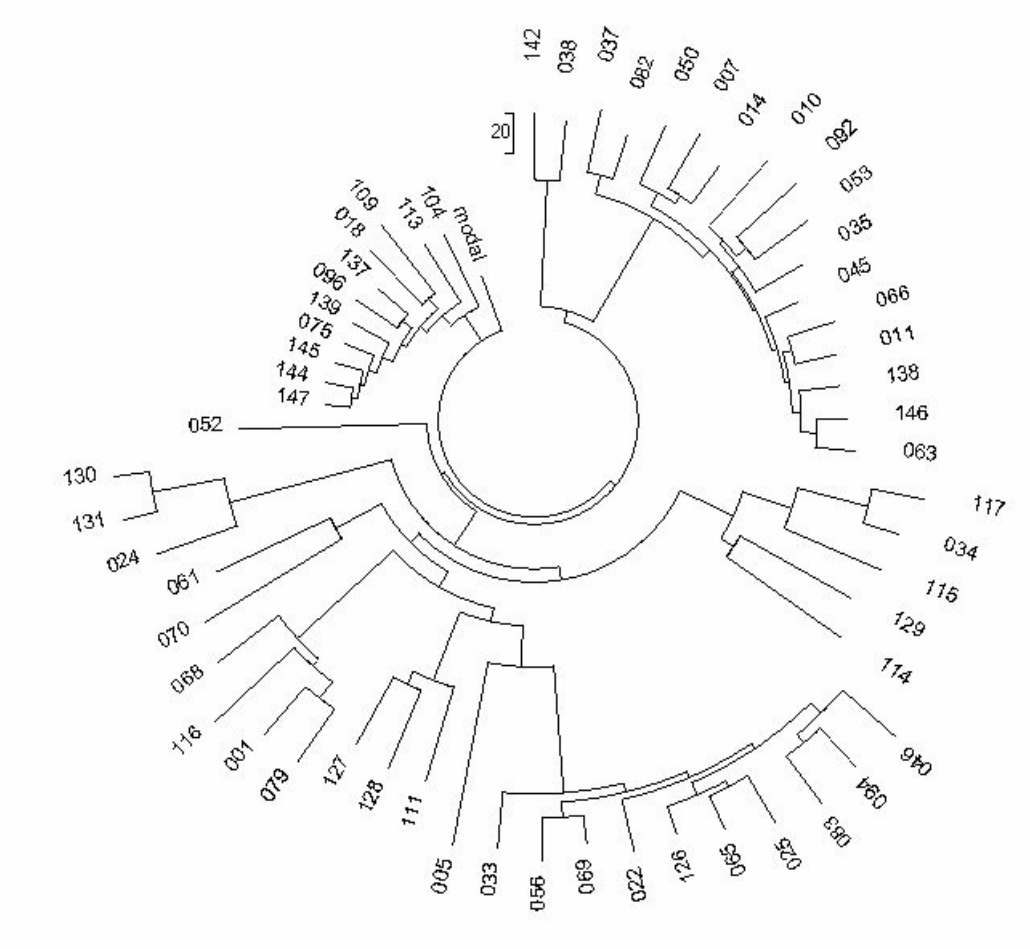

Figure 30. The 37-marker haplotype tree for 57 Jewish haplotypes of haplogroup E1b1b. A "commercial" set (YSearch database)

Each tree consists of three distinct branches of haplotypes, which in turn contain a number of smaller branches.

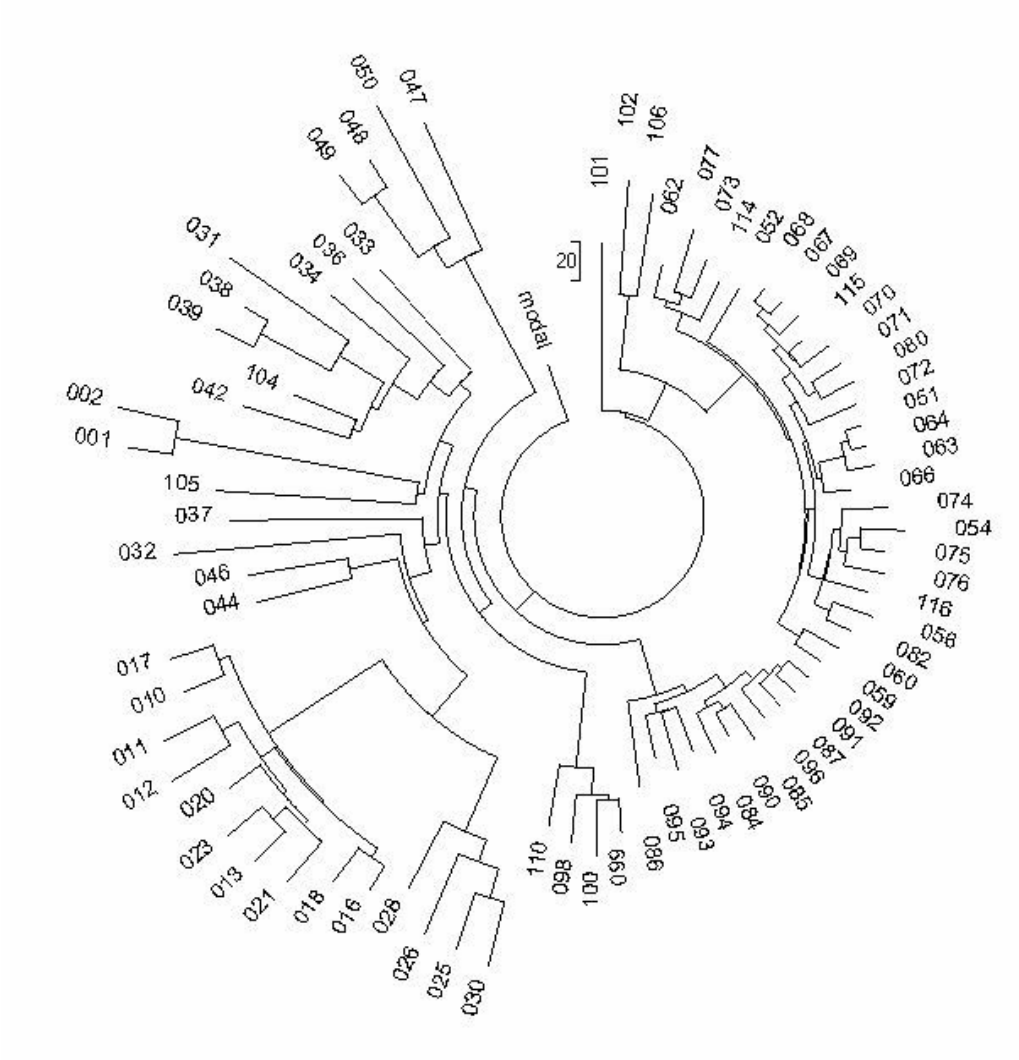

Figure 31. The 37-marker haplotype tree for 77 Jewish haplotypes of haplogroup E1b1b*. Haplotypes assignment to haplotypes is shown in Fig. 26. A "club" set (the Jewish E3b Project).

One of those three principal branches is an ancient branch (6,800±400 years BP), loose and "fluffy", stretching away from the "trunk" of the tree. It takes a lower left corner in Fig. 30 (the "commercial set") and an upper left corner in Fig. 31 (the "club set"). Besides, each of the both sets contains at least three rather recent, flatter branches (1,000±250, 975±100, and 750±200 years BP), some of them are derived from the ancient branch (the second one), and some are protruded from the "trunk" of the tree or close to it (the first and third one), pointing at a different

genesis of mutations. They might have originated from the same ancient ancestor, however, starting from a different mutation and forming a separate branch. As an alternative, they might have originated from a practically unrelated, non-Jewish individual, who belonged to the same E1b1b haplogroup, and furthermore, to the same haplotype as the ancient Jewish ancestor of the same haplogroup. As it was mentioned above, the most frequent 6- and 12-marker haplotype is identical to the ancient 6- and 12-marker ancient Jewish haplotype. In fact, the "mutation-minimized" 25-marker E1b1b* haplotype (out of 800 haplotypes in YSearch database) is only two mutations away from the ancient Jewish E1b1b* haplotype.

An ancient branch (6,800±400 years to a common ancestor)

From half to two-thirds of all Jewish E1b1b haplotypes in the both series of haplotypes are derived from an ancient common ancestor, who had the following haplotype, in a 6- and 12-marker format (the 37-marker haplotype is shown in the Table):

13-12-24-10-11-13

13-24-13-10-16-18-11-12-12-13-11-30

It is identical to the most frequent haplotype in E1b1b* haplogroup, in which Jews account for about 10%:

In the "commercial set" (Fig. 30), the 29-haplotype "fluffy" branch (the lower half of the tree) revealed 154, 299 and 540 mutations in 12-, 25- and 37-marker haplotypes, that corresponds to 284, 289 and 261 generations, respectively, to the common ancestor. In the "club set" the ancient, "fluffy" 48-haplotype branch, having exactly the same base haplotype as shown above, results in 230, 475 and 891 mutations in the branch, that corresponds to 249, 274 and 260 generations to the common ancestor. On average, these six data points give 270±15 generations, that is 6,800±400 years time span to the common ancestor.

In the "commercial set", the majority of present-day descendants of this ancient ancestor were assigned to E1b1b haplogroup with no subgroups, however, five of the 29 current descendants have E1b1b1 subgroup (haplotypes 111, 114, 116, 117, and 126 in Fig. 30), five more have E1b1b1a (haplotypes 79, 115, 127, 128, 129), and two have E1b1b1b2 (haplotypes 130, 131). Obviously, it points at some irregularities with the typing of subgroups. Whatever is the reason(s), even 37-

marker haplotypes are unable to resolve different typing, since haplotypes of different subgroups sit on the same branches next to each other.

The same pattern of subgroups was observed in the "club set" as well. In the 48-haplotype ancient tree 29 were E1b1b1 haplotypes, the rest was a mix of E, E1b1b1a, E1b1b1a1, E1b1b1a2, E1b1b1b2, E1b1b1a3, and E1b1b1c1. Only one rather recent branch (975±100 years BP) consists of only E1b1b1 haplotypes, other were mixed ones.

Even 66-marker trees, which will be considered below, did not resolve subgroups on their branches. The ancient branch in both 66-marker "commercial" and "club" trees contained E, E1b1b, E1b1b1, E1b1b1a, E1b1b1a1, E1b1b1ba, E1b1b1a2, and E1b1b1c1 haplotypes.

However, there are some regularities even in that mix of subgroups. A 19-haplotype "ancient" branch in the upper left corner of the "club set" (Fig. 31) contains all 10 non-E1b1b* subgroups, namely E (haplotypes 001 and 002), all E1b1b1a* (031, 037, 038, 044 and 046), and E1b1b1b2 (047-049). The rest was E1b1b1. There was not a single E1b1b1c1 on this branch, out of seven in the set.

This 19-haplotype branch was a part of the wide-branch ancient haplotype. It had exactly the same ancient base 6- and 12-marker base haplotype, shown above, and had 85, 199 and 378 mutations in its 12-, 25- and 37-marker 19 haplotypes. This translates into 269±35 generations, that is 6,700±900 years to the common ancestor. This fits pretty well to 6,800±400 years BP, determined with 48 haplotypes of the "club set" and 29 haplotypes of the "commercial set" (see above). This branch, being reduced to 12 haplotypes in the 66-marker tree (Fig. 33, the left-hand side) produced 58, 131, 234 and 372 mutations in its 12-, 25-, 37- and 66-marker haplotypes, that translates into 252, 311, 277 or 277 generations, respectively, to the common ancestor, and averages to 279±24 generations, or 7,000±600 years to the common ancestor in the "club" set.

Last, but not least, a similar branch but in the "commercial" set, reduced to 9 haplotypes in the 66-marker tree (Fig. 32, at the bottom), has 41, 84, 164, and 254 mutations, and results in 234, 255, 254, or 247 generations to a common ancestor. Averaging all these eight figures for generations, we obtain 263±24 generations, that is 6,600±600 years BP. All these figures group nicely around 6,800±400 years to the common ancestor, which might be an ultimate ancestor of all presently living Jews of haplogroup E1b1b*.

Some rather recent ancestors (during the Diaspora times)

About half of Jews of E1b1b* haplogroup descended from a few common ancestors, who lived about 1,000±250, 975±100, and 750±200 years BP, and initiated different lineages. These lineages in 6-marker format are derived from the following base (ancestral) haplotypes:

13-12-25-9-11-14
14-12-24-10-11-13
13-12-24-10-11-13

All of them are apparently derived from a common ancestor who lived about 6,800±400 years ago and had a 6-marker haplotype, which is the most frequent haplotype in E1b1b* haplogroup

13-12-24-10-11-13

A recent common ancestor, 1,000±250 years BP (around 10th century AD)

From a quarter to a third of all present day Jews in haplogroup E1b1b* have descended from this common ancestor. The respective haplotypes in both the "commercial" and "club" set form a rather flat branch in the upper right corner of Figs. 30 and 31, containing 15 and 26 haplotypes, respectively. In the both cases the base (ancestral) haplotype in a 6- and 12-marker format is as follows:

13-12-25-9-11-14

14-25-13-9-17-18-11-12-12-13-11-30

These haplotypes on the branches have very few mutations, which itself shows that their ancestor lived rather recently. For example, there is only one mutation in all 15 six-marker haplotypes and in all 26 six-marker haplotypes on these branches, that is only one mutation per 90 and 156 markers, respectively. In their 12-marker haplotypes – on average – is less than one mutation per haplotype with respect to the base haplotype.

Consideration of mutations in 12-, 25- and 37-marker haplotypes showed that a common ancestor of the branch lived 37, 66 or 55 generations BP, respectively ("commercial set") and 28, 46 and 41 generations BP ("club set"). Similar calculations for the respective 66-marker haplotypes gave 34, 41, 37, 35, 18, 40, 38, and 37 generations. An average of these 13 figures (18 generations here is an

obvious outlier, since the 66-marker branch had only few mutations in the first 12-marker panel, out of line with the other panels) gives 41±10 generations, that is approximately 1,000±250 years time span to the common ancestor.

A majority of haplotypes in this branch belong to E1b1b subgroup with a few E1b1b13a haplotypes ("commercial set"), or a third of them are of E1b1b1c1 haplotypes ("club set").

A recent common ancestor, 975±100 years BP (~11 century AD)

About one-seventh of all present day Jews in haplogroup E1b1b* descended from this common ancestor. Their haplotypes are located in a pretty much flat branch in the lower right corner of Fig. 30 ("commercial set") and the lower left corner of Fig. 31 ("club set"). Their ten 6-marker haplotypes in each of the branch

14-12-24-10-11-13

are all identical to each other, there are no mutations at all. Their 12-marker haplotypes have – on average – less than one mutation per haplotype in each case

13-24-14-10-16-17-11-12-13-14-11-32

Mutations in all 37 markers gave the time span to the common ancestor of this branch of 34, 43 or 42 generations (for 12-, 25- and 37-marker haplotypes) in the "commercial set" and 34, 38 or 45 generations for the "club set", that gives on average 39±4 generations to their common ancestor, or 975±100 years BP.

This branch is totally composed of E1b1b1 haplotypes in both "commercial" and "club" sets, and it is the only one completely uniform in terms of subgroups.

A recent common ancestor, 750±200 years BP (around 13th century AD)

About one-sixth of all present day Jews in haplogroup E1b1b* descended from this common ancestor. The respective branches in Figs. 30 and 31 are the most tight (particularly in the "commercial" set, Fig. 30) and located the closest to the trunk of the tree. In both cases, their eleven 6-marker haplotypes in the branch

13-12-24-10-11-13

are the same as those of the common ancient ancestor of half ("commercial set") or two-thirds ("club set") of the individuals whose haplotypes are considered in Figs. 30 and 31. All 11 haplotypes in a 6-marker format have only one mutation in each set, that is one mutation per 66 markers. In a 12-marker format there were 7 mutations per 132 markers in each of the branch, with respect to the base haplotype shown immediately below. This ancestral 12-marker haplotype is as follows (the respective 37-marker base haplotype is shown in Table 1):

13-24-13-10-17-18-11-12-12-13-11-30

This haplotype has only one mutation compared to the ancient haplotype of 6,800±400 years BP, showing 17 in place of 16 for the ancient, ancestral haplotype in DYS#385a.

Overall, eleven 37-marker haplotypes have 38 and 35 mutations per 407 markers in the "commercial" and "club" sets, respectively, that is less than 0.009 mutations per marker, though 37-marker haplotypes mutate faster than 12-marker haplotypes. Mutations in all panels of 37-marker haplotypes give 27, 39, or 40 generations to the common ancestor for the branch in the "commercial" set, and 27, 39 or 37 generations for the "club set". This, along with the respective mutations for both branches in the 66-marker trees, which amounted to 21, 24, 32, 29, 8, 18, 23 and 19, and consider 8 generations as an outlier (there were few mutations in the first 12-panel markers in the "club set"), results in an average of 30±8 generations. Thus, the common ancestor of this branch lived only 750±200 years BP.

From half to two-thirds of this particular branch consists of haplotypes of E1b1b1c1 subgroup (former E3b3a, SNP M34) in the both "commercial" and "club" sets. The other half were E1b1b or E1b1b1 subgroups. Again, despite the tightness of the branch, even 37- and 66-marker haplotypes cannot provide resolution between those subgroups.

Since all these three recent ancestors descended likely from a common ancient ancestor, the pattern of E1b1b* ancestry is apparently related to passing of certain "bottlenecks" in the Jewish population of this haplogroup. Invasions from non-Jewish E1b1b* individuals could not be excluded, since Jewish and non-Jewish haplotypes in this particular haplogroup are so close in sequences. For example, only one mutation away from the above ancient haplotype makes "Harris Modal Haplotype" or "E3b1a2 Modal Haplotype" in YSearch database.

66-marker haplotypes (the YSearch and "club" sets)

The 66-marker haplotype trees are shown in Figs. 32 and 3. Typically, 67-marker haplotypes are determined by testing companies. However, in many haplotypes being considered in this section the marker DYS#425 was missing, hence, 66-marker haplotypes were considered here. The respective base (ancestral) 66-marker haplotypes for the Jews of E1b1b* haplogroup are provided in the Table.

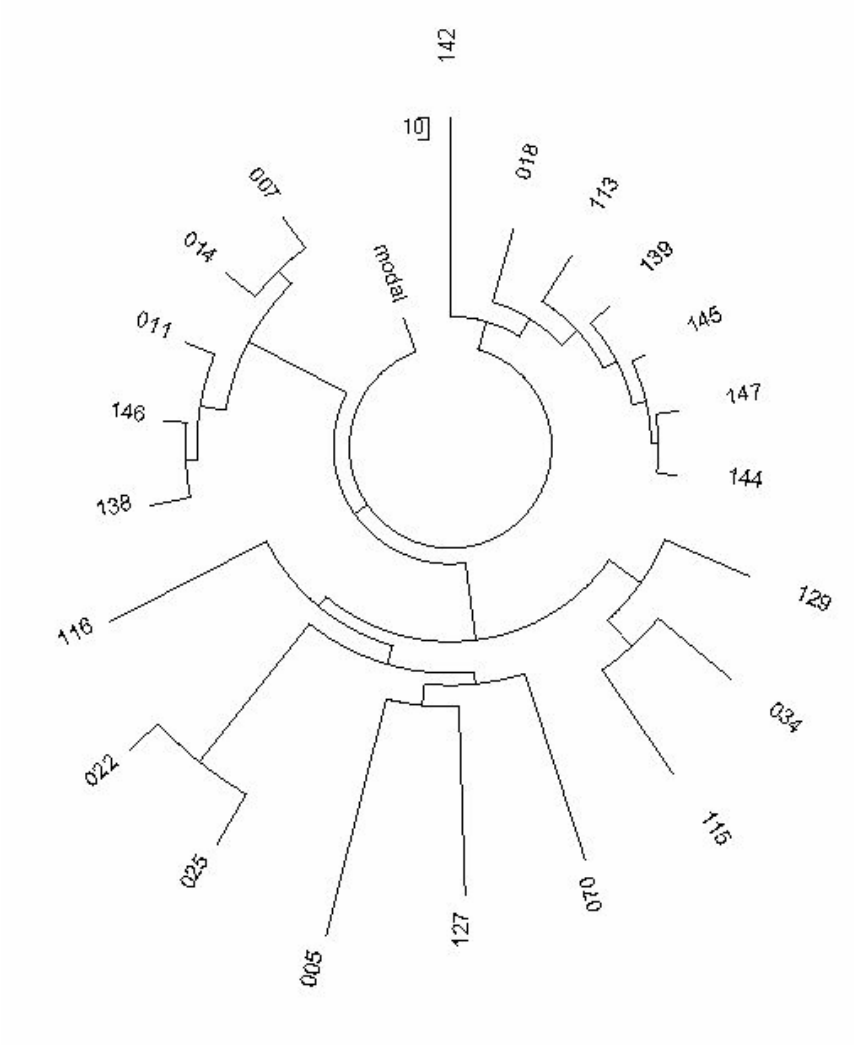

Figure 32. The 66-marker haplotype tree for 21 Jewish haplotypes of haplogroup E1b1b. A "commercial" set (YSearch database)

The 21-haplotype YSearch data based tree (Fig. 32) has three distinct branches. The 32-haplotype "club set" tree (Fig. 33) has two branches, one is very similar with that in Fig. 32, another one is a combination of two branches in Fig. 32. The trees, albeit similar in some details, are not he same, and sizes of their branches are different in the both trees.

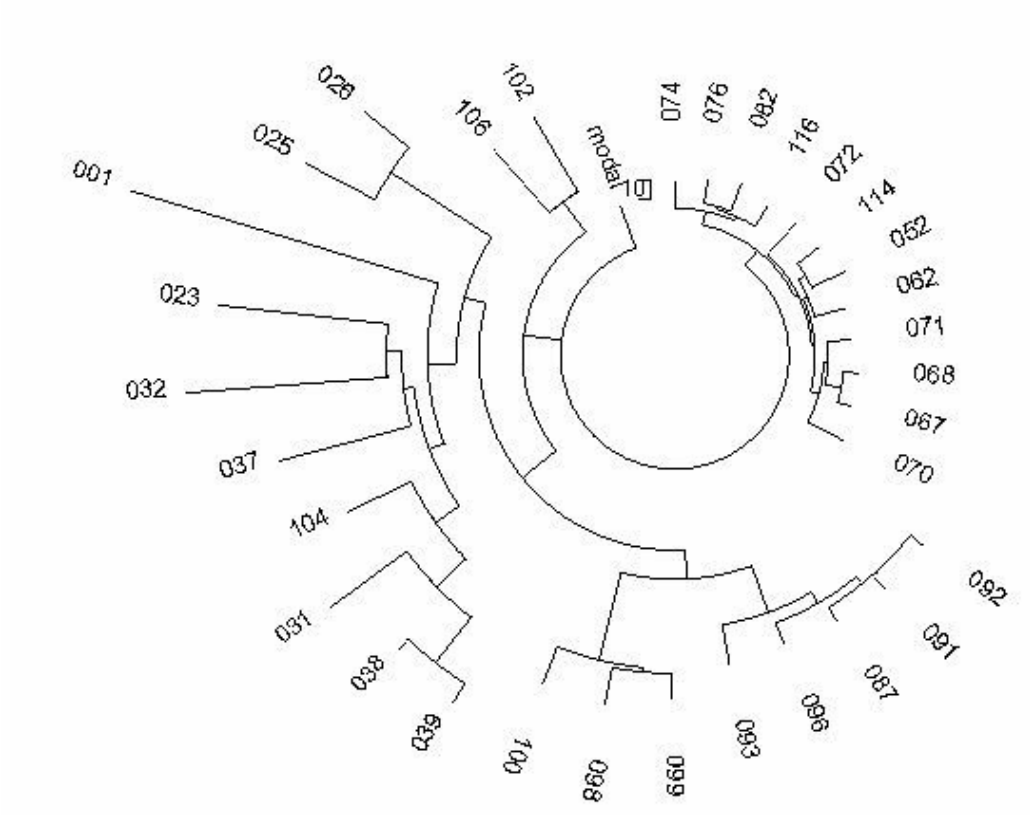

Figure 33. The 66-marker haplotype tree for 32 Jewish haplotypes of haplogroup E1b1b*. Haplotypes assignment to haplotypes is shown in Fig. 26. A "club" set (the Jewish E3b Project).

The both trees are based on the same ancestral haplotypes, which have been discussed in the preceding section. The last panel, of 38-66 markers, added more mutations in each case, however, these mutations fit pretty well into above calculations of common ancestors. That is why only a brief overview of the both trees and their branches is given in this section.

The ancient ancestor

In a similar manner as it was described above, 40% to 50% of haplotypes in the both sets are derived from an ancient common ancestor, who lived 6,800±400 years BP. The respective branches are those at the bottom of Fig. 32 ("commercial set") and on the left in Fig. 33 (the "club set"). They are obviously the most loose ones, and have base (ancestral) 6- and 12-marker haplotypes as follows

13-12-24-10-11-13
13-24-13-10-16-18-11-12-12-13-11-30

Since there are only 9 and 12 haplotypes in these two branches, respectively, time spans to their common ancestors can be considered with a lower reliability compared to more numerous 37-marker haplotypes (29 and 48 haplotypes, respectively). However, the number of mutations in 12-, 25-, 37-, and 66-marker haplotypes in said branches (there are 41, 84, 164, 254 and 58, 131, 234, 372 mutations in said 9- and 12-haplotypes branches) results in 234, 255, 254 or 247 generations to a common ancient ancestor for the "commercial set" and 252, 311, 277 and 277 generations to the common ancestor in the "club set", that in turn gives the average 263±24 generations, that is 6,600±600 years BP. It fits pretty well 6,800±400 years BP, calculated above.

Rather recent ancestors

1,000±250 years BP

Present day descendants of an ancestor who lived 1,000±250 years BP provided 66-marker haplotypes which form flat branches in the upper left corner of Fig. 32 (the 5 haplotype branch in the "commercial set"), and the upper right corner of Fig. 33 (the 12-haolotype branch in the "club set"). 25% to 40% haplotypes in these two sets are derived from this common recent ancestor, who had the following (ancestral) haplotype in 6- and 12-marker formats

13-12-25-9-11-14
14-25-13-9-17-18-11-12-12-13-11-30

Mutation counts and results of calculations of a time span to this ancestor are given in the preceding section.

975±100 years BP

In the 66-marker trees, both "commercial" and "club" sets retained only a few haplotypes of this branch. In fact, these are only two haplotypes (022 and 025) derived from this common ancestor in Fig. 32, and only one haplotype (023) in Fig. 33.

In the 66-marker tree of the "commercial" set of haplotypes (Fig. 32) the two haplotypes sit next to each other and are almost identical. Their 6- and 12-marker haplotypes are:

14-12-24-10-11-13
13-24-14-10-16-17(18)-11-12-13-14-11-32

Hence, they can provide an estimate of 66-marker haplotypes of this branch (Table). Between the two, they have only 8 mutations per 66 markers, that approximately correspond to 31 generations to a common ancestor, that is about 800 years time span. This only shows that their common ancestor was a rather recent indeed, and the figure calculated on only two 66-marker haplotypes are not too distant from that calculated on a more statistically appropriate base.

750±200 years BP

About one-quarter of all haplotypes in the "commercial" 66-marker tree are derived from this common ancestor. The respective 6-haplotype 66-marker branch in Fig. 32 is located in the upper right corner, and the 5-haplotype branch in Fig. 33 – in the lower right corner. Their base (ancestral) 6- and 12-marker haplotypes are as follows,

13-12-24-10-11-13
13-24-13-10-17-18-11-12-12-13-11-30

This 12-marker haplotype has only three mutations in the 6-haplotype "commercial set", and not a single mutation in the 5-haplotype "club set", that is per all 60 markers. This alone shows how recent is the common ancestor indeed.

The base (ancestral) 66-marker haplotype is shown in the Table. It is exactly the same in "commercial" and "club" sets. Mutation counts and results of calculations of a time span to this ancestor are given in the preceding section.

In summary, the Jews of haplogroup E1b1b* have started their lineage from a common ancestor who lived 6,800±400 years BP, and apparently acquired his

haplotype from a general pool of similar haplotypes which are (and have been) the most widely spread haplotypes among those in the Middle East and Europe, who belonged to this haplogroup. Around 10-13th centuries AD this ancient haplogroup has split into three main lineages, and nowadays descendants of them present about half of all the Jews of haplogroup E1b1b*.

Part III

The Jews of Haplogroups R1a1, Q, R2, G*, I*, K and T

The Jews of Haplogroup R1a1

On different accounts, haplogroup R1a1 among the Jews occurs with 7% to 10% probability. I was able to identify at least 44 Jewish haplotypes of R1a/R1a1 combined haplogroup, that was 2.6% of the total haplotypes in YSearch database. Haplogroup R1a1 (SNP M17) includes haplotypes descended from an ancient tribe which allegedly came onto a land between Dnepr-river and Volga-river, and, possibly, Ural-river, in the South of nowadays Russia and Ukraine, about 10-15 thousand (or maybe around 8 thousand) years ago. Where the people of the R1a1 tribe came to this area from is still a matter of conjectures. Some say from the "Ukrainian refuge" (to which there is no data), some say from the East, from Asia, from South Siberia.

As it was shown in the preceding paper in this issue, the most ancient common ancestor of R1a1 was identified in the Balkans, in the former Yugoslavia, and he lived about 8,900 years BP. Apparently in about two thousand years after that time, bearers of Balkan R1a1 moved East, through present day Ukraine, Russia, Kazakhstan, Southern Siberia, and then to China and, about 3,400 years BP, to India.

According to present-day views (rather, conjectures) R1a1 – apparently, in the southern steppes of modern Russia and Ukraine – had domesticated horses, invented chariots, began horse-riding and, hence, very fast – on historical measures – had spread in the East to Eastern Siberia, in the West to Central Europe and to Atlantic ocean, in the North to Scandinavia, in the South to Asia Minor, Apennine peninsula and other Mediterranean coastlines and islands, the Middle East and Egypt, and then entered from the West to Persia (Iran) and Afghanistan, and from the North to India under the name of Aryans, as it had been written in the Vedas.

Thereby the R1a1 tribe and its descendants has spread haplotypes of R1a1 haplogroup onto huge territories, and their maximum frequencies are currently observed in the Eastern Europe: in Russia (up to 70% of the population), Ukraine, Belorus, Poland (up to 50-60%), and on the land of the Eastern wing of the tribe some 4000-3300 BP – in Kyrgyzstan, Tadzhikistan (up to 50-70%), in the Pamir Valley, along the path of the Aryans to India (a small Ishkashim group, up to 70%), and in India itself (about 20-30% of all Indians bear R1a1 haplogroup, particularly those in the North of India, the Kashmir region, as well as in some rather remote tribes, such as Chenchu, which have acquired the haplogroup some 2,900 years BP, as it was shown in the preceding paper in this issue). Haplogroup R1a1 in the Western Europe is observed from about 4-8% of the population in Holland to about 13-20% of that in Denmark, Sweden, Iceland.

Three sources of R1a1/R1a haplotypes (these two haplogroups could not be resolved up to 67-marker haplotype trees and branches, hence, they are considered here as a common pool of R1a1 haplotypes) employed in this paper are: (a) the publication of 2003 (Behar et al, 2003) (the "scientific set"), (b) a data set from YSearch database (2008) (the "commercial set"), and (c) a collection of haplotypes of the Jews who belong to haplogroup R1a/R1a1 and form the "Ashkenazi-Levite DNA Project" (see Ref.) (the "club set").

The most widespread haplotypes in Eastern Europe in haplogroup R1a and R1a1 in 6- and 12- marker formats are

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

It is just one mutation away (in DYS#391) from the Russian R1a1 haplotype shown in the preceding paper in this issue. Again, in haplogroups R1a and R1a1 in YSearch data base these haplotypes are exactly the same. These are referred not to Jewish haplotypes, but to all available haplotypes from said haplogroups, total number of haplotypes 1890 and 672, respectively (YSearch database).

The 6- and 12-marker base R1a/R1a1 haplotypes shown above are the exact match of the base haplotypes in all three Jewish haplotype sets – "scientific", "commercial", and "club".

For the record, the first four the most popular R1a/R1a1 6-marker haplotypes in the total YSearch database are

16-12-25-10-11-13

15-12-25-11-11-13
15-12-25-10-11-13
16-12-25-11-11-13

Before we go into details, a summary of findings in this section can be condensed as follows:

- A single common ancestor of R1a1 Jews lived about 1,350±300 years BP, around the 7th century AD. A "single" in terms of DNA genealogy is equivalent to a group of individuals who lived at about the same time and have the same base (ancestral) haplotype. They could be brothers, for example, or an uncle and his nephew.
- His ancestral lineage gave an offshoot some 700 years later, in about the middle of 14th century, allegedly by passing a haplotype bottleneck, such as when Jews of R1a1 haplogroup as well as other desperate people that time fled from black plague in Central Europe to East Europe.

Let us consider these haplotypes, branches and haplotype trees in more detail.

6-marker haplotypes ("scientific", "commercial", and a "club" sets)

Figs. 34, 35 and 36 show 6-marker haplotype trees constructed of 42 haplotypes published in (Behar et al, 2003), 44 haplotypes of R1a1/R1a Jews extracted from YSearch database, and 100 haplotypes of "Jewish R1a/R1a1 Project", respectively. It is noticeable that all the three set are somewhat different, however, as it will be shown later, this is a feature of R1a1 haplotypes, which accumulate mutations mainly between 25 and 67 markers. It means that 6- and 12-marker panels give a reduced number of mutations, hence, an underestimated time span to a common ancestor. More extended haplotypes provided better statistics, and resulted in reproducible figures from all the three sources of haplotypes, listed above.

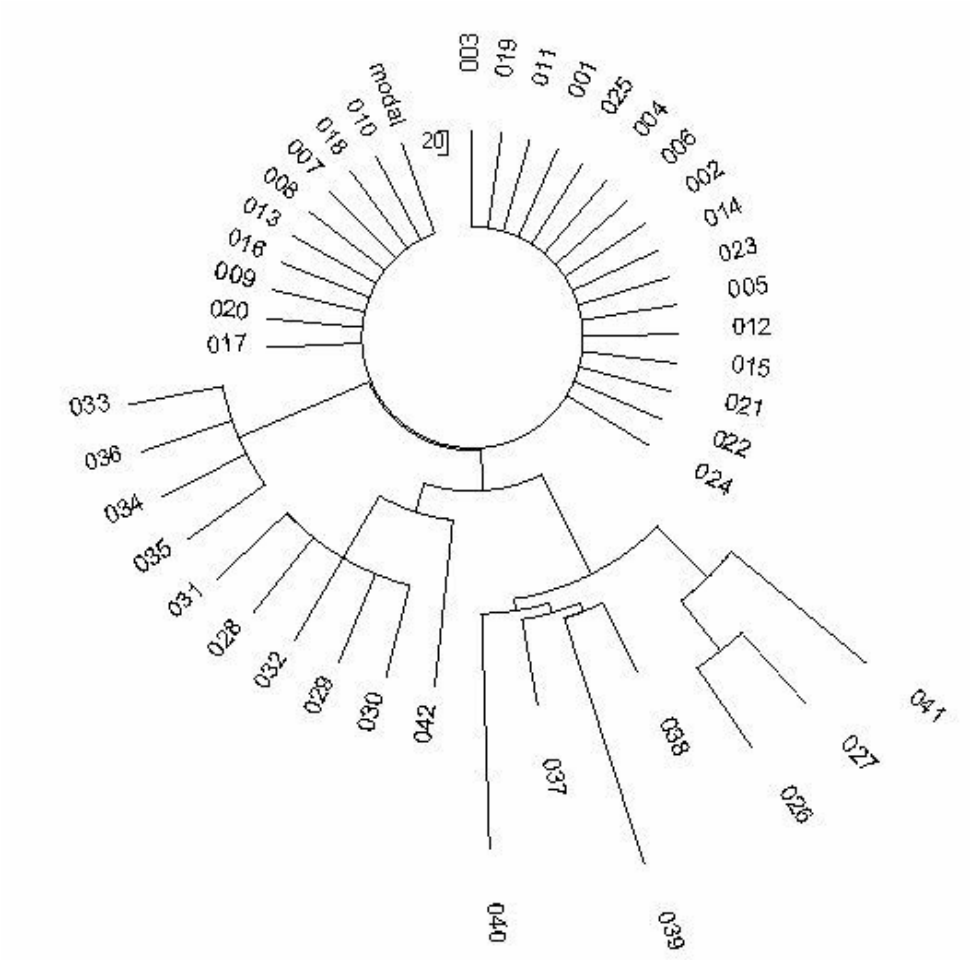

Figure 34. The 6-marker haplotype tree for 42 Jewish haplotypes of haplogroup R1a1. A "scientific" set (Behar et al, 2003)

At the same time, all the three sets clearly point at a single ancestor for each set of data. As it was described in the preceding paper in this issue, the principal criterion of a single ancestor in a set of haplotypes is the same number of generations calculated from a fraction of residual base haplotypes, on the one hand, and that calculated from a number of mutations in the same set of haplotypes. Naturally, this criterion is applicable only to haplotype sets derived from rather recent ancestors, since only in those cases there are at least several identical, base haplotypes in the set. Present day haplotypes derived directly from ancient ancestors are commonly mutated.

The tree in Fig. 34 (the "scientific" set of haplotypes) has 25 base (ancestral) haplotypes, identical to each other

16-12-25-10-11-13

in the 42-haplotype set. Hence, an estimate gives ln(42/25)/0.0096 = 54 generations to a common ancestor. The whole set contains 22 mutations with respect to the above base haplotype. It gives 22/42/0.0096 = 54.6 generations, if to be absurdly precise. This indicates that the Jews of R1a/R1a1 haplogroup descended indeed from a single common ancestor who lived about 1,400 years BP, in the 7th century AD, during the Diaspora.

Behar et al (2003) have shown that 31 of those present day 42 Jewish haplotypes belonged to Askenazi-Levites (74%). Among the other 11 Jews of haplogroup R1a1, one was Ashkenazim-Cohen, four Sephardim-Cohen, one Sephardim-Levite, four Askenazim-Israelite, and one Separdim-Israelite. It is very likely that that the penetration of haplogroup R1a1 into Jewry happened via Levites and after a separation of Ashkenazim and Sephardim into their populations. An alternative explanation would be that only a Levite has passed a genealogical bottleneck, hence, R1a1 Levites as his descendants. This explanation is elaborated below.

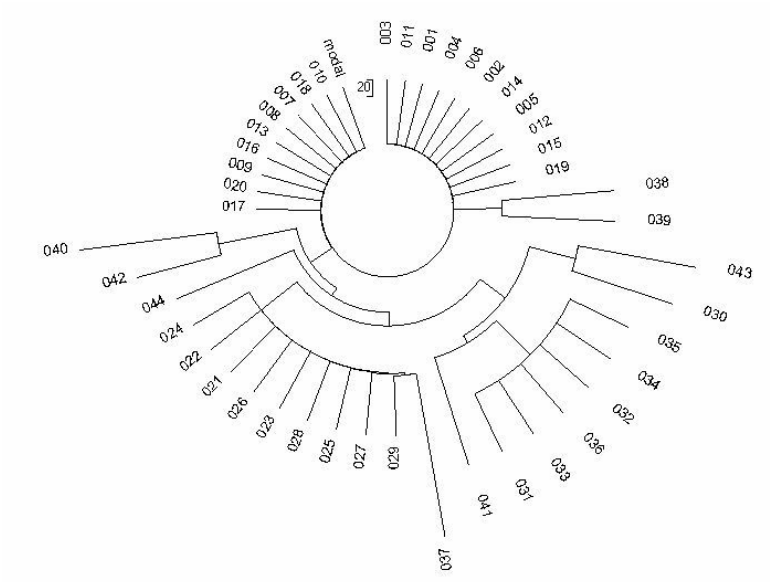

Figure 35. The 6-marker haplotype tree for 44 Jewish haplotypes of haplogroup R1a/R1a1. A "commercial" set (YSearch database)

The “commercial” set (Fig. 35) shows a smaller fraction of the base haplotypes, though the 6-marker haplotypes are exactly the same as shown above, in the “scientific” set. It has 20 base haplotypes from the total of 44, that is an estimated number of generations to a common ancestor is ln(44/20)/0.0096 = 82. However, a number of mutations (29) in the set gives 29/44/0.0096 = 69 generations. Hence, the 6-marker set is somewhat disbalanced in terms of mutations and residual base haplotypes, and points at more than one common ancestor of the haplotypes in the set. Indeed, the 12-marker set from the same source shows exactly that (the next section).

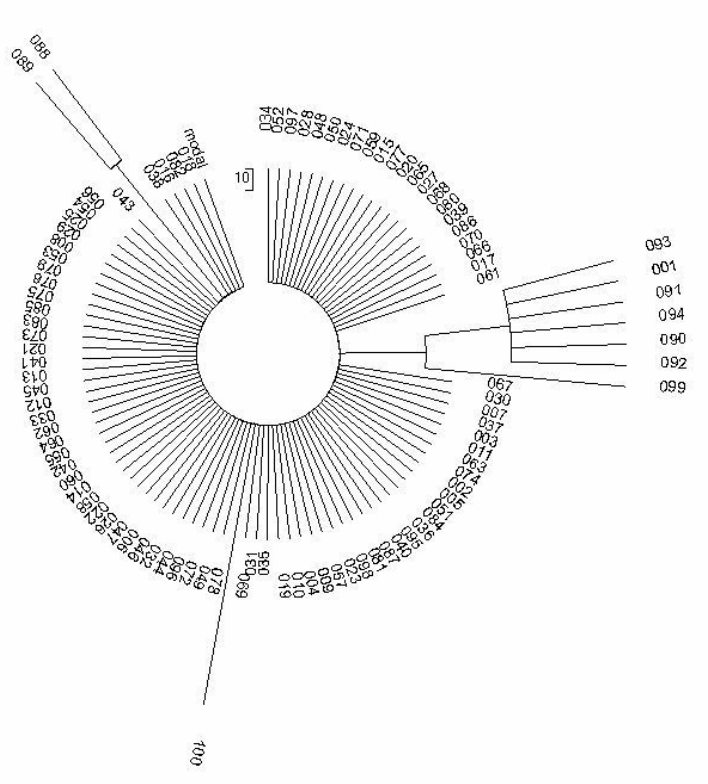

Figure 36. The 6-marker haplotype tree for 100 Jewish haplotypes of haplogroup R1a/R1a1. A “club” set (the Jewish R1a/R1a1 Project).

The “club” set (Fig. 36) looks like the “youngest” one among the three. It shows 90 base haplotypes from the total 100, which gives an estimate of ln(100/90)/0.0096 = 11 generations only. However, since the whole set contains

only 10 mutations in all 100 haplotypes, it gives 10/100/0.0096 = 10.4 generations, if again to be absurdly precise. Hence, despite the low number of generations, the figures again point at a single ancestor in the set. This is a rare case, since a majority of haplotype sets are derived from various ancestors, and typically show a significant mismatch between a number of generations found from residual base haplotypes and mutations in the set.

12-marker haplotypes ("commercial" and a "club" sets)

There are no 12-marker "scientific" sets of R1a/R1a1 haplotypes published in the literature. Therefore, we have to turn to other data bases.

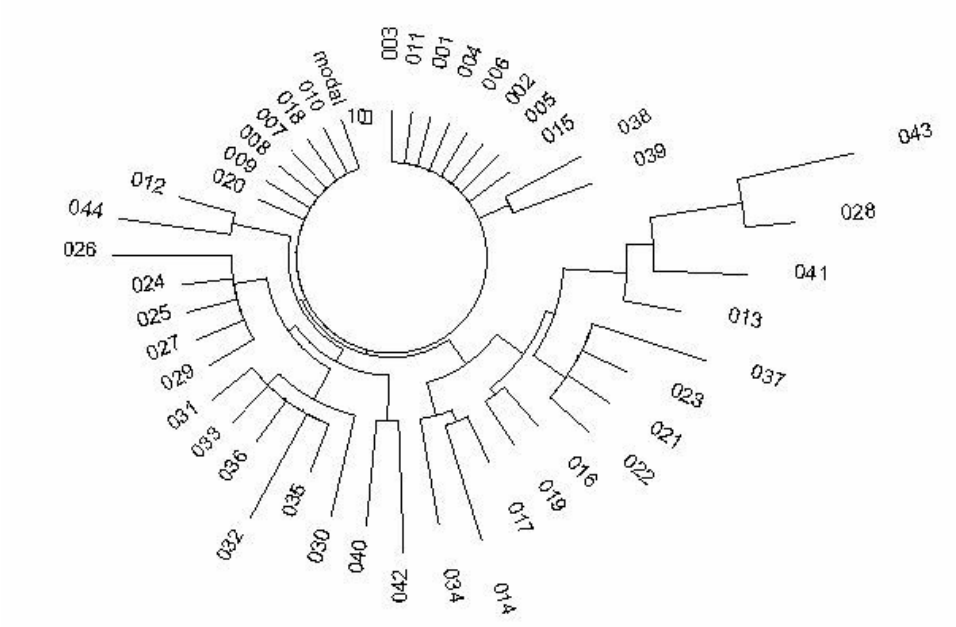

Figure 37. The 12-marker haplotype tree for 44 Jewish haplotypes of haplogroup R1a/R1a1. A "commercial" set (YSearch database)

Let us consider a 12-marker "commercial" haplotype set from YSearch database (Fig. 37). There are 14 base haplotypes in the 44-haplotype tree, seen as a "comb" at the tree top

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

This gives ln(44/14)/0.024 = 48 generations to a common ancestor. There are 57 mutations in all these 44 haplotypes, that gives 57/44/0.024 = 54 generations to a common ancestor. It is rather impressive, how both 6-marker "scientific" and 12-marker "commercial" set of haplotypes give the same results in terms of the shape of the haplotype trees and, hence, a number of generations to a common ancestor.

The above 12-marker Jewish base R1a1/R1a haplotype is identical to the most widespread 12-marker R1a1/R1a haplotype in Eastern Europe (mostly non-Jewish), shown above.

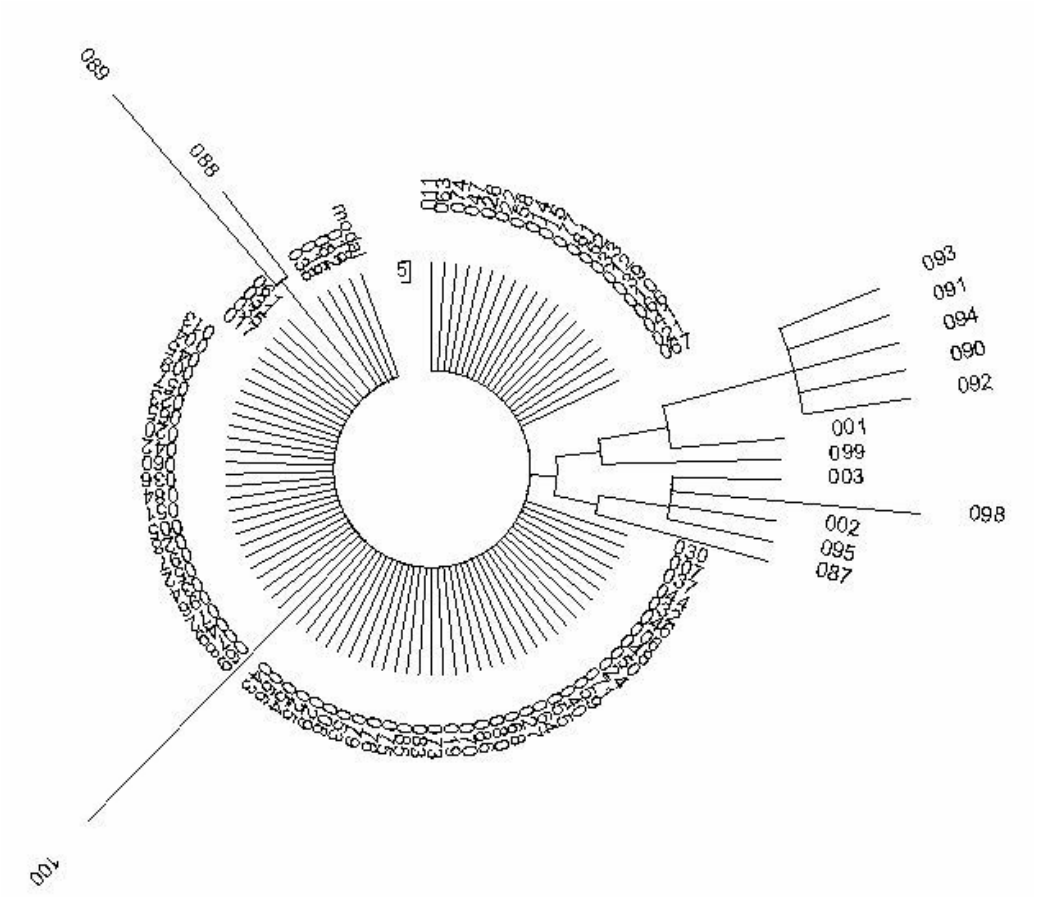

Figure 38. The 12-marker haplotype tree for 100 Jewish haplotypes of haplogroup R1a/R1a1. A "club" set (the Jewish R1a/R1a1 Project).

The “club set” series of R1a1/R1a haplotypes arranged as a haplotype tree is shown in Fig. 38. Again, it shows too few mutations in the whole tree (only 22 mutations in 100 12-marker haplotypes), and the whole tree contains 85 base haplotypes. Despite the clearly underestimated number of generations to a common ancestor, the set gives again the same number of the generations based on both mutations and the residual base haplotypes, hence, points again to a single ancestor of the whole tree.

25-marker haplotypes (“commercial” and a “club” sets)

The “commercial” set, from YSearch database, consists of 30 haplotypes (Fig. 39). There are 3 base haplotypes in the tree

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

Based on a fraction of the base haplotypes on the whole tree, there are 50 generations to a common ancestor of all 30 individuals in the group, and a number of mutations give 49 generations to the common ancestor. Again, the data point to a single common ancestor of R1a/R1a1 Jews living today, or at least those Jews who have entered their haplotypes to the database.

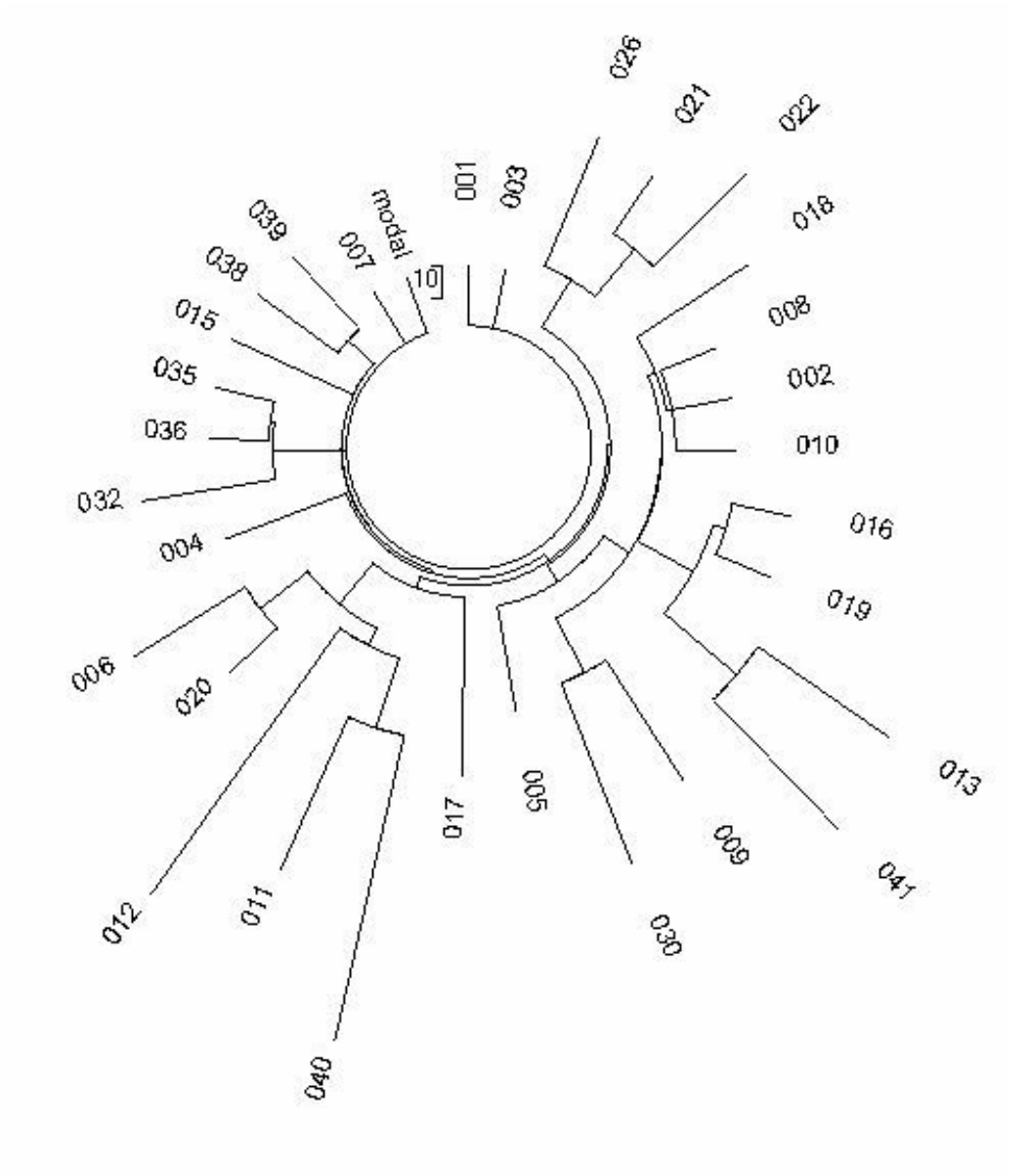

Figure 39. The 25-marker haplotype tree for 30 Jewish haplotypes of haplogroup R1a/R1a1. A "commercial" set (YSearch database)

However, it seems that the set of these 30 haplotypes hints at an offshoot of a group of ancestors who have DYS#459b = 11 (the 15th marker in the haplotype above). There are 10 haplotypes of the total 30, in which this particular allele is 11. It does not look as a statistical mutation, and rather reflects a split of the

haplotype with allele 11 and the following active proliferation of it. This thought is elaborated below.

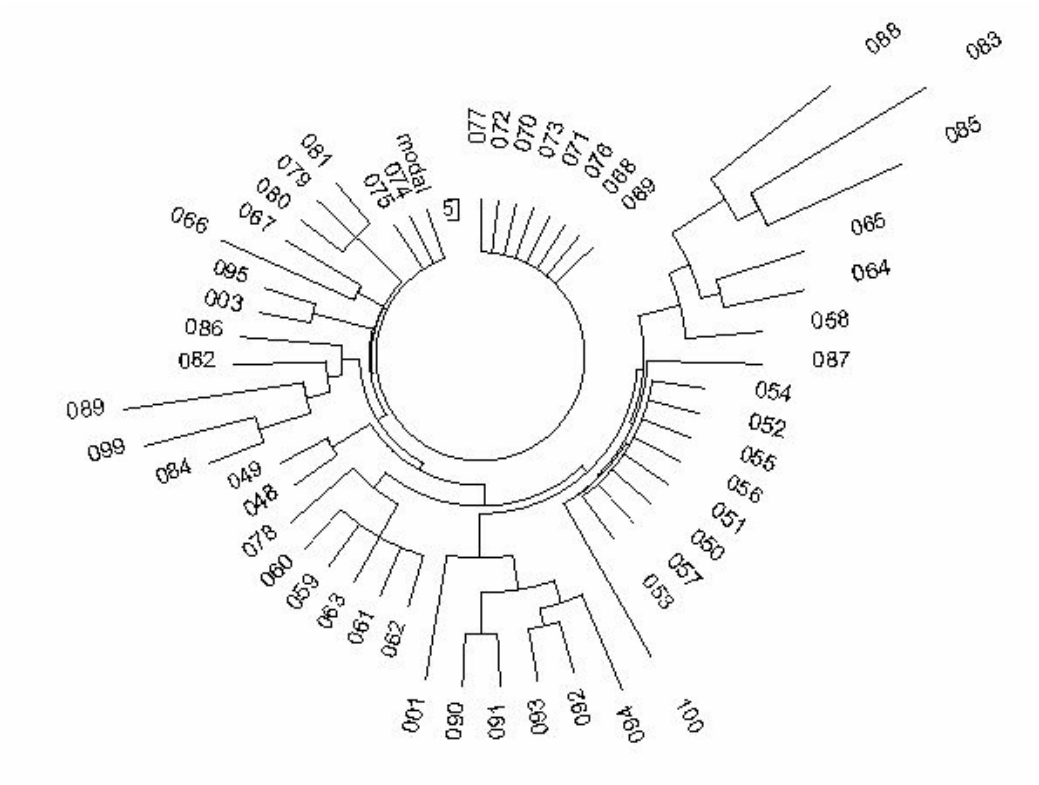

Figure 40. The 25-marker haplotype tree for 52 Jewish haplotypes of haplogroup R1a/R1a1. A "club" set (the Jewish R1a/R1a1 Project).

The "club" set consists of 52 haplotypes, with 10 base haplotypes, sitting on top of the tree (Fig. 40). They are identical to the above 25-marker base haplotypes of the "commercial" set. The "club" set still shows a reduced number of generations to a common ancestor compared to the "commercial" and "scientific" sets, calculated from the residual base haplotypes and from the number of mutations, and found to be 36 generations in both cases. In this 52-haplotype set half of all haplotypes have allele 10, and other half – 11 in DYS#459b. Obviously, it is not a random mutation, but a sign of a split in descendants of the above base haplotype, as it was suggested above.

37-marker haplotypes ("commercial" and a "club" sets)

The "commercial" 37-haplotype tree (Fig. 41) shows the split into two equal groups of haplotypes. This split is unrelated to a difference between R1a and R1a1 haplogroups, since both of them equally presented in the each half of the tree.

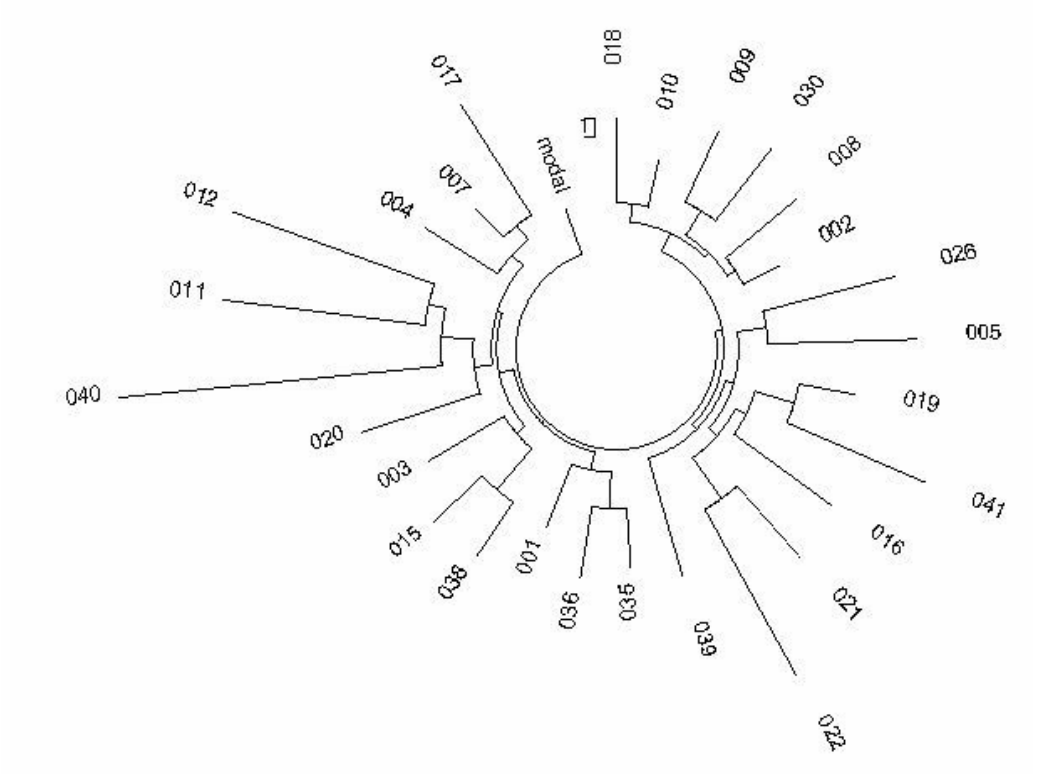

Figure 41. The 37-marker haplotype tree for 27 Jewish haplotypes of haplogroup R1a/R1a1. A "commercial" set (YSearch database)

The 37-marker base haplotype of the tree is shown in the Table. The first, 12-marker panel of all 27 haplotypes on the tree contains 20 mutations, the 25-marker panel contains 63 mutations, the total 37-marker panel contains 150 mutations with respect to the base haplotype. On average, it gives 50±17 generations to a common ancestor. If to calculate mutations separately, for each half of the tree, as shown in Fig. 41, haplotypes of the left-hand side branch belong to the lineage originated by a common ancestor who lived 54±28 generations BP, and those on the right-hand side belong to its derivative lineage

which was initiated 40±6 generations BP. The last wide branch consists of two sub-branches.

The main reason why these two lineages went by two opposite sides of the haplotype tree in Fig. 41 is that these two lineages are deviated in only two markers in 37-markers haplotypes, in DYS##459b and 570. Haplotypes in the parent lineage (on the left in Fig. 41) have there alleles all 10 and predominantly 20, and haplotypes in the derivative, "younger" lineage (on the right-hand side) have these alleles predominantly 11 and 21. The two sib-branches of the "younger" lineage split mainly by DYS## 459b and CDYb, which has alleles all 11 and all 38, respectively, in one sub-branch, in the upper right-hand side, and a mix in the other.

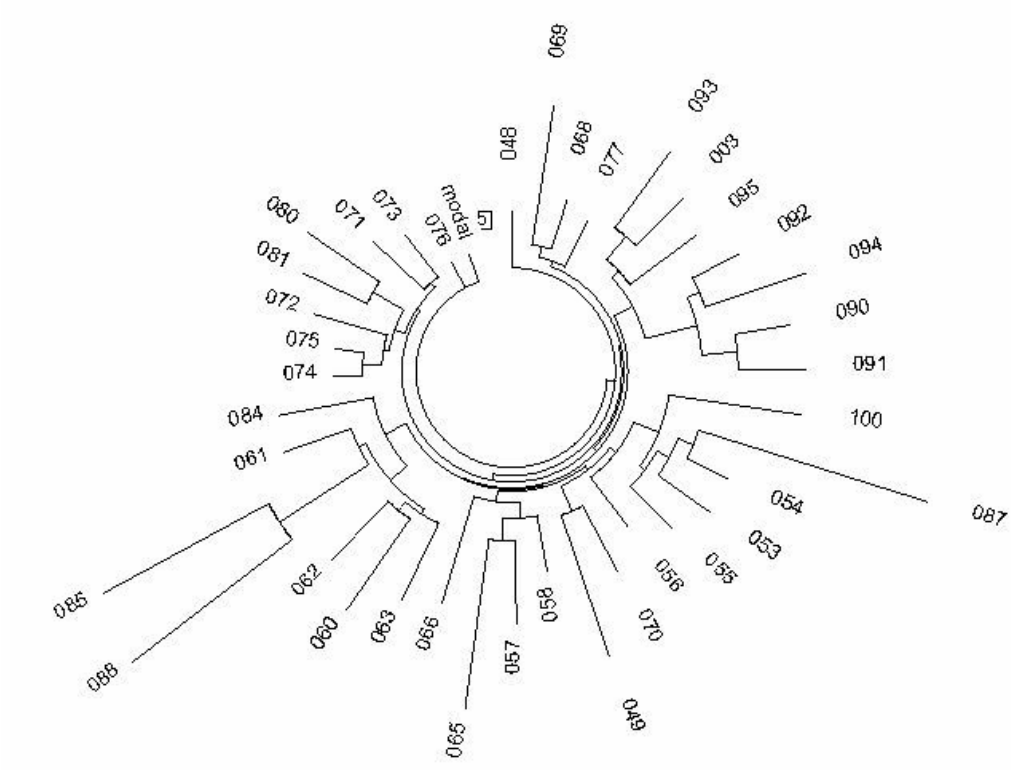

Figure 42. The 37-marker haplotype tree for 38 Jewish haplotypes of haplogroup R1a/R1a1. A "club" set (the Jewish R1a/R1a1 Project).

This split is not pronounced in the "club" set (Fig. 42), containing 38 haplotypes. Its first, 12-marker panel of all 38 haplotypes contains 16 mutations, the 25-marker panel contains 68 mutations, the total 37-marker panel contains 196

mutations with respect to the base haplotype. On average, it gives 40±22 generations to a common ancestor. It seems that the majority of haplotypes in the "club" series belong to descendants of the derivative branch of R1a1 lineage. Hence, a "smother" tree in Figs. 36 and 38 compared with the trees in Figs. 34, 35, and 37.

67-marker haplotypes ("commercial" and a "club" sets)

67-marker trees give the best resolution of branches available compared with lesser-marker haplotypes.

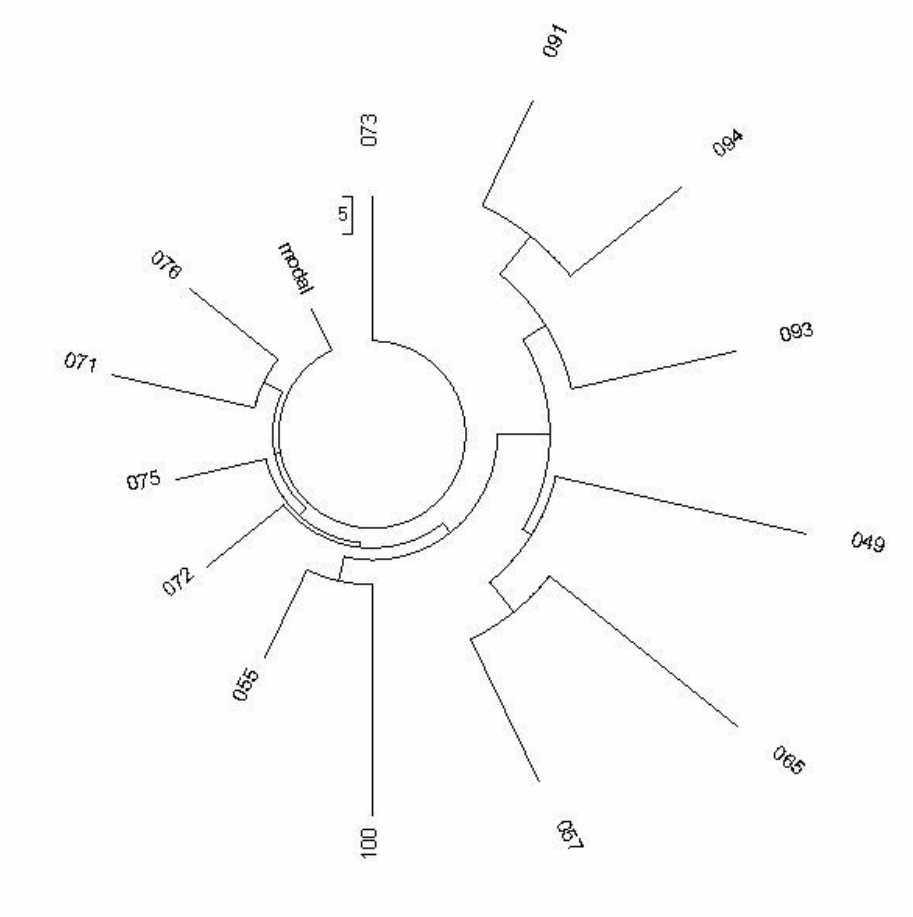

Figure 43. The 67-marker haplotype tree for 13 Jewish haplotypes of haplogroup R1a/R1a1. A "club" set (the Jewish R1a/R1a1 Project).

The 13-haplotype “club” tree (Fig. 43) is clearly separated into two halves which were briefly discussed above. Its 12-, 25-, 37- and 67-marker panels contain 8, 19, 58 and 83 mutations, which translates to 26, 33, 52 and 46 generations to a common ancestor, on average 39±12 generations. However, if to calculate the both halves of the tree, there will be 54±12 and 26±4 generations to their common ancestors.

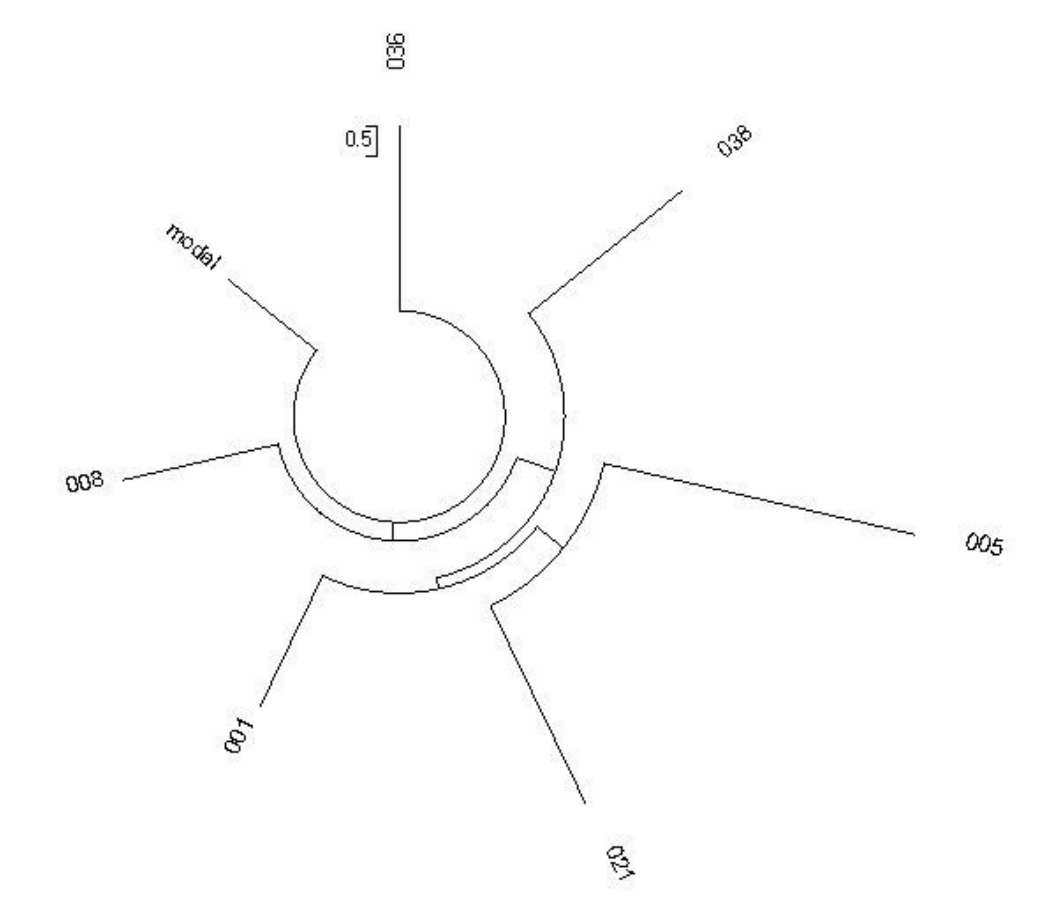

Figure 44. The 67-marker haplotype tree for 6 Jewish haplotypes of haplogroup R1a/R1a1. A “commercial” set (YSearch database)

The “commercial” 67-marker tree has contains only 6 haplotypes, and too small to be analyzed (Fig. 44). Overall, it does not conflict with the “club” tree (Fig. 43). The respective 67-marker base haplotypes are shown in the Table.

Time span to common ancestors in R1a/R1a1 Jewish community. Some historical reminiscences

All twelve figures calculated as a time span to a common ancestor, obtained from 6-, 12-, 25, 37- and 67-haplotype series taken from "scientific", "commercial" ("YSearch") and "club" (Jewish R1a/R1a1 Project) generally fit to the same range, an average number of which is 47±12 generations. The refined numbers pointed at two lineages, the oldest one 54±6 generations, that is about 1,350±300 years BP, around the 7th century AD. It can easily be assigned to Khazar Khaganate, though we do not have direct evidences of such as assignment. It is plausible, however. Around 650 years BP, in the 14th century, a part of the Jewish R1a/R1a1 community had apparently passed through a haplotype bottleneck, and started a new lineage, which now led to a split of R1a1 haplotypes into two approximately equal populations, that is detectable only at and above 37-marker haplotypes, preferably at the 67 marker haplotype level. It might be that this bottleneck reflected a fled from black plague in Central Europe in the middle of the 14th century. The Jewish R1a1 survivors have eventually moved to Eastern Europe, and by the 19th century most of their descendants lived in nowadays Belorus, Ukraine, Poland, Latvia, Lithuania, Russia.

The R1a1 Jewish base haplotype currently is, and it apparently was since around the 7th century, one of the most frequent among R1a1 haplotypes in Eastern Europe. Again, it might have a Khazar heritage, considering tight communications of Khazars with Slavic people of Kiev Rus those times. It is known that Jewish merchants had resided in Kiev, and that after Khazar Khaganate failed, some of the Jews made their way to the West. Why the R1a1 haplotype entered the Jewish community via Levites remains a mystery, and probably will remain. The secret has gone to a grave of a Levite wife, apparently. End of the romantic (hopefully) story.

In fact, R1a1 haplogroup can be met among native people of the Middle East, such as Bedouins, Kurds, Arabs, Jews (see the preceding paper in this issue). In a study (Nebel et al, 2001a) 33 haplotypes of R1a1 haplogroup have been identified in representative of all the four populations living in the Middle East. Only the Jews who currently live in Israel came back after some two millennia in the Diaspora; other three groups of native Middle Easterners presumably continue lineage of their predecessors who never left their homeland for a long time. Indeed, five of the ten Israelite Jews tested positively for R1a1 had the same base haplotype shown above

16-12-25-10-11-13

These ten haplotypes have only 8 mutations with respect to the above base haplotype. Five base haplotypes out of ten haplotypes suggest 72 generations to a common ancestor, while a number of mutations gives an estimate of 83 generations, hence, beginning of AD. Considering a statistically insufficient amount of haplotypes and exactly the same base haplotype as shown above, one can say that a deduced common ancestor lived not too far away in time from that (apparently) Khazar ancestor discussed above.

However, the rest 23 haplotypes of R1a1 haplogroup in Arabs, Kurds and Bedouins (including three Sephardim who live in Turkey, Iraq and North Africa) had only 6 base haplotypes

17-12-25-11-11-13

and as many as 41 mutations in the 23 haplotypes. Obviously, they descended from an ancient ancestor, who lived between about 150 and 180 generations ago, that is 3,800 to 4,500 years ago. If to exclude the Sephardim, the rest 20 individuals have the same base haplotype as indicated above (5 of them) and 36 mutations, that give 160 to 210 generations to a common ancestor, that is about the same figures as above, that is 4,000 to 5,300 years ago. This was time of an expansion of the Aryans to the Middle East, Persia, Afghanistan, India, though many details of that expansion remains to be obscured. This will be a subject of our forthcoming study.

As a conclusion of this section, here are the base, ancestral haplotypes of some populations in haplogroup R1a1 (see also the preceding paper in this issue):

Ashkenazi Jews in Europe, 54±6 generations, that is 1,350±300 years BP:

16-12-25-10-11-13

Russians (Eastern Slavs), 185±8 generations, that is 4,600±200 years BP:

16-12-25-11-11-13

India, 167±8 generations, that is 4,200±200 years BP:

16-12-25-11-11-13

Armenians, 164±6 generation, that is 4,100±150 years BP:

16-12-25-11-11-13

Middle East, Arabs, 121±6 generations, that is about 3,000±150 years BP:

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

Western European - the Donald family, 26 generations BP; Swedes, 145 generations, that is about 3,600 years BP:

15-12-25-11-11-13

Obviously, the Jewish R1a1 population is the most recent one compared to other R1a1/R1a populations, except the Donald family, of course, the 6-marker haplotype was given here just for a comparison.

The Jews of Haplogroup Q

Haplogroup Q (SNP M242) is rather rare in Europe. It is believed to be originated in Europe some 20-25 thousand years ago, transferred to Alaska in about those times, and then spread over to the Americas becoming one of a few major haplogroups among American Indians. Haplogroup Q has some occurrence in Central Asia and in the Euro-Asian steppes among Turkic-speaking people, with frequences commonly between 4% and 10%. Some researchers suggest that haplogroup Q has gotten to the Jews in the Khazar Khaganate, after some Khazarians were converted to Jewdaism around 740 AD. Some current estimates show that approximately 5% of Jews carry haplogroup Q.

YSearch database lists 253 individuals having haplogroup Q. 23 of them, that is 9%, have reported their ancestors with common Jewish names such as Yitzchok, Schloime, Avraham, Reuven, Menashe, Nathan, Yankel, Aaron, Yehoshua, Zelk, Leib, Israel, Solomon, and others. Those 23 haplotypes formed the “commercial set” in this study. Another source of Jewish haplotypes was a publication by Behar et al (2004) which contained also 23 haplotypes of haplogroup Q (SNP P-36) (the “scientific set”), but different ones. Yet another source of Jewish haplotypes was the “Jewish Q Project”, which contained 6- and 12-marker 58 haplotypes, along with some more extended haplotypes, described below (the “club set”).

Analysis of these haplotypes showed that a common ancestor of present day Jews of haplogroup Q lived 675±175 years BP, around the 14th century. This is when the Jewish lineage in haplogroup Q was originated. Some sub-branches of this group of haplotypes might refer to subfamilies of the same Jewish Q family formed approximately 275 years BP, in the 18th century.

The most widespread 6-marker haplotype of the overall Q haplogroup is

13-12-23-10-14(15)-13

in which the penultimate marker has practically the same frequency of alleles 14 and 15. This refers to all bearers of Q haplotype, non-Jewish and Jewish. There are 339 of them, including subgroups, in YSearch database. The most frequent 12-marker haplotype among those who have tested and happened to belong to the overall (Jewish and non-Jewish) haplogroup Q, is

13-22-13-10-14-16-12-12-12-13-15-29

There are 331 of the in Q (with subgroups) section in YSearch database.

Let us consider the Jewish haplotypes.

6-marker haplotypes ("scientific", "commercial", and a "club" sets)

All the three sets ("scientific", "commercial" and "club") show that the Jewish Q haplotypes are derived from a recent ancestor. In the "scientific" set (Fig. 45) 19 of all 23 haplotypes are identical to each other, with the base haplotype

13-12-22-10-15-13

There are only four mutations in all those 23 haplotypes, that is among 138 markers. These figures formally refer to a single common ancestor for all the Jews in the group, who – according to these truncated, 6-marker haplotypes – lived only about 19 generations ago, since ln(23/19)/0.0096 = 20, and 4/23/0.0096 = 18.

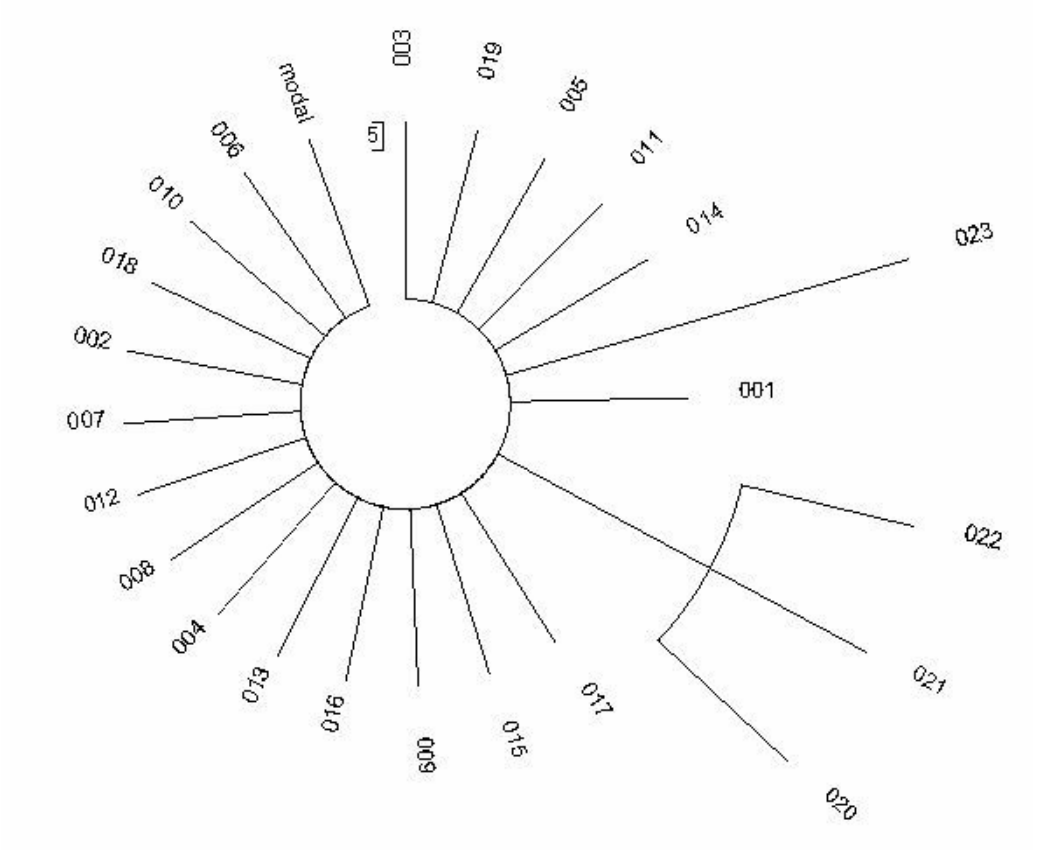

Figure 45. The 6-marker haplotype tree for 23 Jewish haplotypes of haplogroup Q. A "scientific" set (Behar et al, 2004)

To some surprise, it turned out that all the 23 individuals in the "commercial" set, that is in YSearch database, have the same, identical 6-marker haplotype, exactly as that in the "scientific" set and shown immediately above.

In the "club" set 54 out of all 58 haplotypes are identical to each other and to the base haplotype from both the "scientific" and "commercial" sets, shown above. Each of the four "deviated" haplotypes has a single mutation (in fact, each in the same marker, DYS#393). Again, the data point at a single recent ancestor for all 58 individuals in the Jewish group of haplogroup Q, which in this case lived very close to our times.

Obviously, in these cases short, 6-marler haplotypes are practically useless for a detailed analysis. Either the common ancestor lived too recently, and a rough resolution capacity for 6-marker haplotypes is useless in this case, or on some

peculiar statistical reason most of mutations happened in more remote panels of 12-, 25-, 37 or 66-marker haplotypes, or both. It will be shown later in this section.

10-marker haplotypes (the "scientific" set)

The 10-marker 23-haplotype tree is identical to that shown in Fig. 45 for the 6-marker tree. This is because all four mutations in the 10-marker set occurred in the 6-marker area. 19 haplotypes of those 23 were the same:

13-22-13-10-X-Y-12-12-12-13-15-29

It will be shown below that the missing alleles X-Y in the Behar's work (2004) are equal to 14-16.

The data still point at a single common ancestor. This follows from practically the same number of generations calculated from a number of residual base haplotypes, ln(23/19)/0.025 = 8, and from a number of mutations in the whole set, 4/23/0.025 = 7. The figure of 0.025 here is the mutation rate as a number of mutations in a 10-marker haplotype per generation (25 years). This figure was described in the preceding paper in this issue.

12-marker haplotypes ("commercial" and the "club" sets)

The YSearch set of 23 haplotypes is shown as a haplotype tree in Fig. 46, and the "club set" of 58 haplotypes in Fig. 47.

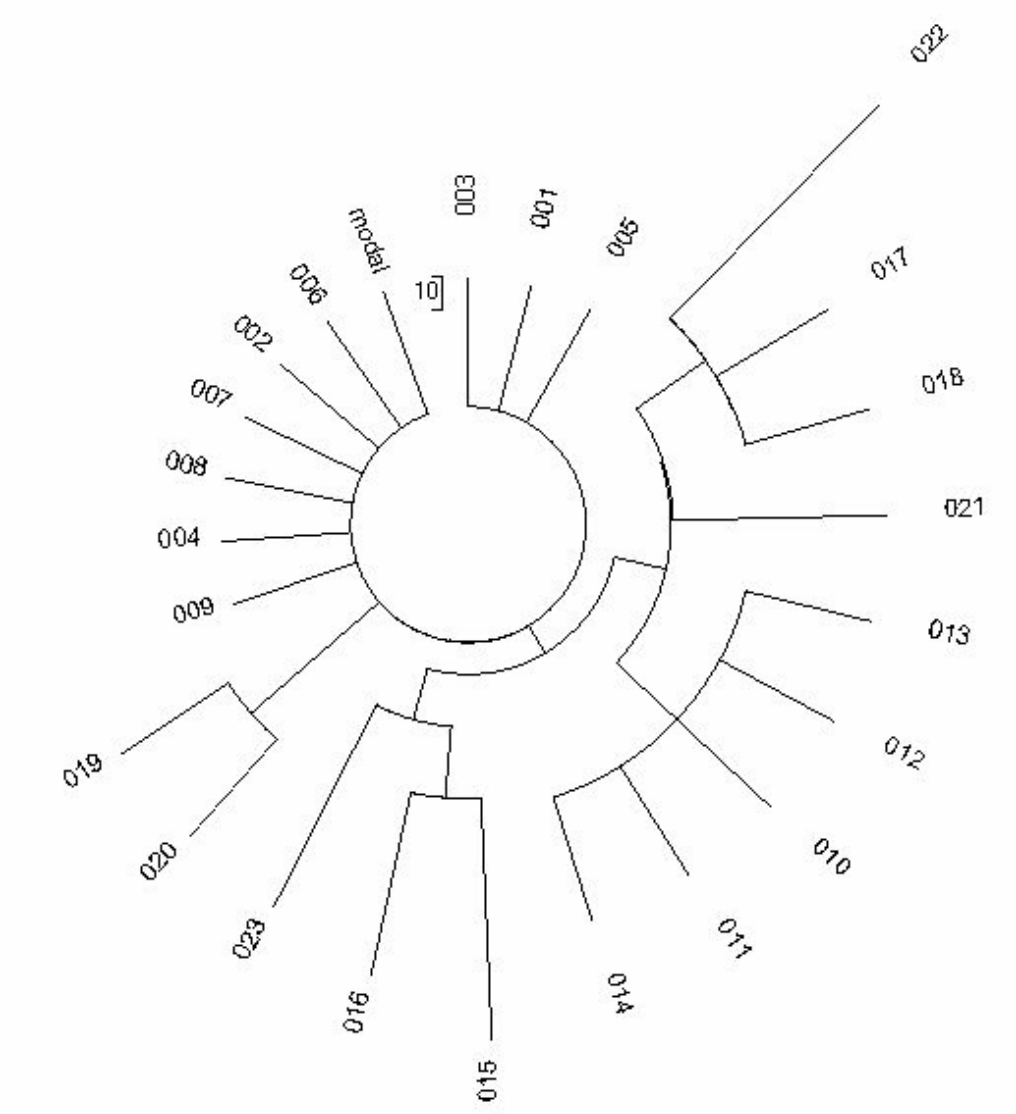

Figure 46. The 12-marker haplotype tree for 23 Jewish haplotypes of haplogroup Q. A "commercial" set (YSearch database)

Both show practically a perfect fit between a number of residual base haplotypes

13-22-13-10-14-16-12-12-12-13-15-29

and a number of mutations in complete sets, translated into a number of generations to a common ancestor for individuals in the both sets. The

"commercial set" contains 9 base haplotypes and 21 mutations, the "club set" contains 30 the same base haplotypes and 40 mutations. This gives ln(23/9)/0.024 = 39 generations and 21/23/0.024 = 38 generations to a common ancestor for the "commercial set", and ln(58/30)/0.024 = 27 generations and 40/58/0.024 = 29 generations to a common ancestor in the "club set". Both sets of data point at a single common ancestor for each set, and both indicate that the ancestor lived rather recently (in terms of DNA genealogy), less than a thousand years BP, in this millennium.

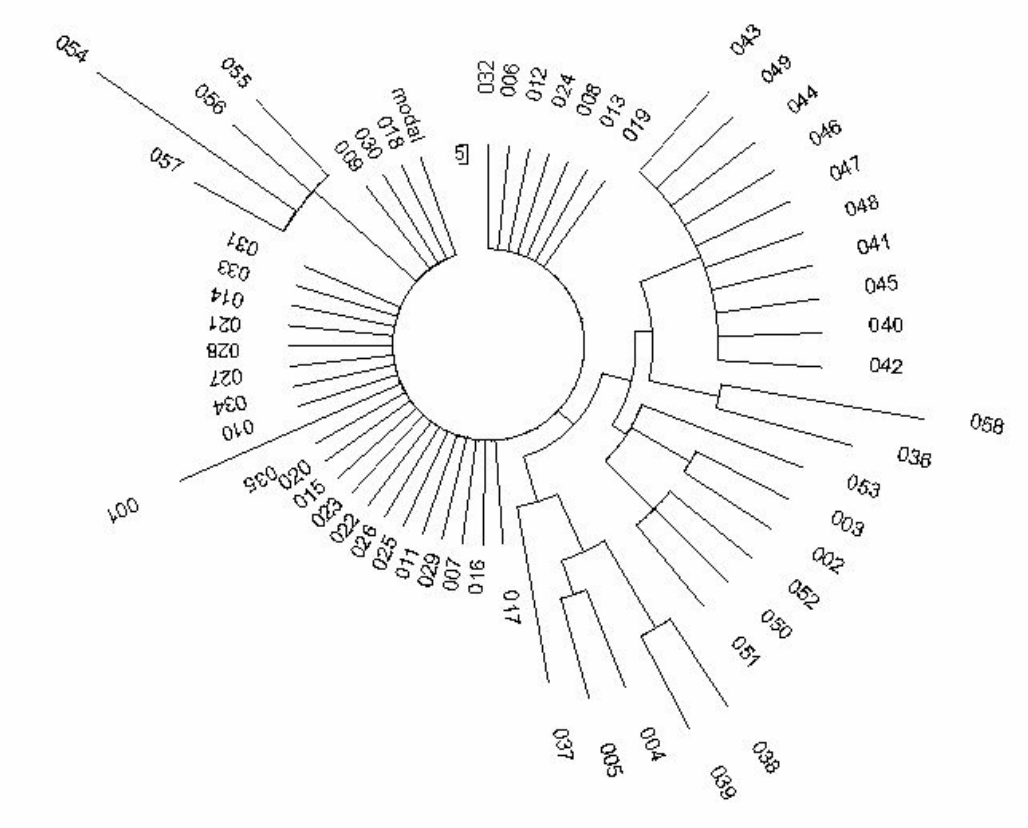

Figure 47. The 12-marker haplotype tree for 58 Jewish haplotypes of haplogroup Q. A "club" set (the Jewish Q Project).

The above Jewish base haplotype is identical with the most common haplotype among all bearers of Q haplogroup recorded in YSearch database (see above). Hence, there is nothing distinctly "Jewish" in the 12-marker base Jewish haplotype.

The following consideration shows that the "commercial set" gives some overestimated number of generations due to higher than an average number of mutations in the marker DYS#385b. This marker provided the most mutations in 12-marker haplotypes, more than half of them in each set (and more than two thirds in the "commercial" set, hence, the longer "mutational" time to a common ancestor). Apparently, as well as in R1a1 Jewish haplotypes, there was a bottleneck for Jewish Q haplotypes several hundred years ago, when a bearer of allele 16 of DYS#385b split/survived and began a new, mutated lineage of Jewish Q haplotypes. We will see it better with more extended haplotypes.

37-marker haplotypes ("commercial" and the "club" sets)

Fig. 48 shows a 37-marker, 29-haplotype tree for the "club set". It can be seen that it contains three sub-branches, one of them is slightly "older" (the upper-right-hand side), than two others. It turned out that all three have the same base haplotype in 37-marker format (Table).

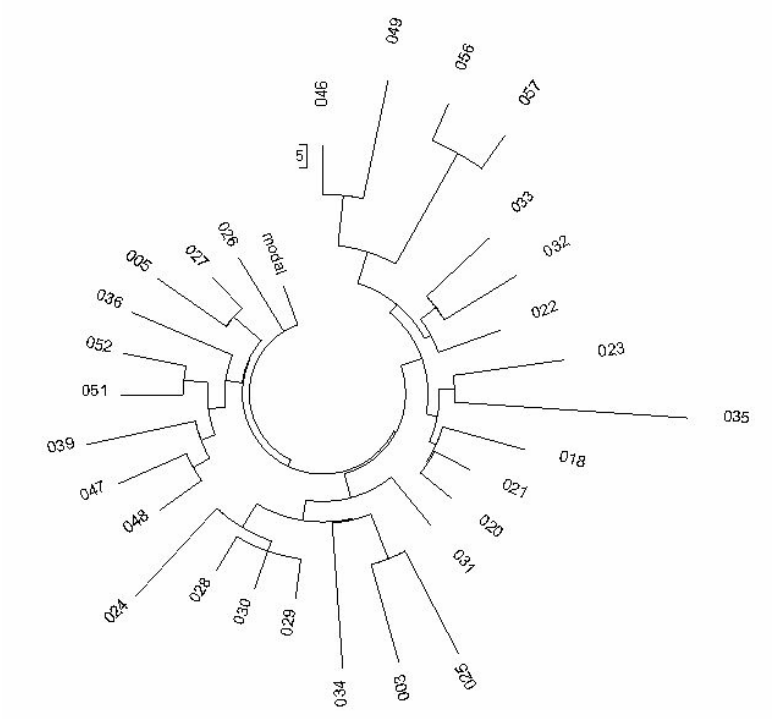

Figure 48. The 37-marker haplotype tree for 29 Jewish haplotypes of haplogroup Q. A "club" set (the Jewish Q Project).

Unfortunately, only five individuals of 23 in the "commercial set" have determined their 37-marker haplotypes. Nevertheless, these five haplotypes provide valuable information, and confirm that a common ancestor lived rather recently. The first 60 markers of the five 37-marker haplotypes (that is, five 12-marker haplotypes) contain only 3 mutations. This corresponds to 25 generations to a common ancestor. When the next panel of 13 markers x 5 haplotypes (65 markers total) is added to those five haplotypes, it adds only 3 mutations again, and corresponds to 26 generations to a common ancestor. The last panel (26 through 37 markers), making five 37-marker haplotypes, adds only 4 mutations, and corresponds to 22 generations to a common ancestor.

Similar calculations with the 29-haplotype 37-marker "club" tree gave 22, 26 and 34 generations to a common ancestor of all 29 individuals. Furthermore, the 1-37- and 1-66 marker panels of fourteen of the 66-marker haplotypes gave 33 and 68 mutations, that corresponds to 27 or 35 generations to the common ancestor. An average of said 25, 26, 22, 22, 26, 34, 27 or 35 generations gives 27±5 generations to a common ancestor, that is 675±125 years BP, around the 14th century. This is when the Jewish lineage in haplogroup Q was originated. The three sub-branches mentioned above might refer to subfamilies of the same Jewish Q family formed approximately 11 generations, or 275 years BP, in the 18th century, based on a number of their mutations.

66-marker haplotypes (the "club" set)

The 14-haplotype 66-marker haplotype tree is shown in Fig. 49.

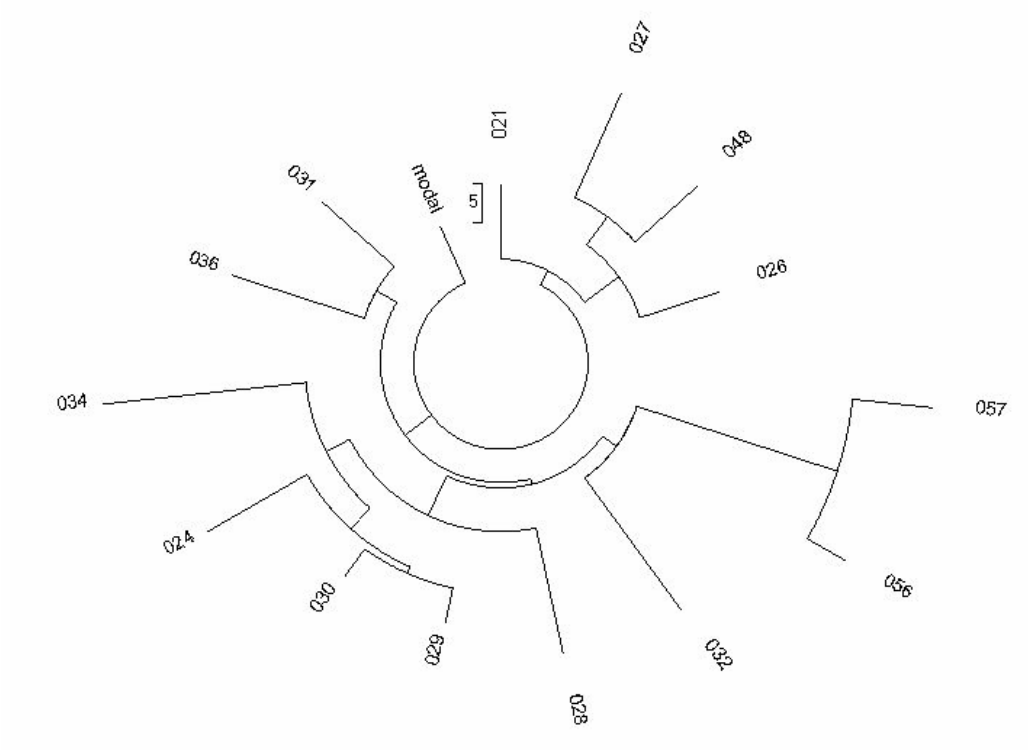

Figure 49. The 66-marker haplotype tree for 14 Jewish haplotypes of haplogroup Q. A "club" set (the Jewish Q Project).

It confirmed an existence of a split of a Jewish sub-family of haplogroup Q some 11 generations ago, apparently in the 18th century. A common ancestor of all 14 individuals, 66-markers of which composed the tree, lived 27±5 generations BP, or around the 14th century. It confirms calculations in the preceding section. This is a bit too late for the Khazar Khaganate, which had dissipated in the 11-12th century AD. However, haplogroup Q could be acquired by the Jews after it. This could be resolved one way or another after considering more haplotypes of the Jews of haplogroup Q.

The Jews of Haplogroup R2

Haplogroup R2 is rather rare outside India, where it accounts for about 90% of all men on Earth, having R2. In India it has been observed in about 10% of male population, in Pakistan – about 7-8%. In Tadzhikistan, neighboring India, haplogroup R2 is met in about 6% of the population. Some singular percentage of population having R2 can be met in the area of Caucasus, among Azerbaidzhanians, Armenians, Georgians, Chechens. It is conjectured that in these areas haplogroup R2 was introduced by the Gypsies, who carry haplogroup R2 with frequency of more than 50% of their population. The next main haplogroup in the Gypsies is H, as it was described in the preceding paper in this issue with an example of Bulgarian Gypsies. It is surmised that haplogroup R2 was originated some 25 thousand years ago.

The Gypsies have brought haplogroup R2 to Europe in medieval times, some 500-700 years ago, apparently first to Bulgaria, Germany and Austria (under the Gypsies names of Sinti and Roma), and then spread over Europe. This haplogroup was recently found among the Jews, and immediately it was suggested by the scholars that it came from the Khazars. No justifications and no time estimates were given.

Recently (Sengupta et al, 2006) a large set of Indian and Pakistani haplotypes was published, including more than 900 haplotypes. 81 of them belonged to haplogroup R2. Since as many as 21 of identical six-marker haplotypes (the base haplotypes) from those 81 are observed, as follows

14-12-23-10-10-14

it is rather obvious that these haplotypes cannot be too old. Indeed, ln (81/21)/0.0096 = 141 generations (163 with correction for back mutations) to a common ancestor. All 81 haplotypes contain 108 mutations from the above base haplotype, which gives 108/81/0.0096 = 139 generations (161 with correction for back mutations). It is a practically absolute fit, indicating that it was a single ancestor who originated the lineage of R2 haplogroup in India 4,000 years BP.

However, it seems that the actual time of origination of R2 haplogroup was much earlier. The R2 section of YSearch data base contains 34 haplotypes of individuals. Half of them are ethnic Indians, plus some Scotts, French, Italians, Armenians. Twelve individuals have names of their predecessors as Abraham, Isaac, Lebe, Mordecai, etc., and some of them presented supplementary information indicating that they are Ashkenazi Jews. The most frequent 6-marker haplotype among those 39 individuals is

14-12-23-10-10-14

which is exactly the same as that the base haplotype of haplogroup R2 in India-Pakistan, shown above. However, if to remove the Jewish haplotypes (which, as it is shown below, are derived from a recent ancestor), the remaining 22 haplotypes contain 35 mutations, that translates into 198 generations from a common ancestor. In 22 of the 12-marker haplotypes there were 101 mutations, which give 236 generations from a common ancestor. In 7 of the 37-marker haplotypes amount of mutations in the 12-, 25- and 37-marker panels corresponded to 282, 259 and 207 generations to a common ancestor. These four figures being averaged give 246±32 generations, that is about 6,200±800 years to a common ancestor of the non-Jewish individuals of R2 haplogroup in YSearch database. This might be a good indication that haplogroup R2 had originated not in India, since the Indian R2 haplogroups were derived from a significantly "younger" ancestor who lived about 4,000 years BP (see above), that is some 2,000 years later that an older bearer of the R2 haplogroup. In any case, this question needs more detailed studies.

And when a common ancestor of the Jews of haplogroup R2 had lived?

The most frequent 12-marker haplotype among those 34 individuals, Jewish and not, is

14-23-14-10-13-20-12-12-11-14-10-29

which is exactly the same as that for the Jewish individuals of haplogroup R2 in YSearch database.

Let us now consider the Jewish haplotypes in more detail.

6-marker haplotypes

11 of the 12 Jewish R2 haplotypes are identical to each other (Fig. 50), and their 6-marker base (ancestral) haplotype is

14-12-23-10-10-14

that is the same as the most popular among known bearers of R2 haplogroup in India and elsewhere in the world.

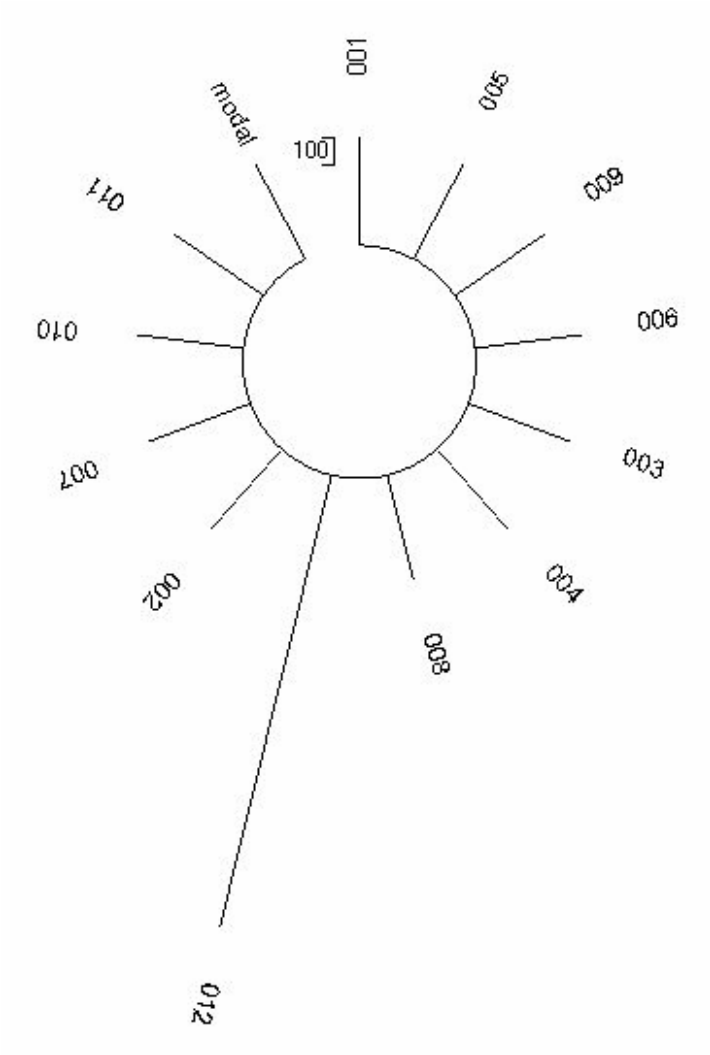

Figure 50. The 6-marker haplotype tree for 12 Jewish haplotypes of haplogroup R2. A "commercial" set (YSearch database)

Formally, 11 base haplotypes out of 12 give ln(12/11)/0.0096 = 9 generations, and one mutation in all twelve 6-marker haplotypes gives 1/12/0.0096 = 9 generations to a common ancestor, and the identity of these figures point out at a single ancestor for all 12 individuals in the set of their 6-marker haplotypes. However, as it has happened before, this tentative conclusion should be examined with more extended haplotypes. It is too often when 6-marker

haplotypes, particularly in small haplotype sets, do not reveal mutations which occur in more extended panels of the haplotype.

12-marker haplotypes

Indeed, a move to the 12-marker tree (Fig. 51) immediately shows that there are two groups of the Jewish haplotypes, with an "older" and a "younger" haplotypes, descending from the same ancestor. Half of all 12 haplotypes still represent the base (ancestral) haplotype

14-23-14-10-13-20-12-12-11-14-10-29

Since their other mutations will be revealed by moving to more extended haplotypes, an estimate of a time span to the common ancestor based on the 12-marker haplotypes will be only tentative.

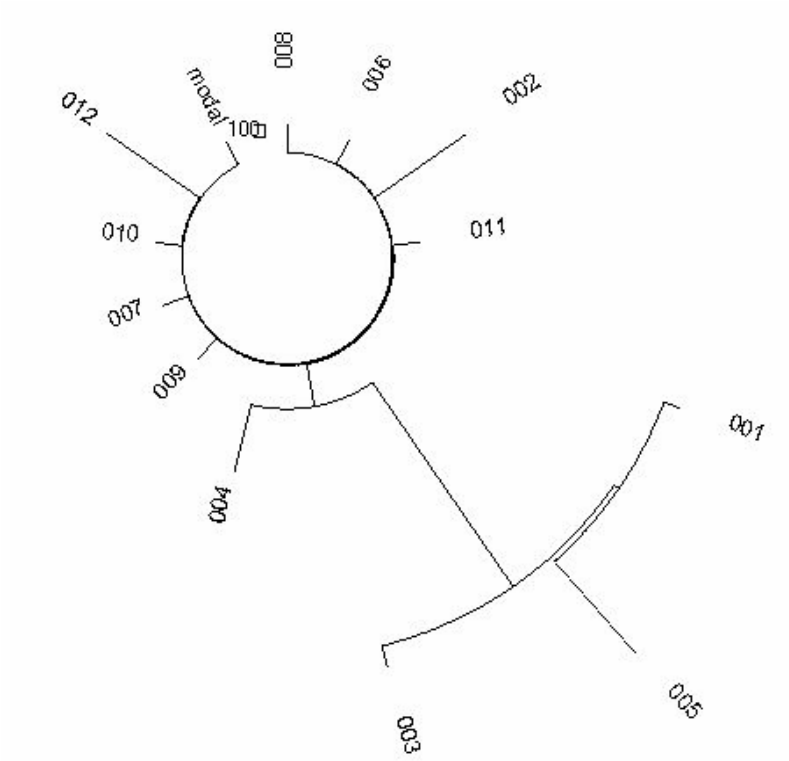

Figure 51. The 12-marker haplotype tree for 12 Jewish haplotypes of haplogroup R2. A "commercial" set (YSearch database)

This 12-marker base haplotype is exactly the same as the most frequent 12-marker haplotype in YSearch database, only one-third of which represent the Jewish haplotypes.

6 base haplotypes from the total 12 Jewish haplotypes would point to 29 generations to a common ancestor, since ln(12/6)/0.024 = 29. The other 6 haplotypes contain 15 mutations with respect to the above base haplotype. This would lead to 15/12/0.024 = 52 generations to a common ancestor. This mismatch (29 and 52) indicates that there were more than just one common ancestor for the Jews in R2 haplogroup. In fact, Fig. 51 makes it rather obvious.

As it turned out, many more mutations have occurred in the 13-37 marker panel of the distant branch, which distinctly separates the two branches. However, even the 12-marker tree shows the principal separation of the two groups of haplotypes.

37-marker haplotypes

The 37-marker haplotype tree is shown in Fig. 52. It reveals a striking feature of R2 Jewish haplotypes. Though there are only 7 haplotypes on the tree, they clearly show that Jewish R2 haplotypes splits indeed into two quite distant groups. Statistics is insufficient to perform detailed analysis, however, there are still enough data available to make some principal conclusions. One group of haplotypes, which are located on the right-hand side of the tree (Fig. 52) and much closer to the trunk of the tree (that is, to the present times), has the same 12-marker base haplotype as shown immediately above and in the Table, and corresponds to the group of the base haplotypes around the 12-marker tree (Fig. 51). This 4-haplotype branch has only 9 mutations with respect to the base 37-marker haplotype, and refers to a common ancestor who lived only 26 generations BP, 650±50 years ago, in the 14th century.

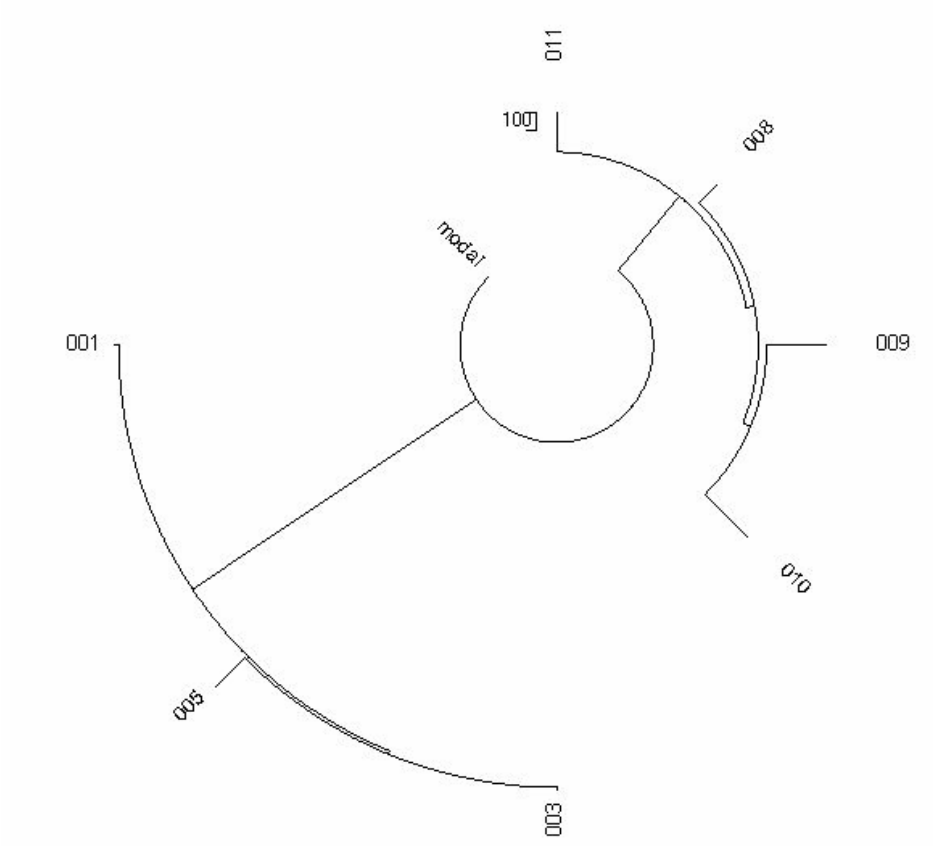

Figure 52. The 37-marker haplotype tree for 7 Jewish haplotypes of haplogroup R2. A "commercial" set (YSearch database)

Another group of Jewish R2 haplotypes, represented with three distant haplotypes on the left-hand side in Fig. 52, shows a base 12-marker haplotype (the 37-marker haplotype is shown in the Table):

14-23-14-10-13-20-12-12-10-13-10-31

It turned out that these three haplotypes (the left-hand side in Fig. 52) have only two mutations in their 37-marker haplotypes, that is among 111 alleles. This formally places their common ancestor only 7 generations BP, that is about two hundred years ago. All three are relatives within seven generations.

Overall, there are 21 mutations between these two base (ancestral) haplotypes in the 37-marker format. This means that these two haplotypes are separated by

thousands of years of separate mutations, and, more specifically, this separation is approximately equivalent to 305 generations between the two, that is about 7,600 years. This places their common ancestors about 4,200 years BP, and fits pretty well with the common ancestor of Indian R2 haplogroup of 4,000 years BP, see above. It is very likely that the both lineages, "young" and the ancient one, are derived from the Gypsies in Europe. The "young" is traced down to the invasion time or a bottleneck time to the Jewish community, and the "older" is traced down to the ancient common ancestor in India.

At any rate, both Jewish ancestral haplotypes, shown above in their 12-marker format and in the Table 2 in a 37-marker format are derived from two quite unrelated individuals, whose haplotypes had evolved from the very initial survivors in haplogroup R2, but traced down in millennia apparently to India, through the Gypsies.

Some historical conjectures

Here is a plausible story of the Jewish haplotypes of R2 haplogroup. Its ancestral haplotype

14-23-14-10-13-20-12-12-11-14-10-29

shown here in the 12-marker format, is about 4,200 years old, that corresponds to the age of this haplotype in India (sere above). This haplotype had arrived to Europe apparently with the Gypsies, in the Medieval times, some 800 years BP, and got into the Jewish community. About 30-40% of the present day Jews, bearers of R2 haplogroup, are direct descendant of those Gypsies, or the Indians, on that matter. Approximately 650 years ago, apparently during the Black Plague times, in the 14th century, a bearer of this haplogroup, albeit in the mutated form had survived and fled to the Eastern Europe. This was a bottleneck for this particular haplotype. Close to half of present day Jews are descendants of that individual.

This story is a mirror one of the Jewish Q haplotypes story (see the preceding section). Apparently, the 14th century, the Black Plague times, created a number of bottleneck situations for the Jews of a number of haplotypes, and not for the Jews only.

The second Jewish R2 haplotype

14-23-14-10-13-20-12-12-10-13-10-31

in the 12-marker format, got to the Jewish community quite recently, merely two hundred years ago. It is very different from the first one. Its three bearers lived in the 19th century in Hungary, Romania and Lithuania. Their current descendants probably do not know that they are rather close relatives. Two of them differ by only 3 mutations in their 66-marker haplotypes.

The Jews of Haplogroup G, and its subgroups G1, G2 and G2c (formerly G5)

Haplogroup G (SNP M201) has appeared – on various estimates – between 10 and 20 thousand years ago, probably in the northern or the eastern part of the Middle East, though estimates regarding both the age and a place of origin of these haplogroups and sub-groups are conflicting and unreliable. It is not very common in Europe and elswhere, though some populations of the Caucasus (Georgians, Kabardinians, Ingushians) have an elevated frequency of haplogroup G*, reportedly up to 30% of their males (Nasidze et al, 2003).

From about 40,000 haplotype entries in YSearch database 536 belong to haplogroup G, 676 to haplogroup G2, 19 and 10 to haplogroups G2c and G1, respectively, that is about 3% overall. A content of these haplogroups among Ashkenazi Jews is about 11%. When haplogroup G and its subgroups got into the Jewish community is currently unknown. This is a subject of this study.

Before we go into detail, this is a brief overview of main results of this section:

- The Jewish haplogroup G apparently does not exist without its sub-groups, such as G1, G2, G2a3 or G2c.Haplogroup G* was originated (or passed a bottleneck) about 15 thousand years BP, however, it has entered the Jewish community between 4,000±200 years BP (haplogroup G2), 3,600±200 years BP (haplogroup G1), reentered as the haplogroup G1 975±175 years BP, and as recently as 575±50 years BP (haplogroup G/G2c) and 625±200 years BP (haplogroup G2a3). Each of these entries had been unrelated to each other, and initiated separate lineages in the Jewish community. Their common ancestor is traced back, as it is indicated above, to 15,200±200 years BP, apparently the date of the origin of G (M201) haplogroup or of passing its bottleneck.

- Haplogroup G1 can be met among about 10% of the Jews of haplogroup G*. G2* accounts for about a quarter of the G* Jews. The most frequent among the G* Jewish people is a sub-group G2c, the youngest one, which account for two-thirds of the Jews in haplogroup G*.

- In this section Jewish haplotypes of G (M201), G1 (M285, M342), G2 (P287), G2a3 (U8), and G2c (M377, formerly G5) are considered. It turned out that haplotypes of these haplogroups and sub-groups are significantly different in patterns of their mutations, hence, in history of these haplogroups in the Jewish community. Jewish haplotypes of haplogroup G are always mixed with those of G1, G2, G2a3 and G2c on the respective branches, hence, it seems that typing of haplotypes G in each of those cases was not complete. The closest to this allegedly "undertyped" G haplogroup is G2c. In fact, these two cannot be separated, and have the same common ancestor at the same time period.

- Not all Jewish G1, G2, G2a3 and G2c haplotypes are distinct from other, non-Jewish haplotypes of these haplogroups. The most widespread G* 6-marker haplotype, deduced from all G* haplotypes (including those of subgroups) in YSearch database is

  15-13-22-10-11-14

  It differs by six mutations from that of haplogroup G1 in the Jews

  14-12-23-11-12-13

  It is identical to that of the "ancient" haplogroup G2 in the Jews (a base haplotype of an ancient lineage, 4,000±200 years BP)

  15-13-22-10-11-14

  It differs by two mutations from that of haplogroup G2a3 in the Jews (a rather recent lineage)

  15-12-22-10-11-13

  and it differs by three mutations from that of haplogroup G2c (the same as "G", a rather recent lineage)

  15-12-23-10-11-13

- It can be concluded that the ancient, ancestral haplotype of haplogroup G2 in the Jewish population had appeared 4,000±200 years BP from a pool of haplotypes of a common, non-Jewish population. Other Jewish base haplotypes either mutated since then and passed their respective "bottlenecks" several hundred years BP, or entered the Jewish population in the Diaspora from bearers of the respective haplogroups in general, non-Jewish population.

- Statistically, three mutations in a 6-marker haplotype (see above) translate into about 380 generations, or about 9,500 years of a separate mutational evolution. Six mutations would more then double this time span. This is a very tentative estimation since a short 6-marker does not provide enough mutations. However, this might indicate how far these haplotypes are distant in time. More detailed estimates are given below in this section.

- The most frequent 12-marker haplotype of the present day general G* population is

  14-22-15-10-14-15-11-13-11-12-11-29

  It differs from that of the ancestral Jewish G1, G2 ("ancient") and G2a3 ("recent") and G2c (= G) haplotypes, which are, respectively

  14-23-15-10-14-16-11-12-11-13-12-29
  14-22-15-10-14-15-11-13-12-12-11-29
  13-22-15-10-11-16-11-12-11-12-11-29
  13-23-15-10-13-16-11-12-11-14-11-32

  by 5, 1, 6, and 10 mutations, respectively, that translates approximately into 5,000, 1,000, 6,000, and 10,000 years of mutational evolution. In other words, the most frequent present day haplotypes of G* haplogroup are the closest to the ancient G2 Jewish haplotypes, and are the most remote from the recent G2c (= G) Jewish haplotypes with an ancestor only a few hundred years back.

6-marker haplotypes (the "scientific" and "commercial" sets)

Fig. 53 shows a combined Jewish G/G2 haplotype tree which was composed based on published data (Behar et al, 2004) for 34 haplotypes G (M201) and 10

haplotypes G2 (P15). 23 of them are base (ancestral) G haplotypes, identical to each other

15-12-23-10-11-13

They sit around the "trunk" of the tree along with only three slightly mutated haplotypes from the same group.

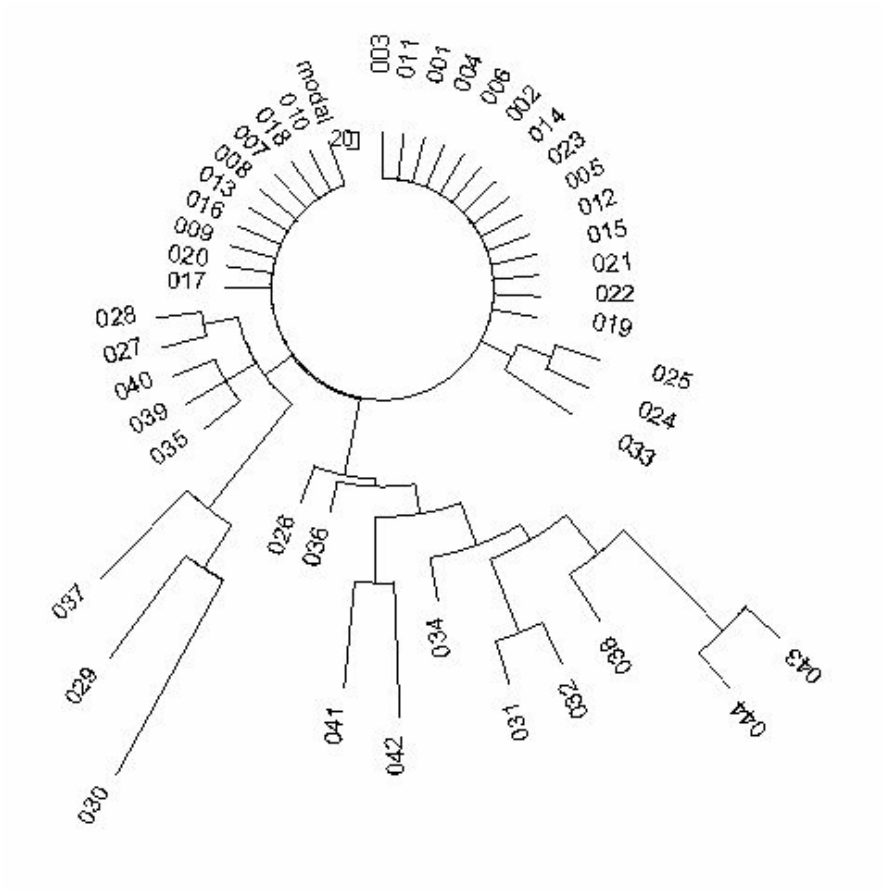

Figure 53. The 6-marker haplotype tree for 44 Jewish haplotypes of haplogroups designated as G (haplotypes 1-34) and G2 (haplotypes 35-44). A "scientific" set (Behar et al, 2004)

All ten haplotypes G2 are located on two branches-protuberances, along with 8 haplotypes designated as G in (Behar et al, 2004). These eight haplotypes are either mistyped and in fact are of G2 haplogroup, or represent a different G lineage, with a separate ancestor, who lived much earlier compared with that for

26 “younger” haplotypes (the latter is more likely). A common ancestor of those 26 haplotypes of haplogroup G lived ln(26/23)/0.0096 = 13 generations ago, or 3/26/0.0096 = 12 generations ago, or only a few centuries BP. In fact, more extended haplotypes change this estimate to 23±2 generations, that is 575±50 years BP. Since the both figures are practically equal to each other (though obtained in two principally different approaches), it points out at a single ancestor of those 26 individuals in haplogroup G.

The G2 branch, located at the lower right corner in Fig. 53, is clearly an ancient one, and has the following 6-marker base haplotype

15-12-23-10-11-14

which likely represents a mix of G and G2 haplotypes. The 10 haplotypes (six G2 and four G) have 17 mutations, which translates into 215 generations, that is 5,400 years to their common ancestor. More extended haplotypes decrease this estimate to 4,000±200 years BP, which is not critical in this context.

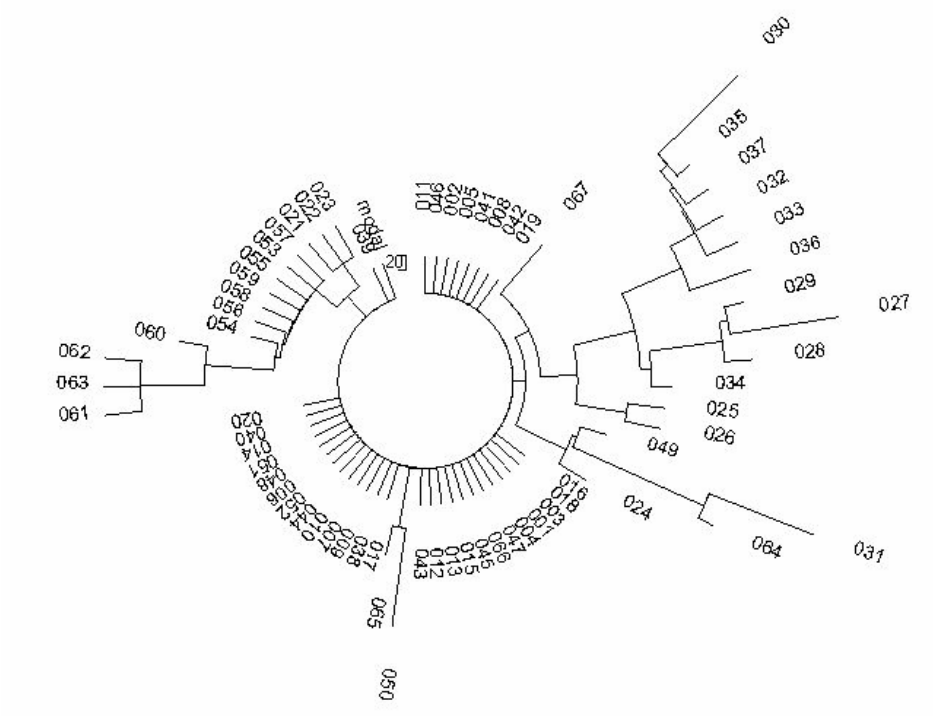

Figure 54. The 6-marker haplotype tree for 67 Jewish haplotypes of haplogroups G, G1, G2 and G2c (former G5). A “commercial” set (YSearch database)

The same in kind pattern is observed with the haplotype tree composed from 66 Jewish G* haplotypes (G, G1, G2, and G2c) extracted from YSearch database (the "commercial set"), and showed in Fig. 54. There are 34 identical, base haplotypes on the tree

15-12-23-10-11-13

out of 39 haplotypes in the series, and all of them belong to G and G2c haplogroups. In fact, these haplogroups cannot be resolved in 6-marker haplotypes, and, as it is shown later, they cannot be resolved in 12-, 37-, and even in 66-marker haplotypes either. It seems that G2c haplogroup and G haplogroup is the same thing, at least in terms of their haplotype structure, location on a tree, their common ancestor, and time span to their common ancestor.

All 39 haplotypes in the G/G2c series contain only 6 mutations. This give a tentative (with 6-marker haplotypes) estimate of ln(39/34)/0.0096 = 14 generations, and 6/39/0.0096 = 16 generations to a common ancestor of these 39 individuals. These 15±2 generations fit pretty well to 13±1 generations, obtained above, in the "scientific" set, that again indicates how similar are results obtained from "scientific" and "commercial" data bases.

Besides G/G5 haplotypes, which sit tightly around the "trunk", the tree contains two more rather wide branches, one obviously ancient, of G1 haplotypes (on the right-hand side of Fig. 54), another one moderately old, of G2 haplotypes (on the left-hand side in the same Figure). The ancient G1 branch also contains a number of G haplotypes – either due to a mistyping, or a poor resolution of haplotypes in the 6-marker format.

This 17-haplotype G1/G branch on the right-hand side contains 39 mutations with respect to the base haplotype

14-12-23-11-12-13

that corresponds to 304 generations, or 7,600 years to a common ancestor. This particularly large time span might reflect a mix of G and G1 haplotypes, hence, a more ancient common ancestor. The 14-haplotype branch on the left is split into two parts, which, as will be shown later using more extended haplotypes, represent both an ancient and a rather recent lineages.

10-marker haplotypes (the "scientific" set)

This is a set of 44 haplotypes of a maximum length for haplogroups G and G2 published in the scientific literature (Behar et al, 2004). The respective haplotype tree is shown in Fig. 55. The tree obviously shows two series of haplotypes, of haplogroup G around the "trunk", with a base haplotype

13-23-15-10-X-Y-11-12-11-14-11-32

and a wide branch of haplogroup G2, with a base haplotype

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

Figure 55. The 10-marker haplotype tree for 44 Jewish haplotypes of haplogroups designated as G (haplotypes 1-34) and G2 (haplotypes 35-44). A "scientific" set (Behar et al, 2004)

There are 14 base haplotypes in the total of 23 haplotypes of haplogroup G. All 23 haplotypes contain only 9 mutations. This gives ln(23/14)/0.025 = 20 generations, and 9/23/0.025 = 16 generations to the common ancestor. This value

of 18±3 generations is a reasonably good fit, and points to a single G ancestor who lived rather recently, in this millennium. It also shows a reasonably good fit with 15±2 and 13±1 generations obtained above from the 6-marker haplotype trees. The two missing alleles for the markers of DYS#385a,b, which have not been identified in the original publication (1), are equal to 13-16 and 14-15, respectively, as will be shown below, when 12-marker haplotypes are considered.

An ancestor of the haplotypes of G2 haplogroup in Fig. 55 lived much longer time ago. The 14-haplotype branch on the right contains 48 mutations, that corresponds to 236 generations, or about 5,900 years BP. This estimate will be further corrected below using more extended haplotypes.

12-marker haplotypes (the YSearch set)

These haplotypes arranged as a tree (Fig. 56) give a much better resolution of the three lineages: G (=G2c), G1, and G2.

G2c haplogroup, a common ancestor of 575±50 years BP

The tightly mixed G/G2c branch on the left-hand side of the tree contains 40 haplotypes, and 23 of these are identical

13-23-15-10-13-16-11-12-11-14-11-32

This base, ancestral haplotype is identical for both G and G2c (former G5) haplogroups. All the 40 haplotypes sitting around the "trunk" of the tree contain 22 mutations. These data lead to ln(40/23)/0.024 = 23 generations and 22/40/0.024 = 23 generations to a common ancestor, that is 575 years BP. An excellent fit between time spans obtained with these quite different methods, using a fraction of residual base haplotypes and mutations in haplotypes, shows that there was a single ancestor for all these 40 individuals of haplogroups G and G2c. Apparently, it is the same haplogroup, and bearers of G haplogroup have not been typed for G2c haplogroup.

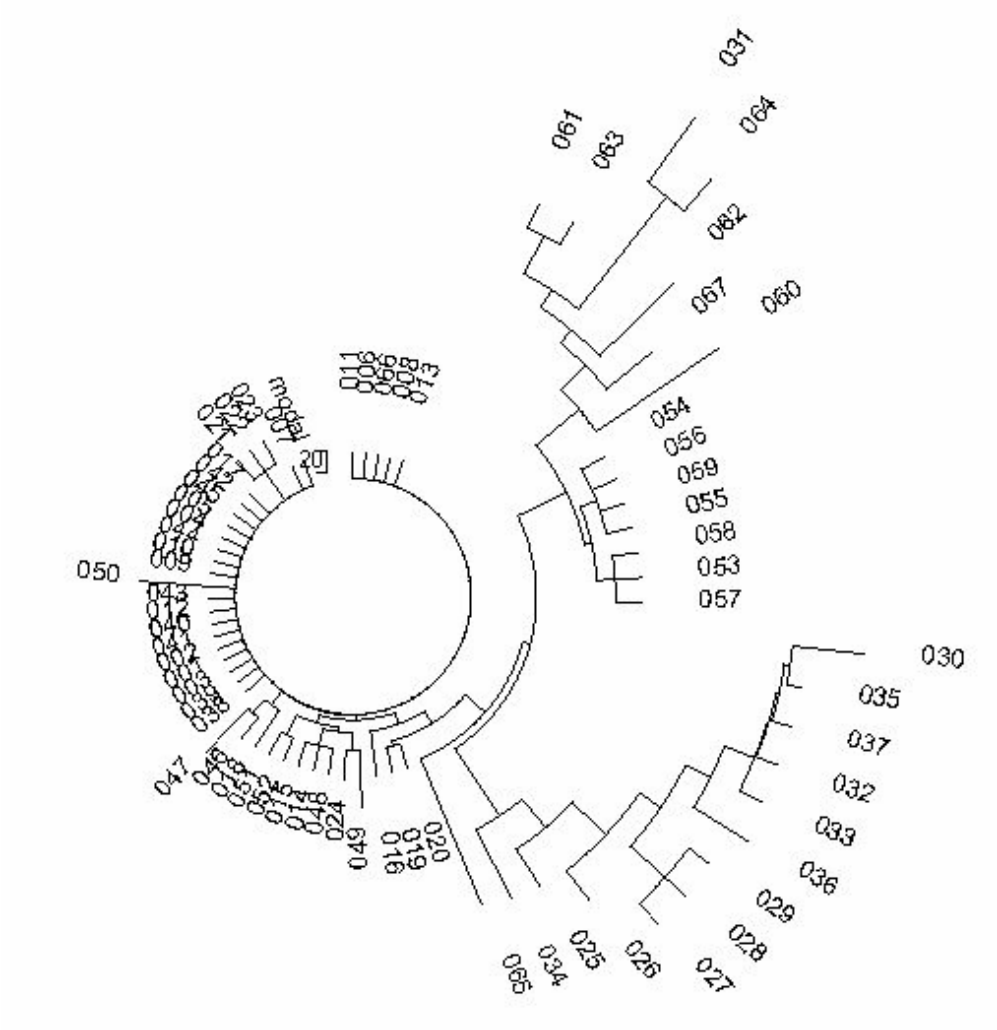

Figure 56. The 12-marker haplotype tree for 67 Jewish haplotypes of haplogroups G, G1, G2 and G2c (former G5). A "commercial" set (YSearch database)

G1 haplogroup, common ancestors of 975±175 and 3,600±200 years BP

The wide branch in the lower right corner of Fig. 56 consists of two parts. The flat part contains three haplotypes of the tested G1 haplogroup (032, 035 and 036). Next to them three more G* haplotypes (030, 033, and 037) are located, apparently being "undertyped" and likely of G1 haplogroup as well. Adjacent to

them are seven more haplotypes, apparently of haplogroup G1 too. A base (ancestral) haplotype of the flat G1 branch is

13-23-14-11-14-15-11-12-11-12-12-28

A base haplotype of the clearly older series of seven haplotypes is two mutations away:

14-23-14-10-14-15-11-12-11-12-12-28

The 6-haplotype flat branch has only 6 mutations, that corresponds to 43 generations to a common ancestor. Along with 43, 40 and 29 generations, obtained with 37-marker haplotypes, it makes an average 39±7 generations, that is 975±175 years from the common ancestor.

The 7-haplotype series of allegedly G1 haplotypes has 21 mutations, which translates to 143±8 generations, that is about 3,600±200 years to a common ancestor. There are no 37-haplotypes available for this branch sub-branch.

A mutational difference between these two base haplotypes is 2,250 years. Hence, the 975±175 yBP ancestor could be very likely derived from the 3,600±200 yBP ancestor.

A difference between common ancestor of the Jews of haplogroups G (=G2c) and of these two G1 is tremendous, 11 mutations away to each one of them, on 12-marker ancestral (base) haplotypes. This formally translates to 872 generations, that is 21,800 years of their mutational difference. Since the G (=G2c) Jewish ancestor lived only 575±50 years BP, and the G1 Jewish ancestors lived 3,600±200 years BP, the G (= G2c) ancestor cannot be derived from the G1. Both G2c and the earlier G1 ancestor descended from an ancestor who lived 13,000±200 years BP.

G2 haplogroup, common ancestors of 4,000±200 years BP and 625±200 years BP

In the upper right-hand side of Fig. 56 a combined branch of G2 haplogroup is located. It consists of a sub-branch of 7 haplotypes, descended from an ancient ancestor, and another sub-branch, 7 haplotypes, descended from a rather recent ancestor. Their base haplotypes are as follows, respectively:

14-22-15-10-14-15-11-13-12-12-11-29

13-22-15-10-12-16-11-12-11-12-11-29

The second, younger base haplotype likely belongs to G2a3 subgroup of haplogroup G*. These two sub-branches contain 23 and 3 mutations, respectively, which translates into 159 and 18. The second figure, of apparently G2a3 group, was then corrected using 37-marker haplotypes to 625±200 years BP. The first figure is related to an ancient ancestor who lived 4,000±200 years BP. A mutational distance between these two ancestral haplotypes is equal to 292 generations, that is 7,300 years. Again, the recent G2 (or G2a3) could not descend from the ancient G2. Their common ancestor lived 6,000±400 years BP.

In the same manner it can be calculated that a distance between common ancestors of Jews of haplogroups G (=G2c) [575±50 yBP] and the "ancient" G2 [4,000±200 yBP] is 806 generations, that is 20,200 years, and their common ancestor lived 12,400±200 years BP.

A distance between common ancestors of Jews of haplogroups G (=G2c) [575±50 yBP] and the rather recent G2a3 [625±200 yBP] is 449 generations, that is 11,200 years, and their common ancestor lived 6,250±150 years BP.

A distance between common ancestors of ancient branches of haplogroup G2 [4,000±200 yBP] and G1 [3,200±500 yBP] is 419 generations, that is 10,500 years. This places a common ancestor of these two haplotypes to 8,800±400 years BP.

Finally, a distance between common ancestors of a rather recent G2a3 and an ancient G1 is 545 generations, that is 13,600 years. This places a common ancestor of these haplogroups to 8,700±500 years BP.

It can be concluded that haplogroup G* with its sub-groups (G1, G2, G2c and possibly G2a3) had emerged 15,200 ±200 years BP. However, these haplogroups had entered the Jewish community much later: G2 at 4,000±200 years BP, G1 at 3,600±200 years BP and re-entered 975±175 years BP, G2c 575±50 years BP, and a recent branch of haplogroup G2a3 had entered the Jewish population 625±200 years BP.

37-marker haplotypes (the YSearch set)

The 37-marker 32-haplotype tree (Fig. 57) provides a better resolution of three nodes of haplotypes from three very distant common ancestors, each for each node. As it was shown in the preceding section, the ancestors of these nodes are practically unrelated within the last several thousand years. All 37-marker base haplotypes of haplogroups G (= G2c), G1, G2, and G2a3 are shown in the Table.

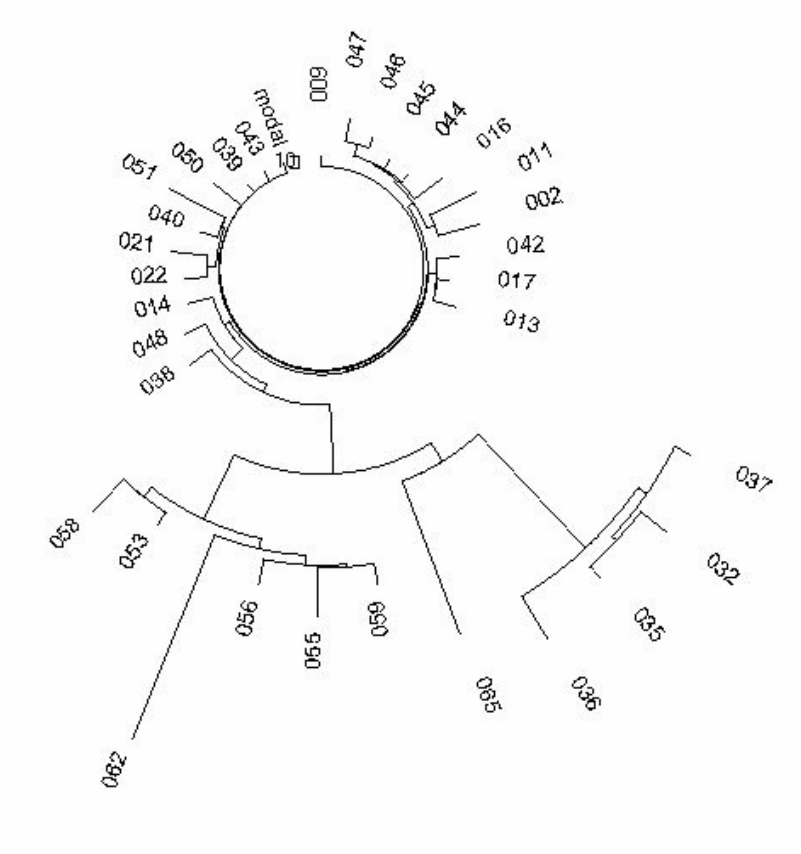

Figure 57. The 37-marker haplotype tree for 32 Jewish haplotypes of haplogroups G, G1, G2 and G2c (former G5). A "commercial" set (YSearch database)

A rather recent ancestor of 575±50 years BP, haplogroup G/G2c

21 haplotypes of the upper node (haplogroup G/G2c) on the tree contain only 11 mutations in the first, 12-marker panel; 21 mutations in the 25-marker panel; and 46 mutations in the complete 37-marker panel. This corresponds to 22, 22 or 25 generations to a common ancestor. The 66-marker haplotypes, considered in the next section, add four more estimates, that is 25, 20, 23 or 25 generations to the common ancestor. The average of these seven figures gives 23±2 generations, that is 575±50 years to the common ancestor for the upper node of haplotypes in Fig. 57. Its base, ancestral haplotype in a 12-marker format is

13-23-15-10-13-16-11-12-11-14-11-32

A rather recent ancestor of 975±175 years BP, haplogroup G1

A flat sub-branch of haplotypes of haplogroup G1, in the lower right-hand side part of the tree in Fig. 57, has only four haplotypes. Its base, ancestral haplotype in a 12-marker format is

13-23-14-11-14-15-11-12-11-12-12-28

Its 37-marker haplotype is shown in the Table. The four haplotypes contain 4 mutations in the first, 12-marker panel; 7 mutations in the 25-marker panel; and 10 mutations in the complete 37-marker panel. This corresponds to 43, 40 or 29 generations, and along with 43 generations obtained with 12-marker haplotypes, gives 39±7 generations, that is 975±175 years to a common ancestor.

An ancient ancestor of 3,600±200 years BP, haplogroup G1

This ancestor has the following 12-marker haplotype

14-23-14-10-14-15-11-12-11-12-12-28

There only one haplotype (065) left on the branch, sitting in an isolation on the tree. The distance to the common ancestor was calculated in the preceding section, using 12-marker haplotypes.

A rather recent ancestor of 625±200 years BP, haplogroup G2a3

Only six G2* haplotypes left on the branch (lower left corner in Fig. 57), five of them are derived from a rather recent ancestor, of G2a3 group, and only one left from an ancient ancestor. These five haplotypes contain only two mutations in the first 12-marker panel (that is, in 60 markers), compared to the base haplotype

13-22-15-10-12-16-11-12-11-12-11-29,

7 mutations in the 25-marker panel, and 14 mutations in all five 37-marker panels. This, including data with 12-marker haplotypes, translates into 17, 31, 32, and 18 generations to a common ancestor, on average 25±8 generations, that is 625±200 years.

66-marker haplotypes (the YSearch set)

Fig. 58 shows the 66-marker haplotype tree, consisting of 17 haplotypes. Those which belong to G (=G2c) cannot be separated even in the 66-marker format. As it was shown in the preceding section, their common ancestor lived 575±50 years BP.

A two-haplotype branch adjacent to it belong to different ancestors, one – of a recent ancestor of haplogroup G2 (haplotype 056), another – of allegedly ancient ancestor of G1 haplogroup (haplotype 065). They are located far away from the "trunk" (G=G2c), since, as we determined before, the difference between them and the "trunk" is about 11,200 years (G2c from G2a3) and 26,300 years (G2c from the ancient G1) of the mutational evolution. However, a common ancestor of the branch had appeared in the Jewish community only 625±200 and 4,000±200 years BP.

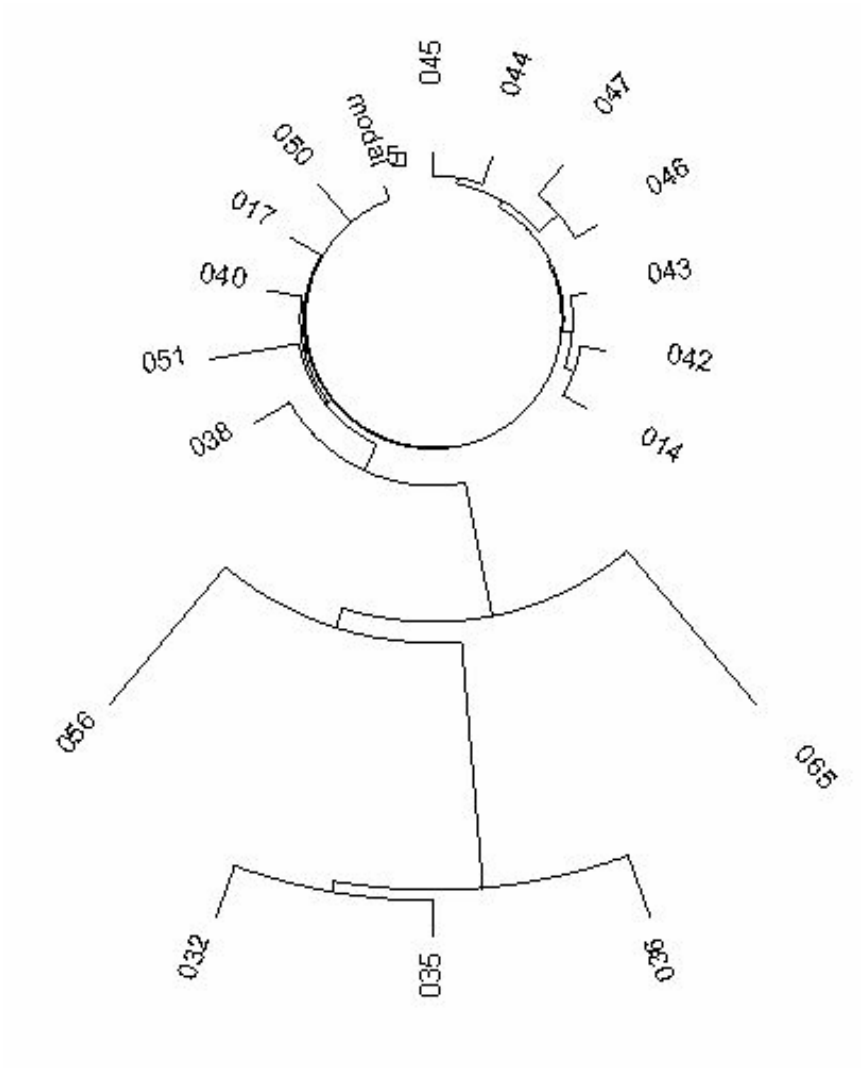

Figure 58. The 66-marker haplotype tree for 17 Jewish haplotypes of haplogroups G, G1, G2 and G2c (former G5). A "commercial" set (YSearch database)

The most remote from the "trunk" are three haplotypes of haplogroup G1 (a rather recent ancestor), with their difference by 11 mutations in their base (ancestral) haplotypes in a 12-marker format. It was determined above that this difference is equal to 26,300 years of the mutational evolution. A common ancestor of this branch had entered the Jewry much later, about 975±175 years BP.

These three principal lineages are continued in the Jewish community, with haplogroup G (= G2c) dominating in its share among the G* haplogroups (G1, G2, G2c).

Only one of these lineages seems to appear at the times of the Khazarian Khaganate, that is the G1 lineage with a common ancestor 975±175 years BP, in the 11th century AD. With some stretch one can conjecture that another haplogroup, recent lineage of G2 (625±200 years BP), within the error margin, could have also entered the Jewish community and established a new lineage in the 12th century, when the Khazars were already moving to Eastern and Central Europe.

Other Jewish G* lineages have entered the Jewish community either much earlier, that of G2 haplogroup at 4,000±200 and G1 at 3,600±200 years BP, or much later, G2c in the 15th century AD (575±50 yBP).

The Jews of Haplogroup I (and subgroups)

Haplogroup I (SNP M170, M258, P19) has appeared about 20-25 thousand years ago in Europe, and spread over the continent, particularly on the Balkans (mainly I2, SNP P215, formerly I1b) and at the Baltic Sea area (mainly I1, SNP M253, M307, M450, formerly I1a). Haplogroup I and its many subgroups are present in YSearch database in an amount exceeding five thousand haplotypes, from which only four haplogroups, namely I, I1, I2, and I2b1 (M223, formerly I1c), constitute for 94% of the total. Haplogroup I* is very rear outside of Europe, and represent nearly 20% of the European population.

The most frequent 6-marker haplotypes in these haplogroups (I*, I1, I2, and I2b1) from a list of 5520, 2006, 675, and 471 haplotypes (in YSearch database), respectively, are:

14-14-22-10-11-13
14-14-22-10-11-13
15-15-23-10-11-13
15-13-23-10-12-14

We will see below how unique are the respective Jewish haplotypes compared with the most frequent ones.

It must have some historical sense, albeit an obscured one, that despite a relatively high presence of haplogroup I* in Europe, there are only a handful of these haplogroups among the Jews. Among more than 5,000 haplotypes in Section I* in YSearch database, there are only 21 of Jewish haplotypes, at least those who bear a Jewish name. Three of them are in haplogroup I, four in haplogroup I1, five in I2, and nine in I2b1. In a work of Behar et al (2004) there are 19 ten-marker haplotypes listed under haplogroup I* (SNP P19) section. As it will be shown below, the 19 haplotypes very likely include those of haplogroups I1 (six haplotypes), I2 (seven haplotypes), and I2b1 (six haplotypes).

In short, the study below showed that most of the Jewish haplotypes of I* haplogroup fall into three haplogroups, I1, I2, and I2b1. Approximately one-third of present day Jews in haplogroup I* belong to each of those subgroups. Haplogroup I1 is an ancient one, and their ancestor among Jews (not necessarily a Jewish ancestor) lived about 4,100±1,900 years BP. Jews of haplogroup I2 descended from a common ancestor who lived 1,400±600 years BP. An ancestor of the Jews of haplogroup I2b1 (former I1c) is the most recent one among other subgroups, and lived 500±125 years BP.

6-marker haplotypes (the "scientific" and "commercial" sets)

Behar et al (2004) published a list of 19 Jewish haplotypes of I (P19) haplogroup, which will be called here the "scientific" set. A haplotype tree composed of those 6-marker haplotypes is shown in Fig. 59. In fact, in (Behar et al, 2004) 35 haplotypes of haplogroup I* are listed, including 16 haplotypes of non-Jewish Hungarians.

Three branches in Fig. 59 produce the following base (ancestral) haplotypes:

14-14-22-10-11-13
16-13-24-11-11-13
16-13-22-10-13-13

The first one is identical with the most frequent haplotype of I* and I1 haplogroups for a general population. This is an ancient base haplotype among the Jews and non-Jews, its branch is located on the right-hand side in Fig. 59, and from 6-marker haplotypes a common ancestor of this branch is estimated to live approximately six thousand years BP. This figure as it turns out is on the upper side of the estimate.

The second one is the base Jewish I2 6-marker haplotype, obtained from the left-hand side in Fig. 59. It does not match the most frequent haplotypes of the general population, listed above, and has – in that sense – a unique 6-marker haplotype. Its ancestor lived 55 generations ago, that is in the first millennium AD. This date will be refined below, with more extended haplotypes.

The third haplotype above is the base Jewish I2b1 haplotype, which is derived from a rather recent ancestor. It sits on top of the tree in Fig. 59 as a short flat branch. An ancestor of this group of the Jews lived quite recently, around 27 generations BP, that is only a few centuries ago. His 6-marker haplotype was also different from the most frequent I* haplotypes in the general population.

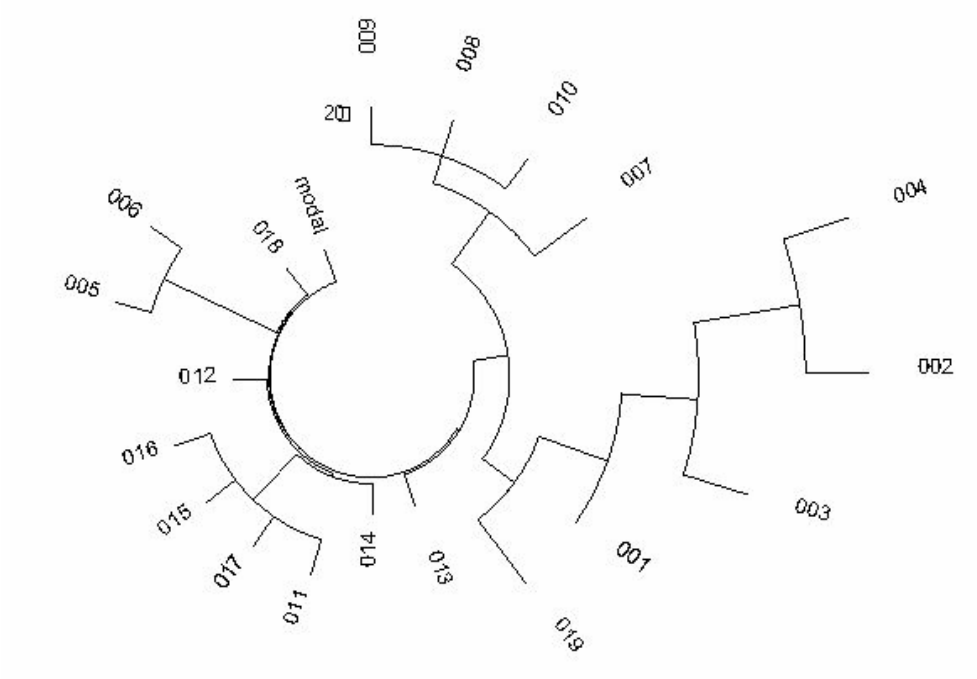

Figure 59. The 6-marker haplotype tree for 19 Jewish haplotypes of haplogroup I. A "scientific" set (Behar et al, 2004)

Fig. 60 shows that the 6-marker Jewish and non-Jewish haplotypes of I* haplogroup are practically mixed. In this Figure the Jewish haplotypes are numbered, and the Hungarian are marked "H". All the three branches have been identified (see below) as the ancient I1 on the left-hand side, I2 on the right-hand side, and I2b1 (former I1c) at the bottom. One can see that the Jewish haplotypes,

including those from the ancient branch on the left do not form any separate group.

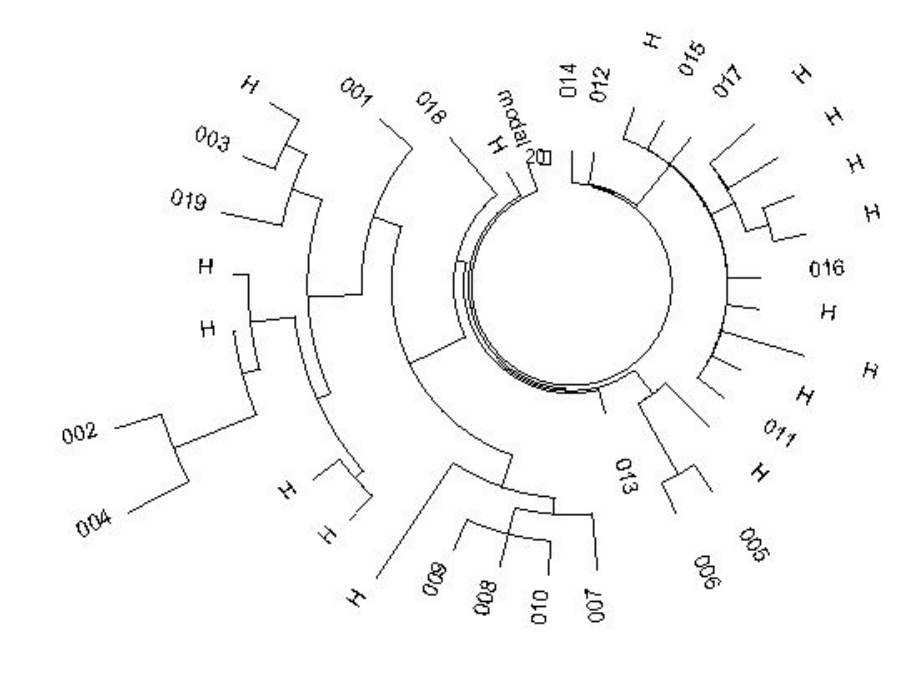

Figure 60. The 6-marker haplotype tree for 19 Jewish haplotypes and 16 non-Jewish haplotypes of haplogroup I. A "scientific" set (Behar et al, 2004)

Another series of 21 Jewish I, I1, I2 and I2b1 haplotypes was extracted from YSearch database (the "commercial" set). Unlike the Behar et al (2004) series, these haplotypes were specifically typed for the said haplogroups, which provided the base for their assignment. The haplotype tree is shown in Fig. 61. As in the "scientific" set, the "commercial" set makes a tree with three similar in kind branches, also of I1, I2, and I1b2 haplogroups. Again, a "scientific" and a "commercial" sets give principally the same results, with three base haplotypes for the branches on in the upper right corner (I1), on the left (I2) and in the lower right corner (I1b2), respectively:

14-14-22-10-11-13
16-13-24-11-11-13
16-13-22-10-13-13

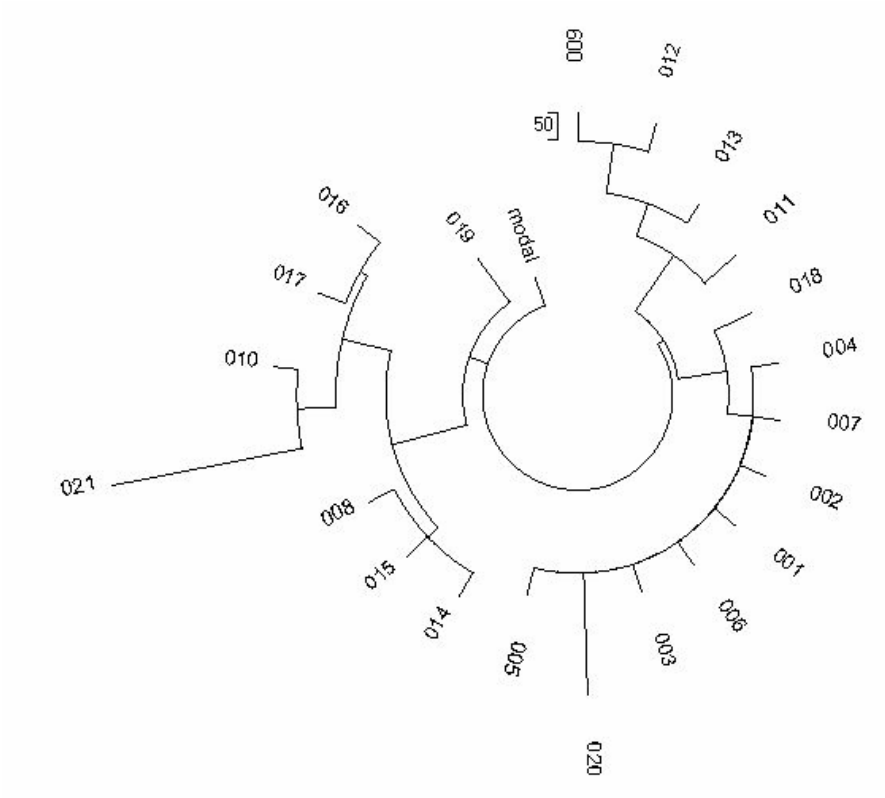

Figure 61. The 6-marker haplotype tree for 21 Jewish haplotypes of haplogroups I, I1, I2 and I2b1. A "commercial" set (YSearch database)

The separation of the haplogroups in the branches is still not complete, hence, time estimates to the common ancestors can be only tentative. In the I1 branch a common ancestor lived approximately 2,100 years ago, on the lower side of the estimate. In the I2 branch (containing also haplotypes of I and I1 haplogroups) a common ancestor lived approximately 43 generations BP, that is about 1,100 years BP. In the I1b2 branch a common ancestor lived only recently, about 500 years BP, in this millennium. The date will be refined using more extended haplotypes.

10-marker haplotypes (the "scientific"set)

Fig. 62 shows a haplotype tree composed from nineteen 10-marker haplotypes published in (Behar et al, 2004). It contains the same three branches of

haplogroup I1, I2 and I2b1, as were identified in the 6-marker haplotype tree in Fig. 59.

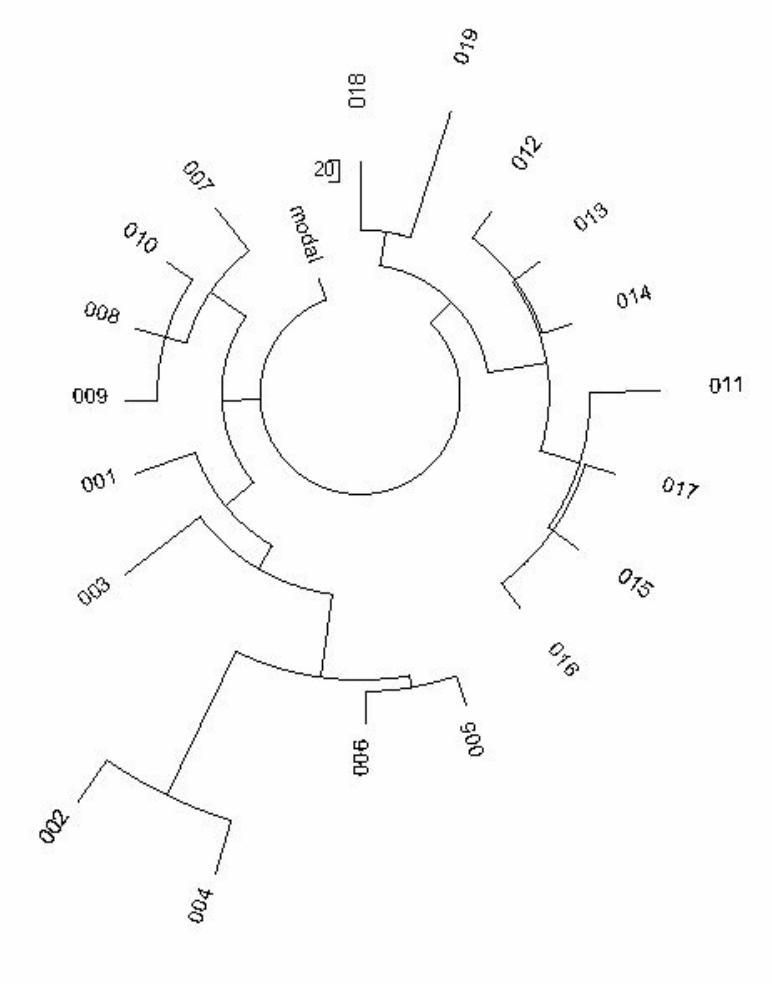

Figure 62. The 10-marker haplotype tree for 19 Jewish haplotypes of haplogroup I. A "scientific" set (Behar et al, 2004)

The base (ancestral) Jewish haplotype in I1 haplogroup is as follows:

13-22-14-10-X-Y-11-14-11-12-11-29(28)

in which missing pair of alleles in DYS#385a,b markers is 13-14, as it was established in the next section. An ancestor of this branch lived 225 generations BP, or about 5,600 years BP.

The base Jewish haplotype in I2 haplogroup is

13-24-16-11-X-Y-11-13-13-13-11-31

in which missing pair of alleles in DYS#385a,b markers is 14-15, as it was established in the next section. An ancestor of this branch lived between 1,000 and 2,000 years BP.

The base Jewish haplotype in I2b1 haplogroup is

13-22-16-10-X-Y-11-13-11-13-13-29

in which missing pair of alleles in DYS#385a,b markers is 15-16, as it was shown in the next section. An ancestor of this branch lived quite recently, since in four 10-marker haplotypes there was only one mutation compared to the above base haplotype, that is a common ancestor lived less than 20 generations BP.

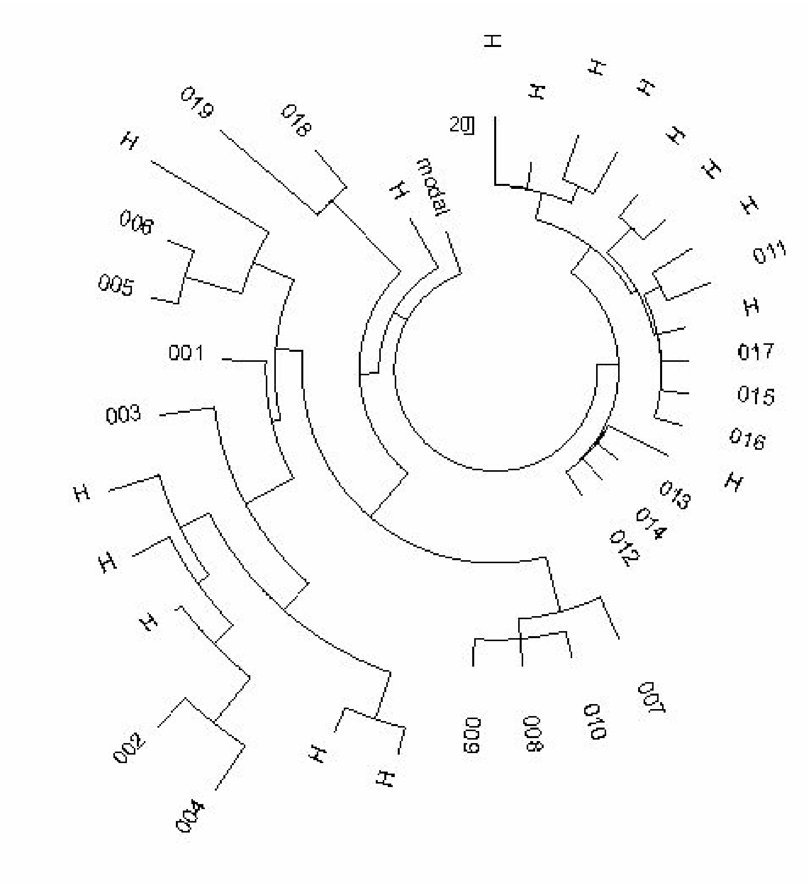

Figure 63. The 10-marker haplotype tree for 19 Jewish haplotypes and 16 non-Jewish haplotypes of haplogroup I. A “scientific” set (Behar et al, 2004)

There was no significant segregation of the Jewish haplotypes and Hungarian I* haplotypes, as shown in Fig. 63, at least with 10-marker haplotypes.

12-marker haplotypes (the YSearch set)

As practically always, a “scientific” and a “commercial” sets give very similar results. The haplotype tree, composed from 21 twelve-marker haplotypes, consists of three branches, assigned to I1, I2 and I2b1 haplogroups (Fig. 64). The respective base haplotypes are

13-22-14-10-13-14-11-14-12-12-11-28

13-24-16-11-14-15-11-13-13-13-11-31
13-22-16-10-15-16-11-13-11-13-13-29

Common ancestors of these branches lived as follows: for the I1 branch, close to 4 thousand years BP; for the I2 branch, about 650 years BP; and for the I2b1 recent branch, only 21 generations, that is about 525 years ago.

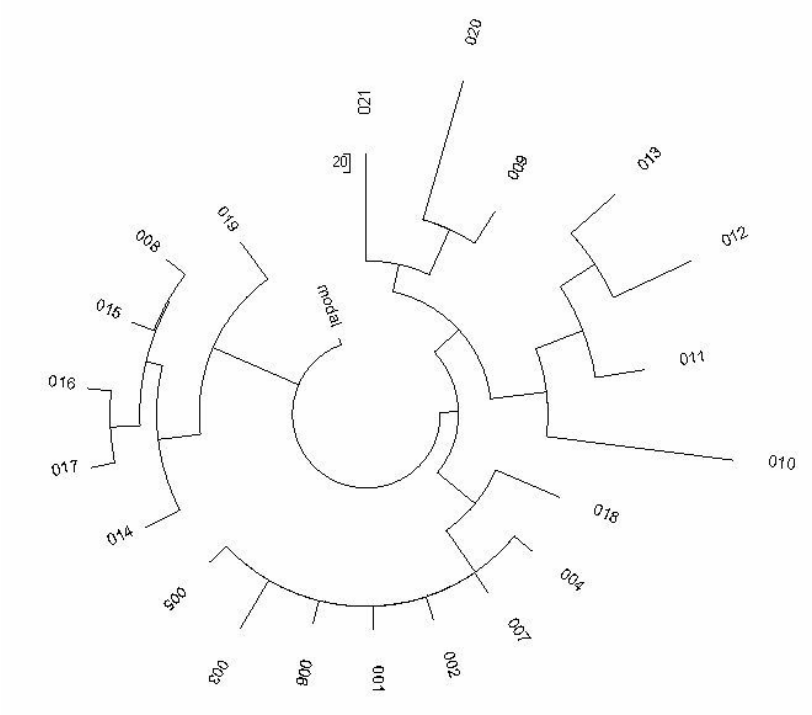

Figure 64. The 12-marker haplotype tree for 21 Jewish haplotypes of haplogroups I1, I2 and I2b1. A "commercial" set (YSearch database)

It is interesting that all the three Jewish base 12-marker haplotypes in haplogroups I1, I2 and I21b differ from each other by as many as 12±1 mutations per haplotype, that is about by 1,100±200 generations, or 27,000±5,000 years mutational distance. This means that these haplotypes all came from different common ancestors, and are not derived from each other. This follows, however, from the fact that they all belong to different haplogroups. According to the mutational distance and time spans to the common ancestors, their common ancestor lived about 16,000±3,000 years BP.

37-marker haplotypes (the YSearch set)

The 13-haplotype 37-marker tree is shown in Fig. 65. It clearly shows that the I2b1 branch, located on the top of the tree, very close to the "trunk", is derived from a very recent ancestor compared with the other two branches. A common ancestor of the I2b1 group of five individuals in the set lived 18 generations BP, when calculated from eight mutations in five 37-marker haplotypes. Considering all four figures obtained from 6-, 10, 12-, and 37-marker haplotypes, namely 27, 21, 14 or 18 generations from a common ancestor of Jewish individuals in haplogroup I2b1, we obtain 20±5 generations, that is 500±125 years to a common ancestor.

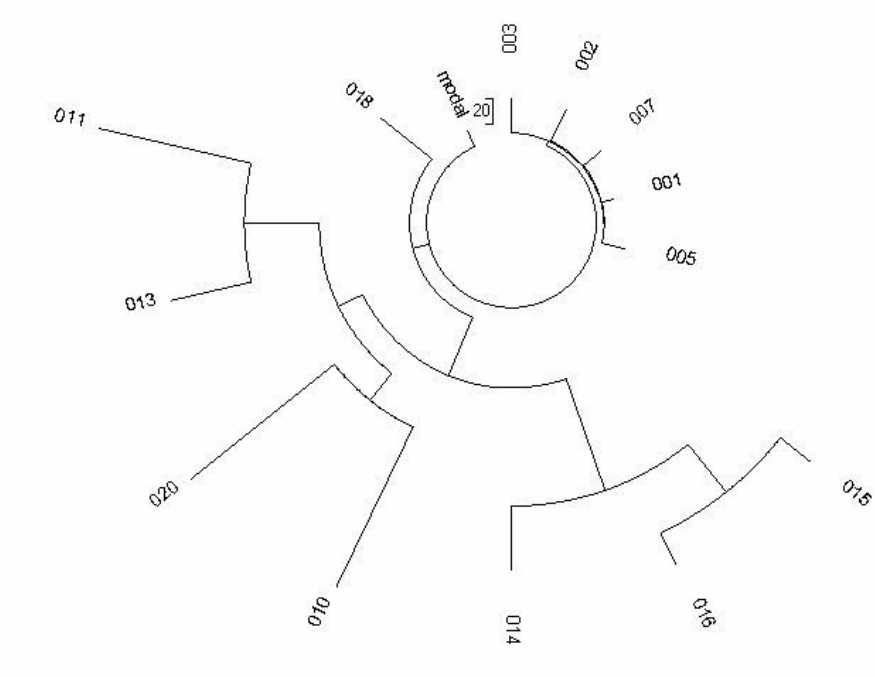

Figure 65. The 37-marker haplotype tree for 13 Jewish haplotypes of haplogroups I1 (haplotypes 010, 011 and 013), I2 (haplotypes 014, 015, 016 and 020) and I2b1 (haplotypes 001, 003, 005, 006, 007). A "commercial" set (YSearch database)

There are only two haplotypes derived from an ancient ancestor in haplogroup I1, and along with other I1 haplotypes in 6-, 12-, 25- and 37-marker haplotypes,

mutations of the group lead to a common ancestor who lived 163±78 generations BP, that is about 4,100±1,900 years to a common ancestor.

A common ancestor of the I2 branch in the lower right-hand side of Fig. 65 lived 55, 94, 43, 26, 61 or 63 generations PB, or in average 57±23 generations, that is 1,400±600 years to a common ancestor.

In a summary, the Jews of haplogroup I* are divided approximately equally between subgroups I1, I2 and I2b1. A common ancestor of the Jews of haplogroup I1 lived about 4,100±1,900 years BP. His 6-marker haplotype

14-14-22-10-11-13

is identical with the most frequent haplotype in a common I1 population.

Almost three thousand years later, already in the Diaspora, a new haplotype,

16-13-24-11-11-13

which belonged to I2 haplogroup, had entered the Jewish community. His bearer was practically unrelated, within many thousand years, to the first common ancestor of the Jewish I1 haplogroup.

Recently, in about the 16th century, a new haplotype, unrelated to the first two

16-13-22-10-13-13

and appeared from the I2b1 population, had invaded the Jewish community. This was apparently in Central or Eastern Europe, in a wave of migration of the Jews to the East. Despite the latter appearance, about one-third of present day Jews of haplogroup I* belong to this, I2b1 subgroup. This is the most compact group of the Jews in haplogroup I*, and all I2b1 representative tested for haplotypes have almost identical 37-marker haplotypes. The first 30-marker haplotypes in five tested individuals (with 19th century ancestors from Ukraine, Russia, Poland) are just identical, and have only a single mutation.

The Jews of Haplogroups K and T

Haplogroup K (SNP M9) is one of the oldest haplogroups on the current list, and has appeared allegedly 40 thousand years ago in Southwestern Asia. Its bearers then had obtained other downstream SNP mutations, and split into many other

haplogroups between L and T. There are relatively few of those individuals who still have the original M9 with no downstream snips. For example, in a set of 496 Jews (Behar et al, 2004) had only 9 haplotypes of haplogroup K, that is less than 2%. In YSearch database a section "Haplogroup K" has only 60 entries, and only 10 of them (17%) are seemingly Jewish, if to judge on first and last names of their predecessors and on comments introduced by the individuals themselves regarding their heritage.

Haplogroup T (SNP M70, M184, M193, M272), that is former haplogroup K2, is met rather rarely in the Middle East, North and West Africa, in Central Europe and in the Mediterranean region. There are 184 entries of K2 haplotypes in the YSearch database, with 20 Jewish haplogroups among them (11%).

It turned out that Jewish haplotypes of K and T haplogroups are very similar, and apparently have the same ancestral (base) haplotypes. On a common haplotype tree they are randomly mixed on the same branches, and cannot be segregated even up to 37-marker haplotypes. There are no 67-marker Jewish haplotypes of K haplogroup available in the YSearch database.

Because of that, the Jewish haplotypes of K and T haplogroups were considered in this work together, on the same haplotype trees.

The most frequent 6-marker haplotype in both K* and T* haplogroups is

14-12-23-10-13-13

and the most frequent 12-marker haplotypes in these haplogroups is

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

These haplotypes were extracted from almost two hundred 6- and 12-marker haplotypes of the K* section of YSearch database.

Our study has indicated that Jews of haplogroups K/T, represented by 39 individuals, descended from at least five common ancestors. One of them lived between 8,500±1,000 years BP. Three lived more than two thousand years BP, in about 3rd and 1st centuries BC, and the most recent one lived only approximately 300 years ago, in about the 17th century.

6-marker haplotypes (combined "scientific" and "commercial" sets)

Only nine Jewish haplotypes of haplogroup K (reportedly M9*) were published in the literature (Behar et al, 2004). Ten more Jewish haplotypes have been extracted from the K haplogroup section of YSearch database. Since it was found that K and T (former K2) Jewish haplotypes have similar sequences and are not segregated on a haplotype tree, 20 more of Jewish T haplotypes were added to the selection and combined with other 19 K and K* haplotypes.

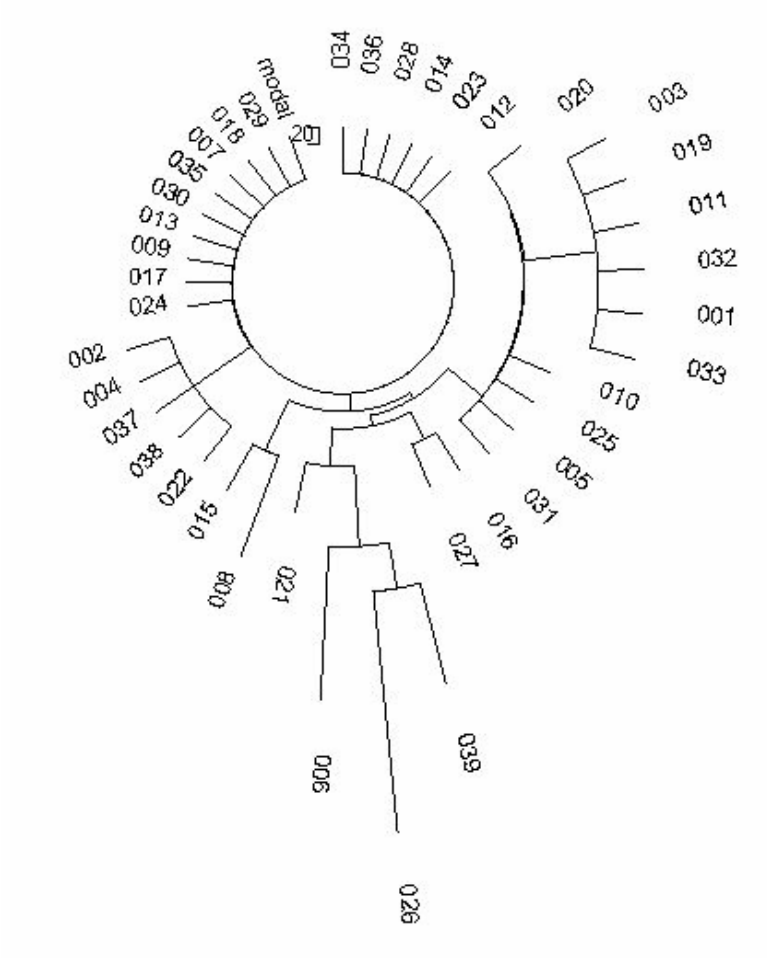

Figure 66. The 6-marker haplotype tree for 39 Jewish haplotypes of haplogroups K and T (former K2). A combined "scientific" set (Behar et al, 2004) and a "commercial" set (YSearch database)

The resulting set of 39 haplotypes is shown as a tree in Fig. 66. The tree contains 15 base haplotypes

14-12-23-10-13-13

which is identical with the most frequent haplotype in K* section of YSearch database (see above). Among these 15 base haplotypes (shown as a "comb" on top of the tree in Fig. 66), two belong to K haplogroup, ten – to T haplogroup (both are from the "commercial" set, and three were taken from the "scientific" set published in (Behar et al, 2004).

15 base 6-marker haplotypes among 39 haplotypes refer to 111 generations from a common ancestor, if a common ancestor is only one in the set: ln(39/15)/0.0096 = 100 generation (without a correction for back mutations, and 111 generations with the correction) to a common ancestor. The whole set contains 45 mutations compared to the above haplotype, that gives 45/39/0.0096 = 120 generations (without a correction for back mutations), or 136 generations (with the correction). This mismatch indicates that there were more than one common ancestor for descendants who provided the haplotypes.

The tree (Fig. 66) gives some indications to possible ancestral haplotypes. Three more can be identified from branches on the tree:

13-12-22-10-13-13
14-14-23-10-13-13
13-12-23-10-13-13

The respective branches are shown in Fig. 66 on the right (a flat 6-haplotype branch), on the left (a flat 5-haplotype branch), and at the lower right (a flat 4-haplotype branch). A 4-haplotype branch at the bottom, including one K, two T, and one K* haplotypes might contain some ancient K* lineages; its all four haplotypes have the last three markers all mutated, and its common ancestor lived approximately 13,000 years BP.

10-marker haplotypes (the "scientific" set)

Nine of 10-marker haplotypes of Jews of haplogroup K* (including T* on that matter) was published in (Behar et al, 2004). They are shown as a tree in Fig. 67.

All four base 6-marker haplotypes shown above are seen in the tree, as part of the following 10-marker base haplotypes:

13-23-14-10-X-Y-11-12-12-12-13-29
13-22-13-10-X-Y-11-12-12-14-13-31
13-23-14-10-X-Y-11-14(12)-11-14-13-31
13-23-13-10-X-Y-11-12-12-13-13-28

These base haplotypes correspond to three haplotypes (004, 005 and 006) in the lower right corner, two haplotypes (002 and 003) at the bottom, two haplotypes (007 and 008) on the left, and one haplotype (001) on the top. As it will be shown below, the missing pairs of alleles of markers DYS#385a,b in publication (Behar et al, 2004) all have been identified using the "commercial" 12-marker haplotype set. They are respectively equal to 15-16, 13-16, 14-16, and 13-15.

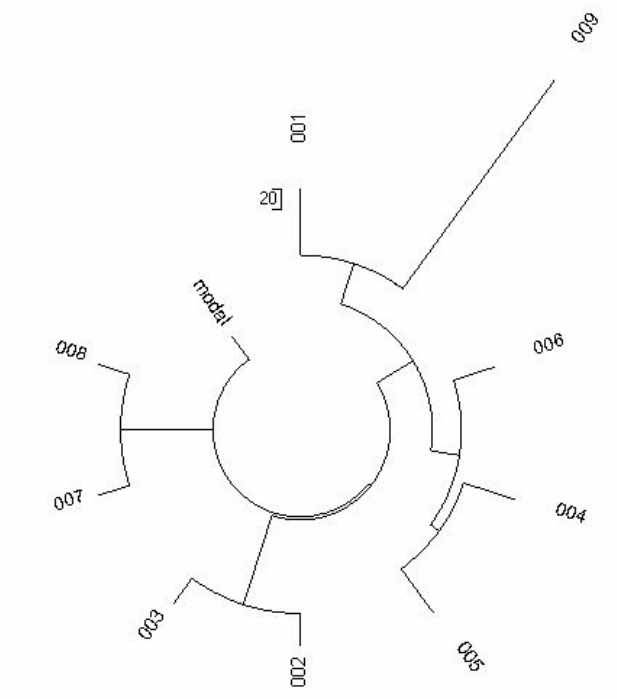

Figure 67. The 10-marker haplotype tree for 9 Jewish haplotypes of haplogroups K and T (former K2). A "scientific" set (Behar et al, 2004)

The tree is too small for any meaningful calculations, and is shown here just to illustrate one more time that "scientific" and "commercial" (YSearch) sets of haplotypes lead to the same haplotypes and practically the same results.

12-marker haplotypes (the YSearch set)

There are four principal branches of haplotypes on the 12-marker set (Fig. 68), containing 30 haplotypes. Ten haplotypes in the set were of K haplogroup, and 20 were T haplogroup.

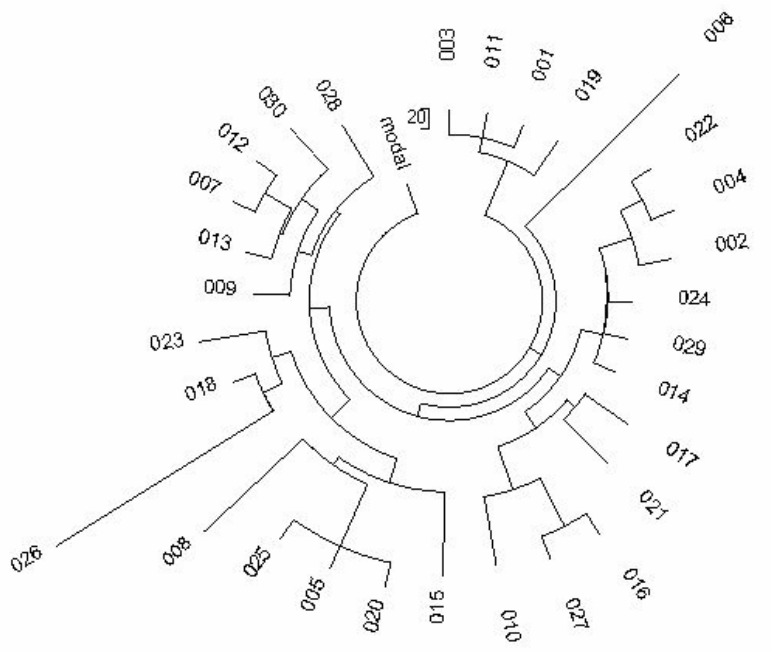

Figure 68. The 12-marker haplotype tree for 30 Jewish haplotypes of haplogroups K (ten haplotypes) and T (former K2, twenty haplotypes). A "commercial" set (YSearch database)

The branch which was a "base" one and was sitting on the "trunk" of the tree partially became a 6-haplotype tight branch in the upper left corner in Fig. 68, with the base haplotype

13-23-14-10-15-16-11-12-12-12-13-29

These 6 haplotypes contain 9 mutations, that is only 1.5 mutations per haplotype. However, the 37-marker format reveals more mutations in these haplotypes, namely from 3.7 to 5.0 mutations per similar size haplotype. This gave on average 181±18 generations to a common ancestor, that is 4,500±500 years BP. The branch contains two haplotypes of K and four haplotypes of T haplogroup.

Four haplotypes on the top are derived from a rather recent ancestor. They have only one mutation in 48 markers, which corresponds to just 300 years to a common ancestor for these four Jewish individuals. He had the following 12-marker (base) haplotype

13-22-13-10-13-16-11-12-12-14-13-31

These four recent haplotypes belong to two K and two T haplogroups. This is one more illustration that 12-marker haplotypes are practically indistinguishable in these two haplogroups.

The 11-haplotype branch on the right-hand side contain two K and nine T haplotypes. Their common base (ancestral) haplotype is

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

A common ancestor of this 11 individuals lived 114 generations BP. When results obtained with 37-marker haplotypes from this branch are added, a number of generations to the common ancestor becomes on average 91±24, that is about 2,300±600 years BP. This 12-marker haplotype is identical to the most frequent haplotype in the general population of bearers of K* haplogroup (see above). Indeed, this branch includes a third of all haplotypes in the set.

The last branch is located on the lower left corner, and contains two K and three T haplotypes. Its base haplotype is

13-23-13-10-13-15-11-12-12-13-13-28

All the five haplotypes contain 11 mutations compared to the base haplotype, which translates to 101±8 generations, that is about 2,500±200 years to a common ancestor.

37-marker haplotypes (the YSearch set)

Only ten Jewish haplotypes are available in the 37-marker format from the initial thirty in the 12-marker format in section K and K2 (now T) in YSearch database. A respective haplotypes tree is shown in Fig. 69.

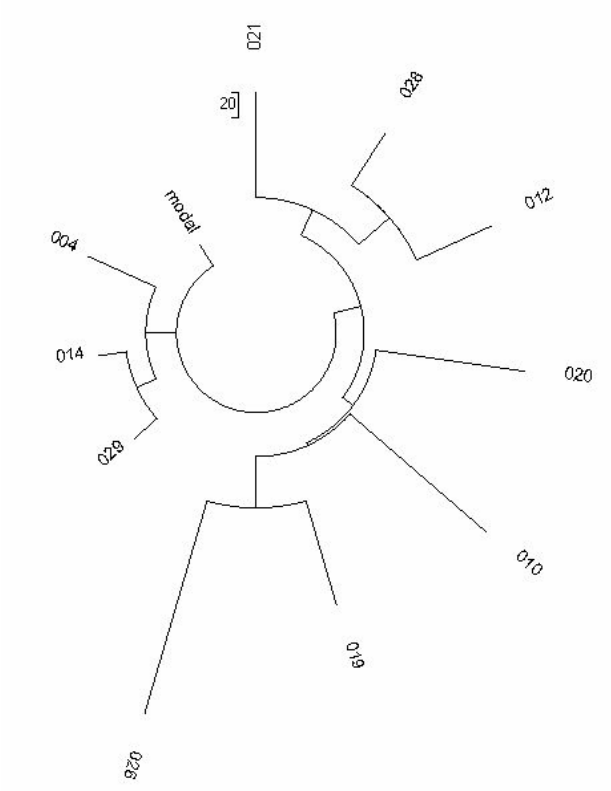

Figure 69. The 37-marker haplotype tree for 10 Jewish haplotypes of haplogroups K (haplotypes 004 and 026) and T (haplotypes 010, 012, 014, 019, 020, 021, 028 and 029). A "commercial" set (YSearch database)

The tree contains two main branches, one of them (on the right-hand side) splits into two sub-branches. These branches were indentified in the preceding section (12-marker haplotypes) in terms of their base (ancestral) haplotypes and a time span to their common ancestors, and now they the assignments will be refined. The 3-haplotype branch on the left-hand side consists of two T haplotypes and one K haplotype. Obviously, there is no resolution in this particular case between these two haplogroups even in a 37-marker format. A common ancestor of these three individuals (whose 19th-century paternal line goes to different areas of

Ukraine and Austria, some are reportedly Rabbinical) lived – by different calculations based on pattern of their mutations – 58, 104, or 89 generations BP, which along with 114 generations obtained from 12-marker haplotypes (see above) gives an average value of 91±24 generations, that is about 2,300±600 years BP. His 12-marker ancestral haplotype was

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

and the 37-marker haplotype is shown in the Table.

A 5-haplotype branch in the lower left corner in Fig. 45 (12-marker haplotypes) reduced to only one haplotype in Fig. 69. In 12-marker format it gave an estimate of 2,500±200 years to the common ancestor. However, it was a principally different common ancestor from that one who lived 2,300±600 years BP (see above) and had a different 12-marker haplotype

13-23-13-10-13-15-11-12-12-13-13-28

In a 12-marker format both haplotypes differ by 8 mutations, which corresponds to about 12 thousand years of a separate mutational history. In their refined, 37-marker format (see Table 1) these two haplotypes differ by 26 mutations, which translated to about 10,000 years of a different mutational evolution. Hence, these two haplotypes have different ancestors.

Another branch, in the upper right corner in Fig. 69, also contains three haplotypes, all three belong to T haplogroup. Their paternal line, dated by the 19th century, goes from Latvia and Ukraine, and their common ancestor lived 180, 200 or 164 generations BP, or on average 181±18 generations, that is 4,500±500 years BP. He had the following 12-marker haplotype

13-23-14-10-15-16-11-12-12-12-13-29

and his 37-marker haplotype is given in the Table.

The last one, a wide 4-haplotype branch is located at the bottom in Fig. 69. It contains three haplotypes of haplogroup T, and one haplotype of haplogroup K* (026). The last one is very mutated, and might represent one of deeper subclades of K*. Overall, this branch contains as many as 22 mutations in the first 12-marker panel, 50 mutations in the 25-marker panel, and 92 mutations in all four 37-marker haplotypes. On average, this gives 338±39 generations, that is about 8,500±1,000 years to a common ancestor. Paternal lines of these four individuals, dated by the 19th century, go from Austria, Russia, Ukraine and

Burma, and one individual noted his lost Ashkenazi identity). The ancient common ancestor had the following 12-marker haplotype

13-23(22)-13-10-12(13)-16(15)-11-12-12-13(14)-13-30(28)(31)

In summary, there are not many available Jewish haplotypes of K and T haplogroups, and their typing sometimes leaves uncertainties. Nevertheless, some tentative conclusions can be made. The earliest common ancestors for some Jewish individuals among 39 considered in this study, lived between 8 and 13 thousand years BP. Their descendants, initiating new lineages within K and T haplogroups, lived 4,500±500, 2,500±200 and 2,300±600 years BP. Each of these lineages include about one-third of presently living Jews of haplogroups K and T. All three lineages came from different ancestors and cross as long as many thousands years BP. Indeed, their 12-marker haplotypes differ by 6 and 8 mutations, that correspond to about 8 to 12 thousand years of mutational difference between them (with a correction for back mutations). Several individuals from the list have a common ancestor who lived as recently as three hundred years ago. They all have different last names and live in different, albeit neighboring countries.

A Conclusion

A large fraction of haplotypes of the Jews have an ancient origin, and have appeared in the Middle East. First of all this is related to tribes/haplogroups which are designated with letters J1, J2, E3b, and embrace about two third of nowadays Jews. These tribes can be traced up to 500-600 generations in depth, that is approximately 12 to 14 thousand years BP. They have left descendants whose descendants survived and live all over the world. Naturally, the tribes themselves could have appeared much earlier, however, in many cases their offspring and/or their descendants did not survive.

Another third of the Jews according to their Y-chromosome belong to other tribes/haplogroups which historically are not commonly associated with bearers of the Jewish traditions, customs, religion. These tribes are such as R1a1, R1b, R2, Q, G, K, I. In each case a story of a penetration of those tribes to a Jewish community presents a natural mystery, a secret of which was carried to a grave long ago. It could be a romantic story, it could be a story far from a romantic one. Some of the stories happened three-to-four thousand years ago, when Jewry was only being established. It was time of an early unification and consolidation of different peoples based on common beliefs, on common lifestyle, on common

values. In some cases an entry of a new haplogroup and new haplotypes to a Jewry happened to be in 8-11th century AD; it well might be times of consolidation of some part of the Jewish Diaspora in the Khazar Kaganat. In some cases new entries of haplogroups and haplotypes happened only 300-400 years ago; these were times of shaping of Jewish communities in Eastern Europe, and there were sometimes accompanied either with demographic expansions among the Jewry, or with passing by them a population bottleneck. Some striking cases explained in this study have been apparently related to the Black Plague in Europe in the 14th century.

Amazing as it is, all of it is written in our DNA. Rather, a scribble on the cuff of the DNA.

There is a deep meaning that these scribbles, if to looks at them for a single individual, without comparisons with others, do not say much. They represent just a string of numbers. However, when compared with those in other people, these scribbles start talking stories. These stories are about origins of mankind, appearance of tribes, their migrations, about our ancestors, their contribution to current populations.

This is what this story on the origin of the Jews is about. In the beginning of each tribe of the Jewry, as well as of each other tribe on that matter, was a single individual, who had survived against many odds and left an offspring, some of them also survived. We all are living proofs of it. For example, in the beginning of all contemporary Jewish people of about 13 million, half of them are males, there were only ten individuals, ancestors of each haplogroup.

And an attempt to clarify “origin of the Jews” is an attempt to visualize those ancestors, times of their lives, places where they lived, where they came from to those places. It is an attempt to describe tribal communities of the modern day Jews, their present numbers, their place in the contemporary world in terms of DNA genealogy.

This was apparently the first attempt of this kind. I apologize before the readers for errors I have inevitably made. The statistics is often poor, haplotype sets are often small, calculations are sometimes rather primitive. I have done my best (well, close to it) and those who can do better let do it better.

Table

Base haplotypes of the Jews in various haplogroups, and their main lineages

Haplogroup	Base haplotype number (see notes at the Table below)	Base haplotype	Generations (years) to a common ancestor
J1	1	**14-16-24-10-11-12**	444±24 (11,100±600)
	2	**14-16-24-10-11-12** 12-24-14-10-12-17-11-16-13-13-11-30-18-8-9-11-11-26-14-20-29-14-15-16-16-10-9-20-22-15-13-17-18-32-36-12-10	368±50 (9,200±1,200)
	3	**15-15-23-10-11-12**	248±16 (6,200±400)
	4	**15-15-24-10-11-12**	204±20 (5,100±500)
	5	**15-16-24-9-11-12**	140±8 (3,500±200)
	6	**14-15-23-10-11-12** 12-23-14-10-13-19-11-15-13-14-11-31-17-8-9-11-11-26-14-20-25-13-15-16-16-11-10-19-22-14-12-16-18-33-35-12-10	72±8 (1,800±200)
	7	**14-16-25-10-13-12** 12-25-14-10-16-18-11-16-10-13-13-29-17-8-9-11-11-26-14-20-28-13-14-14-17-11-9-19-22-17-14-18-18-31-37-12-11	65±9 (1,600±200)
	8	**14-16-23-10-11-12** **(The "older Cohen Modal Haplotype")**	66±14 (1,650±350)

		12-23-14-10-13-17-11-16-11-13-11-31-18-8-9-11-11-25-14-20-25-12-14-16-17-11-10-22-22-15-13-17-17-32-36-12-10	
	9	**14-16-23-10-11-12** **(The "recent Cohen Modal 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	54±4 (1,350±100)
	10	**13-15-24-10-11-12** 12-24-13-10-12-19-10-15-13-12-11-29-18-8-9-11-11-24-14-20-32-14-15-15-16-10-9-19-21-15-13-17-17-32-38-14-10	20±2 (510±50)
J2	11	**15-15-23-10-11-12** 12-23-15-10-14-16-11-15-12-13-11-29-15-8-9-11-11-26-15-20-30-13-13-15-16-10-10-19-22-16-14-17-18-36-37-12-9	240±40 (6,000±1,000)
	12	**15-16-23-9-11-12** 12-23-15-9-14-16-11-16-12-13-11-29-14-8-9-11-11-26-14-19-30-13-13-15-15-10-10-19-22-16-14-17-18-38-38-12-9	124±8 (3,100±200)
	13	**14-16-23-10-11-12** **(The Cohen Modal Haplotype)** 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	49±9 (1,200±200)
	14	**15-15-23-10-11-12** 12-23-15-10-14-17-11-15-12-13-11-29-	35±4 (875±125)

		15-8-9-11-11-24-15-21-32-12-13-17-17-10-11-19-23-16-14-18-19-36-36-12-9	
	15	**14-15-22-10-11-12** 12-22-14-10-14-15-11-15-12-14-11-31-15-9-9-11-11-26-15-20-29-13-13-16-16-10-10-19-22-15-12-15-17-34-36-12-9	29±2 (725±50)
R1b	16	**14-12-24-11-13-13** **(The Atlantic Modal Haplotype)** 13-24-14-11-11-14-12-12-12-13-13-29-17-9-10-11-11-25-15-19-30-15-15-16-17-11-11-19-23-16-16-18-18-37-37-12-12	200±5 (5,000±200)
	17	**14-12-24-11-13-13** **(The Atlantic Modal Haplotype)** 13-24-14-11-11-14-12-12-12-13-13-29-17-9-10-11-11-25-15-19-28-15-15-16-17-11-12-19-23-15-15-18-18-36-38-12-12	216±20 (5,400±500)
	18	**14-12-24-10-14-12** 12-24-14-10-11-14-12-12-12-13-14-29-17-9-10-11-11-25-15-19-29-15-15-16-17-11-11-19-23-16-16-18-17-37-38-12-12	196±4 (4,900±100)
	19	**14-12-24-10-14-12** 12-24-14-10-11-14-11-12-12-13-14-29-17-9-10-11-11-25-15-19-30-15-15-16-16-11-10-19-23-17-16-18-17-37-38-12-12	112±12 (2,800±300)
	20	**14-12-24-10-14-12**	43±10

		12-24-14-10-11-14-11-12-12-13-14-29-17-9-10-11-11-25-15-19-30-15-15-16-16-11-10-19-23-17-16-17-17-38-38-12-12	(1,100±250)
	21	**14-12-24-11-13-13** 13-24-14-11-11-14-12-12-11-13-13-28-17-9-10-11-11-26-15-19-30-15-15-15-18-11-11-19-23-16-15-18-19-36-37-12-11	38±8 (950±200)
	22	**14-12-24-10-14-12** 12-24-14-10-11-15-12-12-12-13-14-28-15-9-9-11-11-25-14-19-29-15-15-16-17-11-12-19-23-16-17-19-17-34-38-12-12	31±6 (775±150)
E1b1b (former E3b)	23	**13-12-24-10-11-13** 13-24-13-10-16-18-11-12-12-13-11-30-16-9-9-11-11-26-14-20-33-15-16-16-17-11-11-19-22-15-12-18-18-33-35-11-10-10-8-15-15-8-11-10-8-11-10-22-23-18-11-12-12-16-7-12-23-18-13-13-12-14-11-11-11-11	270±15 (6,800±400)
	24	**13-12-25-9-11-14** 14-25-13-9-17-18-11-12-12-13-11-30-16-9-9-11-11-21-14-20-28-14-14-15-17-10-10-19-19-15-13-12-16-33-34-13-10-10-8-15-15-7-11-11-8-10-10-20-22-19-11-12-12-14-7-11-25-18-13-13-12-14-11-12-10-11	41±10 (1,000±250)
	25	**14-12-24-10-11-13** 13-24-14-10-16-17-11-12-13-14-11-32-	39±4 (975±100)

		16-9-9-11-11-26-14-20-33-16-17-17-17-11-11-21-22-16-12-18-18-33-36-12-10--11-8-15-15-8-10-9-9-11-(9,11)-22-23-20-11-12-12-15-7-11-(24,25)-18-13-13-12-14-11-11-11-11	
	26	**13-12-24-10-11-13** 13-24-13-10-17-18-11-12-12-13-11-30-19-9-9-11-11-27-14-20-32-15-16-16-17-10-10-19-22-17-13-17-17-31-33-13-10-10-8-15-15-7-10-10-8-10-10-21-24-18-11-12-13-17-7-11-25-21-15-13-12-14-10-12-10-11	30±8 (750±200)
R1a1	27	**16-12-25-10-11-13** 13-25-16-10-11-14-12-12-10-13-11-30-14-9-10-11-11-24-14-20-30-12-12-15-15-11-11-19-23-14-16-19-20-35-38-14-11-11-8-17-17-8-12-10-8-11-10-12-22-22-15-10-12-12-14-8-14-23-21-12-12-11-13-10-11-12-13	47±12 (1,200±300)
	28	**16-12-25-10-11-13** 13-25-16-10-11-14-12-12-10-13-11-30-14-9-11-11-11-24-14-20-30-12-12-15-15-11-11-19-23-14-16-19-20-35-38-14-11-11-8-17-17-8-12-10-8-11-10-12-22-22-15-10-12-12-14-8-14-23-21-12-12-11-13-10-11-12-13	54±12 (1,350±300)
	29	**16-12-25-10-11-13** 13-25-16-10-11-14-12-12-10-13-11-30-14-9-10-11-11-24-14-20-30-12-12-15-15-11-11-19-23-14-16-19-21-35-38-14-11-11-8-17-17-8-12-10-8-11-10-12-22-22-15-10-12-12-14-8-14-23-21-12-12-	26±4 (650±100)

		11-13-10-11-12-13	
Q	30	**13-12-22-10-15-13** 13-22-13-10-14-16-12-12-12-13-15-29-17-9-9-11-11-25-14-19-29-14-15-15-16-10-9-19-19-15-14-17-16-33-38-12-11-11-8-15-19-8-11-10-8-12-11-22-22-16-11-12-12-17-8-12-24-16-14-12-11-13-10-12-13-13	27±5 (675±125)
R2	31	**14-12-23-10-10-14** 14-23-14-10-13-20-12-12-11-14-10-29-16-9-9-11-11-25(26)-14-19-33-12-12-15-16-10-10-20-20-15-15-17-19-34-34-12-11	26±2 (650±50)
	32	**14-12-23-10-10-14** 14-23-14-10-13-20-12-12-10-13-10-31-17-8-9-11-11-28-16-19-33-12-13-14-16-10-10-19-21-15-15-18-17-33-36-12-11-11-9(8)-15-17-8-10-10-8-11-10-12-20-22-14-11-12-12-16-8-12-26-22-13-12-11-14-11-11-13	7±2 (175±50)
G1	33	**14-12-23-10-12-14** 14-23-14-10-14-15-11-12-11-12-12-28	3,600±200 (875±175)
	34	**14-12-23-11-12-13** 13-23-14-11-14-15-11-12-11-12-12-28-17-9-9-11-11-24-16-20-28-11-13-14-16-9-10-20-20-16-13-17-19-36-36(35)-11-10-11-8-16-16-8-11-10-8-11-10-20-21-15-10-12-12-17-8-11-21-20-15-12-11-	39±7 (975±175)

		13-10-11-11-11	
G2	35	**15-12-22-10-11-14** 14-22-15-10-14-15-11-13-12-12-11-29-18-9-9-11-11-23-16-21-30-12-12-12-12-11-10-21-21-14-14-16-19-33-33-11-10	160±8 (4,000±200)
G2a3	36	**15-12-22-10-11-13** 13-22-15-10-12-16-11-12-11-12-11-29-16-9-9-11-11-24-16-21-29-12-13-14-14-11-11-20-21-16-14-18-19-34-35-11-10-11-8-16-16-8-11-10-8-11-10-20-21-15-10-12-12-17-8-11-21-20-15-12-11-13-10-11-11-11	25±8 (625±200)
G/G2c	37	**15-12-23-10-11-13** 13-23-15-10-13-16-11-12-11-14-11-32-18-8-9-11-11-24-16-21-28-13-14-15-15-10-11-20-20-17-16-15-16-34-35-12-10-11-8-16-16-8-11-10-8-12-10-21-22-14-10-11-12-14-9-13-19-21-18-13-11-13-10-11-11-12	23±2 (575±50)
I1	38	**14-14-22-10-11-13** 13-22-14-10-13-14-11-14-12-12-11-28-16(15)-7(8)-8(9)-8-11-23-16-20-29(30)-12-14-14-16-11-10(11)-19-21(22)-15-14-17(16)-19(20)-36(35)-38(36)-12(14)-10	163±78 (4,100±1,900)
I2	39	**16-13-24-11-11-13**	57±23

		13-24-16-11-14-15-11-13-13-13-11-31-18-8-10-11-11-25-15-20-31-13-14-15-16-10-10-21-21-14-12-19-17-34-35-11-10	(1,400±600)
I2b1	40	**16-13-22-10-13-13** 13-22-16-10-15-16-11-13-11-13-13-29-16-8-9-11-11-23-14-20-27-14-15-16-16-11-10-19-21-15-16-19-22-35-36-11-10	20±5 (500±125)
K, T	41	**13-12-23(22)-10-13-13** 13-23(22)-13-10-12(13)-16(15)-11-12-12-13(14)-13-30(28)(31)-16(17)-9-9-11-12-26-14(15)-19-34(33)-13-13-15-16-10-11-23-23-16-13(15)-18(17)-18(19)-34-36-12-9	338±39 (8,500±1,000)
	42	**14-12-23-10-13-13** 13-23-14-10-15-16-11-12-12-12-13-29-17-9-9-11-13-26-14-18-33-11-13-15-19-10-9-22-23-16-14-17-16-35(36)-39-11-9-11-8-16-17-8-11-10-8-12-9-12-20-20(22)-17-10-12-12-14-8-12-23-19-14-11-11-13-11-11-9-11	181±18 (4,500±500)
	43	**13-12-23-10-13-13** 13-23-13-10-13-15-11-12-12-13-13-28-17-9-9-11-12-26-14-19-34-13-13-15-15-10-10-23-24-16-15-15-18-33-36-11-9	101±8 (2,500±200)
	44	**14-12-23-10-13-13** 13-23-14-10-14-16-11-12-11-14-13-31-16-9-9-11-13-28-14-19-33-11-11-16-16-10-10-23-25-16-13-16-18-35-36-12-9-	91±24 (2,300±600)

		11-8-17-17-8-11-10-8-12-10-12-20-20-16-10-12-12-15-8-12-22-19-16-11-12-13-10-11-12-11	
	45	**13-12-22-10-13-13** 13-22-13-10-13-16-11-12-12-14-13-31-18-9-9-11-12-26-15-20-34-11-11-15-16-12-11-23-25-16-13-19-19-34-35-12-9	12±2 (300±50)

Haplotype number, haplotype length, Figure, branch, position, number of haplotypes in the Table, referred to the respective haplotype number:

J1, J*

1 – 6-marker haplotypes: Fig.1, upper right-hand side, 20 haplotypes
2 – 12-marker haplotypes: Fig. 5, lower right-hand side and bottom, 24 haplotypes;
37-marker haplotypes: Fig. 7, left and upper left-hand side, 11 haplotypes
3 – 6-marker haplotypes: Fig.1, left-hand side, 40 haplotypes
4 – 6-marker haplotypes, Fig.1, upper left-hand side, 29 haplotypes
5 – 6-marker haplotypes, Fig. 1, lower right-hand side, 12 haplotypes
6 – 12-marker haplotypes, Fig. 5, upper right-hand side, 9 haplotypes
37-marker haplotypes, Fig. 7, lower right-hand side, 4 haplotypes
7 – 12-marker haplotypes, Fig. 5, bottom, 15 haplotypes
25-marker haplotypes, Fig. 6, lower right-hand side, 10 haplotypes
37-marker haplotypes, Fig. 7, left-hand side, 7 haplotypes
8 – 37-marker haplotype, Fig. 7, bottom, 11 haplotypes
9 – 37-marker haplotype, Fig. 7, right-hand side, 16 haplotypes
10 – 12-marker haplotypes, Fig. 5, lower right-hand side, 5 haplotypes
25-marker haplotypes, Fig. 6, lower left-hand side, 2 haplotypes.
37-marker haplotypes, Fig. 7, upper left-hand side, 2 haplotypes.

J2

11 – 10-marker haplotypes, Fig. 12, left-hand side, 35 haplotypes
12-marker haplotypes, Fig. 13, all 75 haplotypes except 22 CMHs
37-marker haplotypes, Fig. 14, all 25 haplotypes, except 10 CMHs
12 – 37-marker haplotypes, Fig. 14, lower left-hand side,6 haplotypes
13 – 10-marker haplotypes, Fig.12, top, 16 haplotypes
12-marker haplotypes, Fig.13, lower right-hand side, 18 haplotypes
37-marker haplotypes, Fig.14, upper right-hand side, 10 haplotypes
14 – 37-marker haplotypes, Fig. 14, lower right-hand side, 4 haplotypes
15 – 6-marker haplotypes, Fig.10, right-hand side, 24 haplotypes;
12-marker haplotypes, Fig.13, upper right-hand side, 10 haplotypes
37-marker haplotypes, Fig.14, upper left-hand side, 3 haplotypes

R1b

16 – 37 haplotypes, Fig. 22, the whole tree, 56 haplotypes
17 – 37 haplotypes, Fig. 22, right-hand side, 20 haplotypes

18 – 37 haplotypes, Fig. 22, the whole tree minus the upper left-hand side recent branch, 42 haplotypes
19 – 37 haplotypes, Fig. 22, bottom, 22 haplotypes
20 - 12-marker haplotypes, Fig.21, middle right-hand side, 16 identical haplotypes
37-marker haplotypes, Fig.22, lower left-hand side, 14 haplotypes
21 – 12-marker haplotypes, Fig.21, lower left-hand side, 12 identical haplotypes
37-marker haplotypes, Fig. 22, upper left-hand side, 14 haplotypes
22 – 12-marker haplotypes, Fig.21, lower right-hand side, 8 identical haplotypes
37-marker haplotypes, Fig. 22, lower right-hand side, 7 haplotypes

E1b1b (former E3b)

23 – 37-marker haplotypes, Fig. 30, lower half, 29 haplotypes; ("YSearch set); Fig. 31, left, bottom, and lower right-hand side, 48 haplotypes ("club set"); Fig. 31, upper left corner, 19 haplotypes ('club set")
66-marker haplotypes, Fig. 32, bottom, 9 haplotypes, ("YSearch set"); Fig. 33, left-hand side, 12
haplotypes ("club set")
24 – 37-marker haplotypes, Fig. 30, upper right corner, 15 haplotypes ("YSearch set); Fig. 31, upper right-hand side, 26 haplotypes ("club set")
66-marker haplotypes, Fig. 32, upper left corner, 5 haplotypes ("YSearch set"); Fig. 33, upper right- hand side, 12 haplotypes ("club set")
25 – 37-marker haplotypes, Fig. 30, lower right corner, 10 haplotypes ("YSearch set); Fig. 31, lower left-hand side, 10 haplotypes ("club set")
66-marker haplotypes, a deduced sequence in 38^{th}-66^{th} markers
26 – 37-marker haplotypes, Fig. 30, upper left-hand side, 11 haplotypes ("YSearch set"); Fig. 31, lower right-hand side, 11 haplotypes ("club set")
66-marker haplotypes, Fig. 32, upper right-hand side, 6 haplotypes (YSearch set"); Fig. 10, lower right-hand side, 5 haplotypes ("club set")

R1a1

27 – 6-marker haplotypes, Figs. 34-36; 42, 44 and 100 haplotypes ("scientific", YSearch, and "club" sets, respectively)
12-marker haplotypes, Figs. 37 and 38, 44 and 100 haplotypes (YSearch and "club" sets, respectively)
25-marker haplotypes, Figs. 39 and 40, 30 and 52 haplotypes (YSearch and "club" sets, respectively)
37-marker haplotypes, Figs. 41 and 42, 27 and 38 haplotypes (YSearch and "club" sets, respectively)
67-marker haplotypes, Figs. 43 and 44, 13 and 6 haplotypes ("club" and YSearch sets, respectively)
28 – 37-marker haplotypes, Fig. 41, left-hand side, 13 haplotypes (YSearch set)
67-marker haplotypes, Fig. 43, right-hand side, 6 haplotypes ("club" set)
29 – 37-marker haplotypes, Fig. 41, right-hand side, 14 haplotypes (YSearch set)
67-marker haplotypes, Fig. 43, left-hand side, 7 haplotypes ("club" set)

Q

30 – 6-marker haplotypes, Fig. 45; 23 haplotypes ("scientific" set)
12-marker haplotypes, Figs. 46 and 47; 23 and 58 haplotypes (YSearch and "club" sets, respectively)
37-marker haplotypes, Fig. 48; 29 haplotypes ("club set")

66-marker haplotypes, Fig. 49; 14 haplotypes ("club set")

R2

31 – 6-marker haplotypes, Fig. 50, haplotypes around the "trunk", 11 haplotypes
12-marker haplotypes, Fig. 51, haplotypes around the "trunk", 9 haplotypes
37-marker haplotypes, Fig. 52, the right-hand side, 4 haplotypes
32 – 12-marker haplotypes, Fig. 51, bottom, 3 haplotypes
37-marker haplotypes, Fig. 52, the left-hand side, 3 haplotypes
66-marker haplotypes, 3-mutation difference in two haplotypes (001 and 005)

G1

33 – 12-marker haplotype, Fig. 56, lower right-hand side area, 7 haplotypes
34 – 12-marker haplotypes, Fig. 56, lower right-hand side, the flat branch, 6 haplotypes
37-marker haplotypes, Fig. 57, lower right-hand side, 4 haplotypes
66-marker haplotypes, Fig. 58, bottom, 3 haplotypes

G2

35 – 12-marker haplotypes, Fig. 56, upper far right-hand side, 7 haplotypes
37-marker haplotypes, Fig. 57, only one haplotype left (062)

G2a3

36 – 12-marker haplotypes, Fig. 56, upper right-hand side, flat branch, 7 haplotypes
37-marker haplotypes, Fig. 57, lower left-hand side, 5 haplotypes
66-marker haplotypes, Fig. 58, only one haplotype left (056)

G/G2c

37 – 12-marker haplotypes, Fig. 56; 40 haplotypes around the "trunk"
37-marker haplotypes, Fig. 57; 21 haplotypes around the "trunk"
66-marker haplotypes, Fig. 58; 12 haplotypes around the "trunk"

I1

38 – 6-marker haplotypes, Fig. 59, lower right-hand side, 5 haplotypes ("scientific" set); Fig. 61, upper right-hand side, 4 haplotypes (YSearch set)
10-marker haplotypes, Fig. 62, lower left-hand side, 6 haplotypes ("scientific" set)
12-marker haplotypes, Fig. 64, upper right-hand side, 7 haplotypes (YSearch set)
37-marker haplotypes, Fig. 65, left-hand side, 4 haplotypes (YSearch set)

I2

39 – 6-marker haplotypes, Fig. 59, lower left-hand side, 10 haplotypes ("scientific" set); Fig. 61, left-hand side, 7 haplotypes (YSearch set)
10-marker haplotypes, Fig. 62, right-hand side, 7 haplotypes ("scientific" set)
12-marker haplotypes, Fig. 64, upper left-hand side, 6 haplotypes (YSearch set)
37-marker haplotypes, Fig. 65, lower right-hand side, 3 haplotypes (YSearch set)

I2b1

40 – 6-marker haplotypes, Fig. 59, upper right-hand side, 4 haplotypes ("scientific" set); Fig. 61, lower right-hand side, 9 haplotypes (YSearch set)
10-marker haplotypes, Fig. 62, upper left-hand side, 4 haplotypes ("scientific" set)
12-marker haplotypes, Fig. 64, bottom, 7 haplotypes (YSearch set)

37-marker haplotypes, Fig. 65, upper right-hand side, 5 haplotypes (YSearch set)

K, T

41 – 37-marker haplotypes, Fig. 69, lower right-hand side, 4 haplotypes (YSearch set)
42 – 12-marker haplotypes, Fig. 68, upper left-hand side, 6 haplotypes (YSearch set)
37-marker haplotypes, Fig. 69, upper right-hand side, 3 haplotypes (YSearch set)
43 – 12-marker haplotypes, Fig. 68, lower left-hand side, 5 haplotypes (YSearch set)
37-marker haplotypes, Fig. 69, only one haplotype left (020)
44 – 12-marker haplotypes, Fig. 68, right-hand side, 11 haplotypes (YSearch set)
37-marker haplotypes, Fig. 69, left-hand side, 3 haplotypes (YSearch set)
45 – 12-marker haplotypes, Fig. 68, top of the tree, 4 haplotypes (YSearch set)

References

Behar, D.M., Thomas, M.G., Skorecki, K., Hammer, M.F., Bulygina, E., Rosengarten, D., Jones, A.L., Held, K., Moses, V., Goldstein, D., Bradman, N and Weale, M.E. Multiple origins of Ashkenazi Levites: Y chromosome evidence for both Near Eastern and European ancestries. Am. J. Hum. Genet. 73, 768-779 (2003).

Behar, D.M., Garrigan, D., Kaplan, M.E., Mobasher, Z., Rosengarten, D., Karafet, T.M., Quintana-Murci, L., Oster, H., Skorecki, K. and Hammer, M.F. Contrasting patterns of Y chromosome variation in Ashkenazi Jewish and host non-Jewish European populations. Hum. Genet. 114, 354-365 (2004).

Bonne-Tamir, B., Korostishevsky, M., Redd, A.J., Pel-Or, Y., Kaplan, M.E. and Hammer, M.F. Maternal and paternal lineages of the Samaritan isolate: mutation rates and time to most recent common male ancestor. Annals of Human Genetics 67, 153-164 (2003)

Eisen, Y. Miraculous Journey. Feldheim Publishers, Jerusalem, 2004, 712 pp. ISBN 1-5687-1-323-1)

Jewish Ashkenazi-Levite DNA Project, http://www.familytreedna.com/public/Ashkenazi-Levite/index.aspx?fixed_columns

Jewish E3b Project, http://www.familytreedna.com/public/JewishE3bProject/

Jewish Q Project, http://www.familytreedna.com/public/Jewish_Q

Levy-Coffman, E. A mosaic of people: the Jewish story and a reassessment of the DNA evidence. Journal of Genetic Genealogy 1, 12-33 (2005)

Nasidze, I., Sarkisian, T., Kerimov, A. and Stoneking, M. Testing hypotheses of language replacement in the Caucasus: evidence from the Y-chromosome. Hum. Genet. 112, 255-261 (2003)

Nebel, A., Filon, D., Weiss, D.A., Weale, M., Faerman, M., Oppenheim, A. and Thomas, M. High-resolution Y chromosome haplotypes of Israeli and Palestinian Arabs reveal geographic substructure and substantial overlap with haplotypes of Jews. Hum. Genet. 107, 630-641 (2000).

Nebel, A., Filon, D., Brinkmann, B., Majumder, P.P., Faerman, M. and Oppenheim, A. The Y chromosome pool of Jews as part of the genetic landscape of the Middle East. Am. J. Hum. Genet. 69, 1095-1112 (2001a).

Nebel, A., Filon, D., Hohoff, C., Faerman, M., Brinkmann, B. and Oppenheim, A. Haplogroup-specific deviation from the stepwise mutation model at the microsatellite loci DYS388 and DYS392. European J. Hum. Genetics 9, 22-26 (2001b).

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., Rani, M.V.U., Thakur, C.M., Cavalli-Sforza, L.L., Majumder, P.P., and Underhill, P.A. 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 (2006)

Shen, P., Lavi, T., Kivisild, T., Chou, V., Sengun, D., Gefel, D., Shpirer, I., Woolf, E., Hillel, J., Feldman, M.W. and Oefner, P.J. Reconstruction of patrilimeages and matrilineages of Samaritans and other Israeli populations from Y-chromosome and mitochondrial DNA sequence variation**.** Human Mutation 24, 248-260 (2004).

Skorecki, K., Selig, S., Blazer, S., Bradman, R., Bradman, N., Warburton, P.J., Ismajlowicz, M and Hammer, M.F. Y chromosomes of Jewish Priests. Nature 385, 32 (1997)

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

Thomas, M.G., Parfitt, T., Weiss, D.A., Skorecki, K., Wilson, J.F., le Roux, M., Bradman, N. and Goldstein, D.B. 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 (2000).

CONTENTS

June 2008
Volume 1
No. 1
Special Edition