Ultima ratio

Proceedings of the Russian Academy of DNA Genealogy

2011 June
Volume 4, No. 6

The Russian Academy of DNA Genealogy

ISSN 1942-7484

Proceedings of the Russian Academy of DNA Genealogy.
An ongoing scientific edition in the area of DNA genealogy, history, linguistics, with implications in family studies. Published by Lulu Inc., 2011.

Publisher
The Russian Academy of DNA Genealogy
Anatole A. Klyosov
Boston, USA

Cover by
Pavel Shvarev

Second Edition

CONTENTS

Haplotypes of R1b1a2-P312 and related subclades: origin and "ages" of most recent common ancestors

Anatole A. Klyosov
http://aklyosov.home.comcast.net

SUMMARY

A detailed analysis of 2299 67 marker haplotype dataset provided by Michael Walsh, administrator of R1b1a2-L21 site http://www.familytreedna.com/public/R-L21/default.aspx?section=yresults has allowed to identify base (ancestral) 67 marker haplotypes and timespans to their common ancestors as follows:

R1b1a2	L265, M269, M520, S3, S10, S13, S17	**~ 7000 ybp[1]**
R1b1a2a	L23/S141, L49**.1**	**~ 6200 ybp[1]**
R1b1a2a1a	L51/M412/S167	**5300±700 ybp[1]**
R1b1a2a1a1	L11, P310, P311	**~ 4800 ybp (in Europe)[2]**
R1b1a2a1a1a	M405/S21/U106	**4175±430 ybp**
null mutation U106		**3325±450 ybp**
R1b1a2a1a1b	P312/S116	**~ 4800 ybp**
null mutation P312		**3575±400 ybp**
R1b1a2a1a1b1	***M65***	**~ 1800 ybp**
R1b1a2a1a1b2	M153	**~ 3640 ybp**
R1b1a2a1a1b3	S28/U152	**4125±450 ybp[3]**
null mutation U152		**3525±460 ybp**
R1b1a2a1a1b3c	L2/S139	**4025±410 ybp**
R1b1a2a1a1b3c1	L20	**3650±400 ybp**
R1b1a2a1a1b3d	L4/S178	**1275±290 ybp**
R1b1a2a1a1b4	L21/M529/S145	**3750±380 ybp**
null mutation L21		**3025±460 ybp**
null mutation L21		**1500±325 ybp**
Unidentified clade		**1650±175 ybp**
R1b1a2a1a1b4b	M222	**1450±160 ybp**
R1b1a2a1a1b4e	L144	**~ 4000 ybp**
R1b1a2a1a1b4f	**L159.2**	**1775±200 ybp**
R1b1a2a1a1b4g	L193	**1275±170 ybp**
R1b1a2a1a1b4h	L226	**1500±170 ybp**
R1b1a2a1a1b4i	**P314.2**	**2225±300 ybp**
R1b1a2a1a1b5	**L176.2/S179.2**	**3675±560 ybp**
R1b1a2a1a1b5a	SRY2627	**3150±320 ybp**
null mutation SRY		**~ 2800 ybp**
R1b1a2a1a1b5b	L165/S68	**~ 3000 ybp**
R1b1a2a1a1b6	**L238/S182**	**~ 1000 ybp**

[1] In Asia
[2] 4575±580 ybp (a different dataset)
[3] 3800±380 (a different dataset)

A geographical/regional distribution of R1b1a2-P312 and it's the largest subclade P312-L21 was analyzed. The Bell Beaker movements from the Iberian Peninsula up North, North-East and to the Isles from 4800 years before present (ybp) and upward provides the best fit with the data of DNA genealogy.

Introduction

This article continues a series of studies on DNA genealogy of bearers of R1b haplogroup, published in the Proceedings of the Russian Academy of DNA Genealogy (ISSN 1942-7484) and other editions, a complete list is given in http://aklyosov.home.comcast.net). The list below is a short version in order from the latest publications down to some earlier ones):

Klyosov, A.A. (2011) Haplotypes of R1b1a2-P312 and related subclades: origin and "ages" of most recent common ancestors. Proc. Russian Academy of DNA Genealogy, 4, № 6, 1127-1195

Klyosov, A.A. (2011) DNA genealogy of the major haplogroups of Y chromosome. Proc. Russian Academy of DNA Genealogy, 4, № 5, 988-1014; this issue, pp. 1257-1282.

Klyosov, A.A. (2011) Sumers as the ancient bearers of R1b1b2 haplogroup? Proc. Russian Academy of DNA Genealogy, 4, No.4, 762-763.

Klyosov, A.A. (2011) Haplogroup R1a on Comoros Islands and ancient migrations of R1a1 and R1b haplogroups. Proc. Russian Academy of DNA Genealogy, 4, No.1, 33-39.

Klyosov, A.A. (2011) History of R1a and R1b haplogroups in Iran and Lebanon. Proc. Russian Academy of DNA Genealogy, 4, No.1, 20-32.

Klyosov, A.A. (2010) On time and place of origin of R1b haplogroup (reading Wikipedia with disbelief and astonishment). Proc. Russian Academy of DNA Genealogy, 3, No.12, 2084-2109.

Klyosov, A.A. (2010) On "age" of haplogroup R1b1b2 in the Caucasus. Proc. Russian Academy of DNA Genealogy, 3, No.10, 1801-1802.

Klyosov, A.A. (2010) Hypothesis on origin of the Turks, Sino-Caucasian (proto-Turkic?) languages, and haplogroup R1b. Proc. Russian Academy of DNA Genealogy, 3, No.10, 1757-1764.

Klyosov, A.A. (2010) Haplogroup R1b1 and its subclades in Asia. Proc. Russian Academy of DNA Genealogy, 3, No.10, 1676-1695.

Klyosov, A.A. (2010) Migration route of R1b haplogroup to Europe and its populating in Europe. A critical analysis of Myres et al paper "A Major Y-chromosome haplogroup R1b…" (Eur. J. Hum. Genetics, 26 August 2010). Proc. Russian Academy of DNA Genealogy, 3, No.10, 1652-1675.

Klyosov, A.A. (2010) An "age" of R1b1b2-M269 subclade and its downstream subclades L23, L51, L11. Proc. Russian Academy of DNA Genealogy, 3, No. 8, 1310-1315.

Klyosov, A.A. (2010) Subclade R1b1b2-L226 ("Irish III") – the latest update. Proc. Russian Academy of DNA Genealogy, 3, No. 7, 1211-1213.

Klyosov, A.A. (2010) Haplotypes of P312* subclade: a history in letters. Proc. Russian Academy of DNA Genealogy, 3, No. 7, 1165-1183.

Klyosov, A.A. (2010b) Irish haplotypes and haplogroups. Proc. Russian Academy of DNA Genealogy, 3, No. 6, 1029-1053.

Klyosov, A.A. (2010c) Subclade R1b1b2-M153 on the Iberian Peninsula and among the Basques. Proc. Russian Academy of DNA Genealogy, 3, No. 6, 976-982.

Klyosov, A.A. (2010) Migration route of R1b1b2 to Europe (2). A critical consideration of the L. Morelli et al (PLoS ONE) paper. Proc. Russian Academy of DNA Genealogy, 3, No. 6, 903-915.

Klyosov, A.A. (2010) Migration route of R1b1b2 to Europe (1). A critical consideration of the P. Balaresque et al (PLoS Biology) paper. Proc. Russian Academy of DNA Genealogy, 3, No. 6, 896-902.

Klyosov, A.A. (2010) Haplogroup R1b. Part 2. Proc. Russian Academy of DNA Genealogy, 3, № 3, 406-475.

Klyosov, A.A. (2010) Consideration of 37 Irish R1a1 haplotypes and 1036 R1b1 haplotypes in the same series of 67 marker haplotypes: a separation of the R1a1 subclade of the Russian Plain and R-M222, and timespans to their common ancestors. Proc. Russian Academy of DNA Genealogy, 3, No. 3, 398-405.

Klyosov, A.A. (2010) A consideration of a family of (L21+M222-) haplotypes with the "11-13 signature". Proc. Russian Academy of DNA Genealogy, 3, No. 3, 391-397.

Klyosov, A.A. (2010) Haplotypes of the R1b haplogroup in Central Africa. Proc. Russian Academy of DNA Genealogy, 3, No. 3, 369-378.

Klyosov, A.A. (2010) Haplogroup R1b. Part 1. Proc. Russian Academy of DNA Genealogy, 3, No.2, 249-299.

Klyosov, A.A. (2010) Mapping of Europe by R1b1b2-L21 migration times. Proc. Russian Academy of DNA Genealogy, 3, No.2, 191-201

Klyosov, A.A. (2010) The principal mystery in relationships between IndoEuropean and Turkic language families and an attempt to solve it with the help of DNA genealogy: considerations of a non-linguists. Proc. Russian Academy of DNA Genealogy, 3, № 1, 2-57.

Klyosov, A.A. (2009) Iberian haplotypes and analysis of populations of the Basques, Sephards, and other populations of Spain and Portugal. Proc. Russian Academy of DNA Genealogy, 2, № 3, 390-421.

Klyosov, A.A. (2009) DNA Genealogy, mutation rates, and some historical evidences written in Y-chromosome. II. Walking the map. J. Genetic Genealogy. 5, 217 – 256.

Klyosov, A.A. (2008) Mysteries of the "Western European" haplogroup R1b. Proc. Russian Academy of DNA Genealogy, 1, № 4, 568-630.

HISTORY OF R1b HAPLOGROUP from Central Asia to Europe

Appearance of R1b in Central Asia (Altai, Xinjiang, North-Western China), its ancient migrations to Europe; Bashkirs, Caucasian people, Anatolia, Sumer, North Africa, Iberia, the Basques, continental Europe, British Isles), Bell Beakers as R1b1a2-P312 and –U106 and their upstream (L51, L11/L151/P310/P311) and downstream subclades in Europe, and non-Bell Beakers R1b1a2-M269, -L23, and their downstream subclades in Europe

It was shown in a series of more than 30 papers (Klyosov, 2008-2011, see above and the list of references at the end of the paper) that haplogroup R1b1, as well

in Central Asia (Southern Siberia or adjacent regions, such as Altai, Xinjiang, North-Western China). R1a arose around 21,000 years before present (ybp), R1b1 arose next, about 16,000 ybp. Since the very slow (with respect to mutations) 22 marker European R1a1-M17 base (aka deduced ancestral) haplotype [DYS 426, 388, 392, 455, 454, 438, 531, 578, 395S1a, 395S1b, 590, 641, 472, 425, 594, 436, 490, 450, 617, 568, 640, 492]

12 12 11 – 11 11 – 11 – 11 8 17 17 8 10 8 12 10 12 12 8 12 11 11 12 (R1a)

and R1b1a2-M269 base 22 marker haplotype

12 12 **13** – 11 11 – **12** – 11 **9 15 16** 8 10 8 12 10 12 12 8 12 11 11 12 (R1b1a2)

differ by 7 mutations, as marked above (Klyosov, 2011a), which gives 7/0.006 = 1167 generations without correction for back mutations, or 1380 generations with the correction (the correction factor in this particular case equals 1.187, see pp. 1280-1281 in this issue), that is 34,600 years between their common ancestors, THEIR common ancestor lived approximately (34600+5000+7000)/2 = 23,300 years before present. This is a common ancestor of R1 haplogroup. Here 0.0060 is the mutation rate constant for the 22 marker haplotype (in mutation/haplotype/generation of 25 years) [Klyosov, 2011b, and the paper in this issue, pp. 1239-1256].

Indeed, if the R1 common ancestor lived 23,300 ybp, then the R1a1-M17 base haplotype on the Russian Plain appeared (2,300 + 16,000) = 18,300 years after the R1 appearance, and the R1b1b2-M269 arose (7,300 + 9,000) = 16,300 years after the R1 appearance, hence, the difference between the two base haplotypes above equals to (18,300 + 16,300) = 34,600 years, exactly as the two 22 marker base haplotypes show as the seven mutation difference between them.

After haplogroup R1 arose about 23,000 years ago, and its downstream R1b arose in Central Asia 16,000 ago, the latter had migrated across North Kazakhstan (including the later Botai archaeological culture, 6500-5500 ybp), through the territory of the present-day Bashkirs (13-11-8 thousand ybp), established Seroglazovo archaeological culture (13-11 thousand ybp) and then Middle Volga Basin archaeological cultures (Srednevolga culture 8-7 thousand ybp; Samara culture 7 thousand ybp; Srednestog culture 7500-5500 ybp; Khvalyn culture 7-6 thousand ybp, Kama culture 7-5 thousand ybp), that is the cultures which were (tentatively) named "Pre-Kurgan cultures"; R1b1a2-M269 subclade arose around 7,000 ybp, and then R1b1a2-L23, around 6,500 ybp. Both of them came to the Caucasus region around 6,500-6,000 ybp, and possibly earlier. Those R1b1b2 spoke non-IndoEuropean language(s), which can be vaguely traced now under various names. Some call them Proto-Turkic language, some Sino-Caucasian language, some PaleoEurasian language, some call them the Erbin language

(because of R1b haplogroup of their bearers), some call them Western Caucasian or North-Western Caucasian languages, some call them the Basque language, and find some similarities between the Caucasian and Basque languages, etc.

In fact, those similarities do exist, albeit in a rather weak form because thousands of years passed since then. Still now most of R1b1a2 haplotypes in the Caucasus (in Armenia, Dagestan, Georgia) belong to the ancient L23 subclade (with a common ancestor in the Caucasus of around 6,000 ybp), and have a characteristic DYS393=12 allele, unlike DYS393=13 in most of European R1b1b2 haplotypes.

From the Caucasus, R1b1a2-L23 and R1b1a2-M269 bearers went South over the mountains, to Anatolia (a common ancestor of 6,000 ybp), and then split into three major routes. One went further South, to Tigris and Euphrates Rivers, and became the Sumers. Many present-day Assyrians, descendants of ancient Sumers, still have their R1b1a2 haplotypes. Another went westward, across Asia Minor, and came to Europe, to the Balkans and Mediterranean Sea region around 4500 ybp. The third group went across Northern Africa and Egypt (and, incidentally, might have left some R1b1b2 Pharaohs there) to the Atlantic and went across Gibraltar to the Iberian Peninsula around 4800 ybp. They became the Bell Beakers, and moved up North into the continental Europe. The Bell Beaker culture in Europe had lasted between about 4400 and 3800 ybp.

R1b1a2 bearers came to the Pyrenees, apparently, as mainly L51/M412 and/or L11/L151/P310 subclades, and soon split there into U106 and P312 downstream subclades, which went to the continental Europe as said Bell Beakers with downstream subclades such as U198, U152, L2, L20, L4, L21, M222, L226, SRY2627, etc. They left the Basques behind, which still maintain the ancient Erbin (R1b) language, similar in kind with some Caucasian languages, and with the Sumer language, and having some elements of Proto-Turkic languages back to Asia and further to Siberia.

Until the beginning of the 1st millennium BC the R1b1a2 language in Europe was predominantly (or only) non-IndoEuropean (non-IE). There is not a single solid evidence of otherwise, that R1b1a2 in Europe spoke a IE language before the 1st millennium BC. Celtic language, as well as "Proto-Celtic" language, for instance, is placed by linguists not earlier than 800-900 BC, and all more earlier placements are typically groundless. At the same time, we know that it were R1a1 bearers who brought their IE language to India around 3500 ybp, and it is being dated in Europe to at least 6,000 ybp. There were no R1b1a2 in Europe those times. It mush have been R1a1 bearers who spoke proto-IE language(s) in Europe 6,000 ybp and some earlier.

Roots of the Bell Beakers can be traced back by comparing 67 marker base haplotypes of their present-day descendants of "parallel" subclades U106 and P312:

13 **23** 14 11 11 14 12 12 12 13 13 29 – 17 9 10 11 11 25 15 19 29 15 15 17 17 –
11 11 19 23 **16** 15 **17** 17 **37 39** 12 12 – 11 9 15 16 8 10 10 8 10 10 12 23 23 16 10 12 12
15 8 12 22 20 13 12 11 13 11 11 **13** 12 **(U106)**

13 24 14 11 11 14 12 12 12 13 13 29 – 17 9 10 11 11 25 15 19 29 15 15 17 17 –
11 11 19 23 15 15 18 17 36 38 12 12 – 11 9 15 16 8 10 10 8 10 10 12 23 23 16 10 12 12
15 8 12 22 20 13 12 11 13 11 11 12 12 **(P312)**

There are six mutational differences between them (marked in the base U106 base haplotype above), which corresponds to 6/0.12 = 50 → 53 "conditional" generations (25 year per generation), that is 1,325 year time span between their common ancestors. The sign → means a correction for back mutations (see below). Since a common ancestor of U106 and P312 lived 4175±430 ybp and 4100±415 ybp respectively (see below, also 3950±400 ybp, Klyosov, 2010a, which is practically the same figure within the margin of error), THEIR common ancestor lived (1325+4175+4100)/2 = 4800 years ago. This fits to the Bell Beakers timing quite well.

Project description

In this article we consider subclade R1b1a2-P312 and its downstream subclades based on the list of 2299 67 marker P312 haplotypes, provided by Michael Walsh, administrator of R1b1a2-L21 site http://www.familytreedna.com/public/R-L21/default.aspx?section=yresults

The haplogroup/subclade tree is given above (see Summary) for R1b1a2-P312 upstream and downstream subclades which are considered in this paper. The nomenclature described at
http://www.isogg.org/tree/ISOGG_HapgrpR.html, May 2011.

Haplotype analysis. Methodology

The essence of the methodology employed in this study is as follows:
(a) to build a haplotype tree and to resolve lineages/branches,
(b) to calculate each branch separately and to identify a timespan to the most recent common ancestor for each one,
(c) to identify a timespan to a common ancestor (TSCA), if any, for all the branches altogether,
(d) to make calculations with a correction for back mutations, if a timespan is

generations (25 years each, that is 575 years in 23 "generations"),
(e) to compare calculations for 25-, 37, and 67-marker haplotypes, to make sure that the calculations are compatible for all the three haplotype formats,
(f) to verify the data obtained with the logarithmic method (when possible), which does not need mutation counting,
(g) to use calibrated mutation rates and calibrated generation lengths,
(i) to provide margins of error to all results, for their realistic evaluations.

All the above items are illustrated in the text below, those some of them are omitted in order not to make to text too detailed. A few, which need more detailed explanations, are described in this section.

Analysis of mutations and their rates. Principal methodology was described in (Klyosov, 2009a,b). Haplotype trees were constructed using PHYLIP, the Phylogeny Inference Package program, as was explained in detail in (Klyosov, 2009a and references therein). A "comb" around the wheel (a "trunk"), in haplotype trees identifies "base" haplotypes, identical to each other and carrying no mutations compared to their ancestral haplotypes. They typically are observed in 12- and 25 marker haplotype trees, but not in 67 marker trees (where all haplotypes are typically mutated compared with their ancestral haplotypes). The farther the haplotypes lay from the trunk (the wheel), the more mutations they carry compared to the base haplotype, hence, the older the respective branch.

The "base" haplotypes are the ancestral haplotypes in an ideal case. However, since those haplotypes often are deduced ones, it would be inaccurate to call them "ancestral" when that might not necessarily be. Hence, we call them "base" haplotypes.

Timespans to the most recent common ancestor were calculated using both "logarithmic" and "linear" approaches. The logarithmic method is based on the assumption (which largely or practically always holds true) that a transition of the base haplotypes into mutated ones is described by the first-order kinetics:

$N = Ae^{kt,}$

that is

$\ln(N/A) = kt$

where:

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

k = an average mutation rate,
t = a number of generations to a common ancestor.

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

Mutation-counting methods are all based on accumulation of mutations in haplotypes over time. They include the "linear", "quadratic" (ASD) and "permutational" methods (Klyosov, 2009a). In this paper the "linear" method is largely employed, in which a total number of mutations in a set of haplotypes is counted, an average number of mutations per marker is calculated, a correction for back mutations is introduced, either numerically, or using a handy table (Klyosov, 2009a), and a time span to a common ancestor is calculated, either using the table, or applying the respective mutation rates.

It is important that the TSCA (time span to a common ancestor) values obtained by using the linear method and the logarithmic method should be equal to each other within margin of error. It means that the accumulation of mutations in the base (ancestral) haplotype in the course of the TSCA has followed the first order kinetics. In practical terms it means that the dataset indeed has a common ancestor, and it represents a lineage, not a mix of different lineages descending from various common ancestors. If the apparent TSCA obtained by the linear and the logarithmic methods significantly differ from each other (beyond a margin of error), those TSCAs are "phantom" ones, they are incorrect.

Average mutation rates employed in this paper, are published in (Klyosov, 2009a), with some corrections made later for 67 marker haplotypes (Klyosov and Rozhanskii, 2010):

For 12 marker haplotypes – 0.022 mutations per haplotype, 0.00183 mutations per marker,
for 25 marker haplotypes – 0.046 mutations per haplotype, 0.00183 mutations per marker,
for 37 marker haplotypes – 0.090 mutations per haplotype, 0.00243 mutations per marker,
for 67 marker haplotypes – 0.12 mutations per haplotype, 0.00179 mutations per marker.

For 49 marker haplotypes (see below) – 0.080 mutations per haplotype, 0.00163 mutations per marker. The lowest mutation rate constant in this case is the result of removal from the marker panel of 18 markers sensitive to recLOH mutations, null mutations, and other complications.

These mutation rates were calibrated employing 25 years per generation. This is a fixed mathematical figure, not an actual length of generations, which is a "floating" value, depending on many factors, including cultural, demographical, economical, and largely varied between ancient times and today. Time spans to a common ancestor were calculated typically employing 67 marker haplotypes, and/or sometimes the first 25 and/or 37 markers haplotypes, and, when possible, the data were compared to each other. Margins of error for time spans to common ancestors are calculated as described in (Klyosov, 2009a).

There are many confusions in the literature regarding mutation rate constants. Some authors claim that 25 years per generation is incorrect, and suggest different time spans per generations, typically between 20 and 35 years per generation, citing some arbitrarily chosen examples from recent genealogies. However, it is impossible to know if those generation lengths stay the same down through millennia. On the contrary, it would be hard to imagine that. Generation length is a "floating" factor, and depends on times, cultures, historical situations, etc. Therefore in our approach we do not use any arbitrarily chosen generation length. Since DNA genealogy obtains only a product <kt> (as a ratio of a number of mutations divided by a number of haplotypes or markers in the dataset), where k in the mutation rate and t is a number of generations, we can (and should) set a generation length based on an actual number of years for the calibration example. For instance, an earlier dataset the R1a1 Donald Clan ("Red Subgroup") contained 44 mutations in 68 of 12 marker haplotypes, and 69 mutations in 60 of 25 marker haplotypes. In a recent update (2010) there were 64 mutations in 125 of 12 marker haplotypes and 166 mutations in 124 of 25 marker haplotypes. Since we know that the common ancestor of the group, John Lord of the Isles, lived (in the context of this study) 650 years ago (he died in 1386, that is 624 years ago), we can make it 26 generations of 25 years each. In other words, we calibrate the mutation rate constant setting it at the mathematical value of 25 years per generation. We could have set it for 13 generations of 50 years each, it would not change a thing. There will be 650 years anyway. Just a mutation rate constant would be twice as high.

Let us see how we do it. For the four dataset for the "Red Subgroup", setting 26 generations to the common ancestor, we obtain:

12 marker haplotypes (an earlier and a recent count of mutations and haplotypes):

44/68/26 = 0.025±0.005 mutations per haplotype per generation (of 25 years)

64/125/26 = 0.020±0.003 mut/hapl/gen

One can see that these mutation rate constants are the same within the margins of error. The subsequent detailed studies with many datasets set the mutation rate constant as 0.022±0.001 mut/hapl/gen. In fact, the Chandler data (Chandler, 2006) also gave the mutation rate constant 0.022 mut/hpl/generation.

25 marker haplotypes:

69/60/26 = 0.044±0.007 mutations per haplotype per generation (of 25 years)

166/124/26 = 0.051±0.006 mut/hapl/gen

One can see that these mutation rate constants are the same within the margins of error. The subsequent detailed studies with many datasets set the mutation rate constant as 0.046±0.002 mut/hapl/gen. The Chandler data (2006) when combined all 25 markers is much higher; it does not fit its own 12 marker mutation rate constant when applied to real examples. Therefore, we do not use it.

One has to be careful, though, with those kinds of calibration. The latest (May, 2011) inspection of the Donald R1a1 haplotype series (the Red group) revealed that the Association sharply increased their database, recruiting many haplotypes which are obviously derived from "older" ancestors and have non-proportionally many mutations. For example, in May 2011 the "Red Group" contained 148 of 12 and 25 marker haplotypes, which have 98 and 260 mutations, and 102 of 67 marker haplotypes, which have 454 mutations from their base haplotypes. It gives:

98/148/0.022 = 30 generations from their apparent "common ancestor", which is likely a phantom one.

260/148/0.046 = 38 generations (!) from their (certainly) phantom "common ancestor", which allegedly lived 38 → 40 generations, that is 1000 years back. It is as many as 14 generation older than Lord John of the Isles.

454/102/0.12 = 37 → 38 generations, that is 950 years back.

The confusions regarding mutation rate constants in the literature continue. Some authors claim that only father-son pair studies can give the right value for mutation rate constants. This, of course, is the most objective approach, when done with a good statistics, that is with tens of thousands of pairs or more, preferably millions of pairs. However, there is one serious caveat – data would be obtained in mutations per a generation only, hence, again the issue of how

many years per generation should be employed is left unresolved. Again, we need a calibration with actual historical data, that is come back to our approach.

Here is an example. A massive study of almost two thousand of father-son pairs (Ballantyne et al, 2010) revealed 48 mutations in the first 12 markers and 102 mutations in the first 25 markers (unfortunately, 37 and 67 marker haplotypes have not been studied in full, many markers were missing) in 1727±49 and 1704±86 pairs, respectively. This gives mutation rate constants as 0.028±0.005 and 0.060±0.009 for 12 marker and 25 marker haplotypes, respectively. Another way of calculations, per (extrapolated) 10,000 father son-pairs and for each marker gave practically the same values, 0.0277±0.0049 and 0.0595±0.0084 mutations per haplotype per generation. They are exactly the same values as the mutation rate constants employed in our study, but for 32 years per generation, not for 25 years, used in our study. In practical terms, it does not make any difference, except statistics in the Ballantyne et al study was still rather poor. For example, even with almost 2000 father-son pairs, a number of mutations in the first 12 marker haplotypes was, respectively, 3, 2, 7, 5, 3, 6, 0, 0, 6, 9, 1, 6. In the following 13 markers it was 14, 4, 0, 0, 3, 2, 0, 19, 12 (some markers were combined in the cited study). Hence, margin of error for those data was rather wide, namely ±14% and ±10%, respectively, only for a number of mutations observed in the cited study.

Being applied for said Clan Donald "Red Subgroup", the father-son pair mutation rate constants give:

12 marker haplotypes:

44/68/0.028 = 23 generations

64/125/0.028 = 18 generations

25 marker haplotypes:

69/60/0.060 = 19 generation

166/124/0.060 = 22 generations

An average number of generations is 20.5±2.4, that is 656±77 years to a common ancestor of the "Red Subgroup" of R1a1 individuals, at 32 years per generation.

As one can see, for a practical application of father-son data and the respective mutation constants, it is necessary to calibrate the data.

The principal conclusion is that the mutation rate constants used in this study and listed above, are appropriate, and they do not contradict with father-son mutation studies.

Null mutations and recLOH mutations

In case of null mutations they were counted as one mutation compared to the respective base haplotype. However, null mutation haplotypes typically do form their separate branches on the haplotype tree. In that case there is no need to count null mutations, since all haplotypes in the branch and their base haplotype all contain the null mutation.

In case of recLOH mutations they were counted as one mutation regardless how large was a gap between the base allele and a resulting recLOH allele. For example, if the majority of haplotypes in a dataset having one common ancestor (that is verified using the haplotype tree along with the logarithmic method compared with the linear method, see above) have 19-21 in their YCAII loci, and some haplotypes contain 19-19 or 21-21, it is counted each as one mutation, not two. If the base DYS464 alleles are 15-15-17-17, then 15-17-17-17 or 15-15-18-18, is counted as one mutation, not two. If the base DYS385 alleles are 11-14, then 11-11 is counted as one mutation, not three. The same goes for 14-14 in these loci. 12-13 in these loci is counted as two mutations as well. In many cases recLOH or no recLOH mutations give the same result, such as 9-9 or 10-10 in DYS459 with its (typically) base 9-10, it would be one mutation anyway, recLOH or no recLOH.

Correction for back (reverse) mutations

Essentially, when a mutation happens, and they happen now and then, they can equally happen "up" or "down", e.g. from the ancestral 17 to either 18 or 16 (in a descendant). If it mutates to, say, 18, the next move (after 550 generations on average for 25 marker haplotypes or 460 generations for 67 marker haplotypes) can be to 17 or 19, equally probable.

If it mutates back to 17, to the ancestral allele, this would be the "back mutation", and one cannot see it, looking at the resulting haplotype. It was 17 in the ancestral haplotype, and it went 17--> 18 --> 17, that is 17 again. How do we know that ANY allele in a present day haplotype is not back mutated?

In fact, we do not know. However, we can calculate a probability of such an event in all the 67 alleles in a 67 marker haplotype, or in a haplotype of any format. Those back mutations actually slow down the OBSERVED mutations. We

observe, say, 125 mutations in a dataset, and we calculate that in fact there were 137 mutations. This is a correction for back mutations.

There are two principal ways to introduce a correction for back mutation into the calculations. One way is to use the following formula (Adamov & Klyosov, 2008; Klyosov, 2009a)

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

where:

λ_{obs} = observed average number of mutations per marker in a dataset (or in a branch, if the dataset contains several branches/lineages),

λ = average number of mutations per marker corrected for back mutations.

The above formula is applicable for completely symmetrical pattern of mutations, that is for equal number of mutations "up" and "down" from the base (ancestral) haplotype. For asymmetrical series of haplotypes (rather, for mutations in the dataset) a degree of asymmetry should be calculated and a slightly more complicated formula (Klyosov, 2009a) should be used; however, this additional factor is, as a rule, not very significant, and typically fits into a margin of error of calculations.

Let us consider an example of a dataset of 100 of 25 marker haplotypes, containing 400 mutations from the base haplotype. Then 400/100/25 = 0.160 mutations per marker. At the mutation rate of 0.002 it would give 0.160/0.002 = 80 generations, that is 80x25 = 2,000 years to a common ancestor. However, as it was mentioned above, with 24 generations and deeper in time one should introduce a correction for back mutations. At 80 generations it is almost two centuries. Here is how it works:

$$\lambda = \frac{0.160}{2}(1 + \exp(0.160)) = \frac{0.160}{2}(1 + 1.174) = 0.174$$

0.174/0.002 = 87 generations, that is 87x25 = 2,175 years to a common ancestor.

Another way is to use the handy Table (Klyosov, 2009a), which provides two columns of data – one without a correction for back mutations, second for the corrected value. For our example it shows that 80 generations, not corrected for back mutation, correspond to 87 generations after the correction is made. It is

exactly the same value of 87 generations calculated above using the mathematical formula.

As an example of the logarithmic method with a correction for back mutations, let us consider a series of 750 of 19 marker Basque and Iberian R1b1 haplotypes (Adams et al, 2009) containing 16 of identical haplotypes, that is **base** haplotypes in the series. It gives [ln(750/16)]/0.0285 = 135 generations without a correction for back mutations. The correction Table described above immediately gives 135 → 156 generations (corrected), that is 156x25 = 3900 years to a common ancestor of the Basque and Iberian haplotypes, predominantly subclade R1b1b2-P312*. This is within the margin of error with the timespan to a common ancestor of Basque and Iberian haplotypes calculated using the linear method (Klyosov, 2009a).

49 marker haplotypes

Some calculations below have been done using 49 marker haplotypes, which represent a 67 marker panel from which 18 markers were removed due to their inclinations to recLOH mutations, null mutations and other complications. Removed markers were DYS385a,b; DYS389-2; DYS459a,b; DYS464a,b,c,d; YCAIIa,b; CDYa,b; DYS395S1a,b; DYS425; DYS413a,b. The mutation rate constant for the 49 marker panel was equal to 0.080 mutation/haplotype/generation, that is 0.00163 mutation/marker/generation of 25 years. 102 haplotypes of the Clan Donald (R1a1 "Red Group") series in the 49 marker format fit that mutation rate constant within 10% of its value.

Haplotype analysis. Actual 67 marker haplotypes

A series of 2299 of 67 marker haplotypes which is considered in this paper, is too "heavy" for a contemporary personal computer to be calculated altogether into a haplotype tree. Typically, 1100-1300 haplotypes in the 67-marker format is a limit for composing of haplotype trees when PHYLIP program is employed. Therefore, the haplotypes were considered by their subclades.

As a first approximation the logarithmic method was used. The whole series of 2299 of 67 marker haplotypes contained 154 of 12 marker base haplotypes

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

which gives [ln(2299/154)]/0.022 = 123 →141 generations, that is approximately 3525 years to a common ancestor of all 2299 haplotypes.

The same series contained seven base 25 marker haplotypes

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

which gives [ln(2299/7)]/0.046 = 126 →145 generations, that is approximately 3625 years to a common ancestor. This is a quite a reasonable dating for the P312 subclade, which typically varies from 3600 to 4100 ybp for various series of haplotypes.

It should be noted that these two series of 12- and 25 marker base haplotypes obtained using the logarithmic method gave a difference of only four generations, or less than 3% between each other, and generally fit into a typical range of TSCA (timespans to common ancestors) for P312 subclade. This shows that all 2099 of 67 marker haplotypes are derived - with a good probability – from one common ancestor.

Among those 154 base (presumably ancestral haplotypes and other statistically accidental fits through back mutations) 12 marker haplotypes there were:

26 haplotypes from England (8.1%)
26 from Ireland (4.5%)
19 from Scotland (7.0%)
8 from Germany (7.8%)
6 from Spain (10.2%), Wales (9.4%), France (6.5%) each
2 from Denmark, Czech, Finland, Netherlands each
1 from Sweden, Switzerland, Greece, Luxemburg, Poland, Portugal,
and 43 did not have a geographical assignment.

Among seven 25 marker base haplotypes, four were from Scotland, two from Ireland, and one from England.

Statistics with 25 marker haplotypes is not too good to make any meaningful conclusion, however, the highest amount of base 12 marker haplotypes in Spain might be indicative regarding the origin of P312 in Europe (see above). The highest amount of P312* haplotypes is found also in Iberia (Myres et al, 2009).

R1b1a2-P312 and its six immediate downstream subclades

Here is how all six immediate downstream subclades are represented in the 2299 67 marker haplotype dataset (without including their downstream subclades):

L21	1024 (45%)
U152	162 (7%)
M153	5 (0.2%)

L176.2*	5 (0.2%)
L238	2 (0.09%)
M65	1 (0.04%)

With inclusion of some downstream subclades, L176.2 would add L165 (9 haplotypes) and SRY2627 (141 haplotypes) and with its total 155 haplotypes takes 6.7% of all, close to that in U152. However, L21 would add 462 haplotypes of its downstream subclades M222, L144, L159, L193, L226, and P314, and with 1486 haplotypes takes 65% of all, two-thirds of the dataset.

L21 haplotypes

L21 is one of six immediate downstream subclades of R1b1a2-P312 (see the chart above). However, it is the most populous among them all.

Fig. 1 shows a tree of 1024 of 67 marker haplotypes R1b1a2-L21, downstream of P312. Of those the first 220 haplotypes (numbers 1-220) were marked (by the P312 Project) as L21, 701 haplotypes (221-922) were marked L21*, the last 102 haplotypes (923-1024) were marked L21**. All the three marked groups were scattered around the tree.

According to Michael Walsh, asterisk in a subclade index shows that it was confirmed that a haplotype is ancestral for the major downstream SNPs. For example, P312* is P312+ L21- U152- SRY2627- M222- M153-. However, since L21 was not always included in some early testing packages, P312* can in those cases be P312+ L21+.

The whole tree has the following base haplotype:

13 24 14 11 11 14 12 12 12 13 13 29 – 17 9 10 11 11 25 15 19 29 15 15 17 17 –
11 11 19 23 **16** 15 18 17 36 38 12 12 – 11 9 15 16 8 10 10 8 10 10 12 23 23 16 10 12 12
15 8 12 22 20 13 12 11 13 11 11 12 12 (**L21, 3750±380 ybp**)

This haplotype differs by only one mutations (marked, in DYS 456) compared with the parent P312 base haplotype (see above). To be exact, in 1024 of L21 haplotypes an average allele in DYS456 was 15.71, that is close to 16. This places apart common ancestors of P312 and L21 by 1/0.12 = 8 generations only, that is by 200 years, That is, if P312 arose 4100 ybp, L21 should arose approximately 3900 ybp.

Let's check it.

All 1024 haplotypes of R1b1a2-L21 have 16,056 mutations from the base haplotype shown above. It gives 16056/1024/0.12 = 131→150 generations, that is 3750±380 years from the common ancestor of L21 subclade. This value is the same as that (3750 ybp) calculated from a mutational difference from the P312 and L21 base haplotypes.

Those 1024 haplotypes have 4 base 25 marker haplotypes, which gives [ln(1024/4)]/0.046 = 121 → 138 generations, that is approximately 3450 years to the common ancestors. These values, 3750±380 ybp obtained by the "linear" method and 3450 ybp by the logarithmic method are the same within the margin of error.

A small branch of 108 haplotypes within L21 subclade

Fig. 1 reveals one clearly separate branch within L21 subclade, at 7 o'clock, containing 108 haplotypes, with the base haplotype

13 24 14 **10** 11 14 12 12 12 13 13 **30** – **18** 9 10 11 11 25 15 19 **30** 15 15 17 17 –
11 **12** 19 **24** 16 15 18 17 **37** 38 12 12 – **12** 9 15 16 8 10 10 8 10 10 12 **22** 23 16 10 12 12
15 8 **11** 22 20 13 12 11 13 11 11 12 12

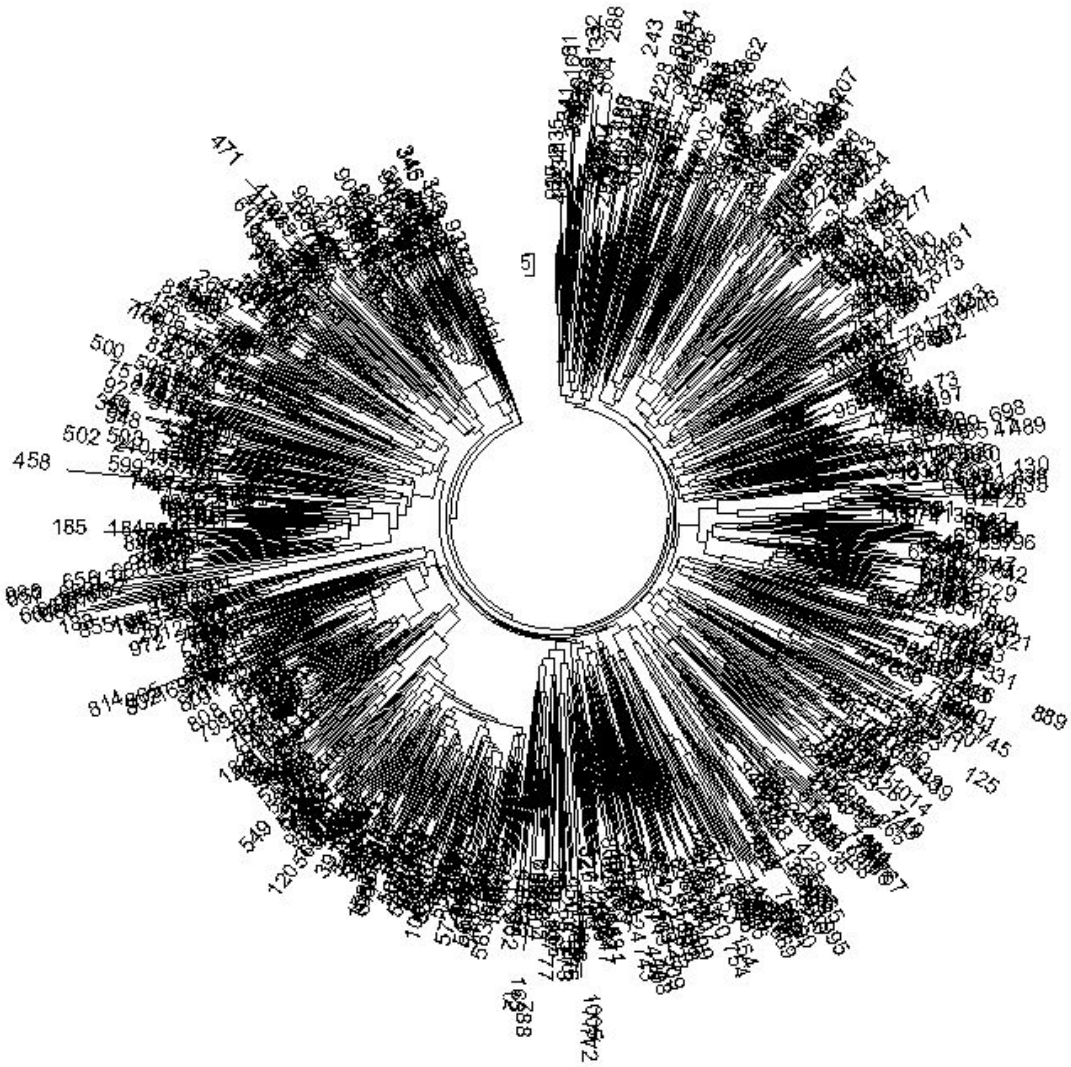


Fig. 1. 67 marker haplotype tree of 1024 haplotypes of R1b1a2-P312-L21. A list of haplotypes were provided by Michael Walsh, administrator of R1b1a2-L21 site
http://www.familytreedna.com/public/R-L21/default.aspx?section=yresults**)**

This branch composed as a separate tree is shown in Fig. 2.

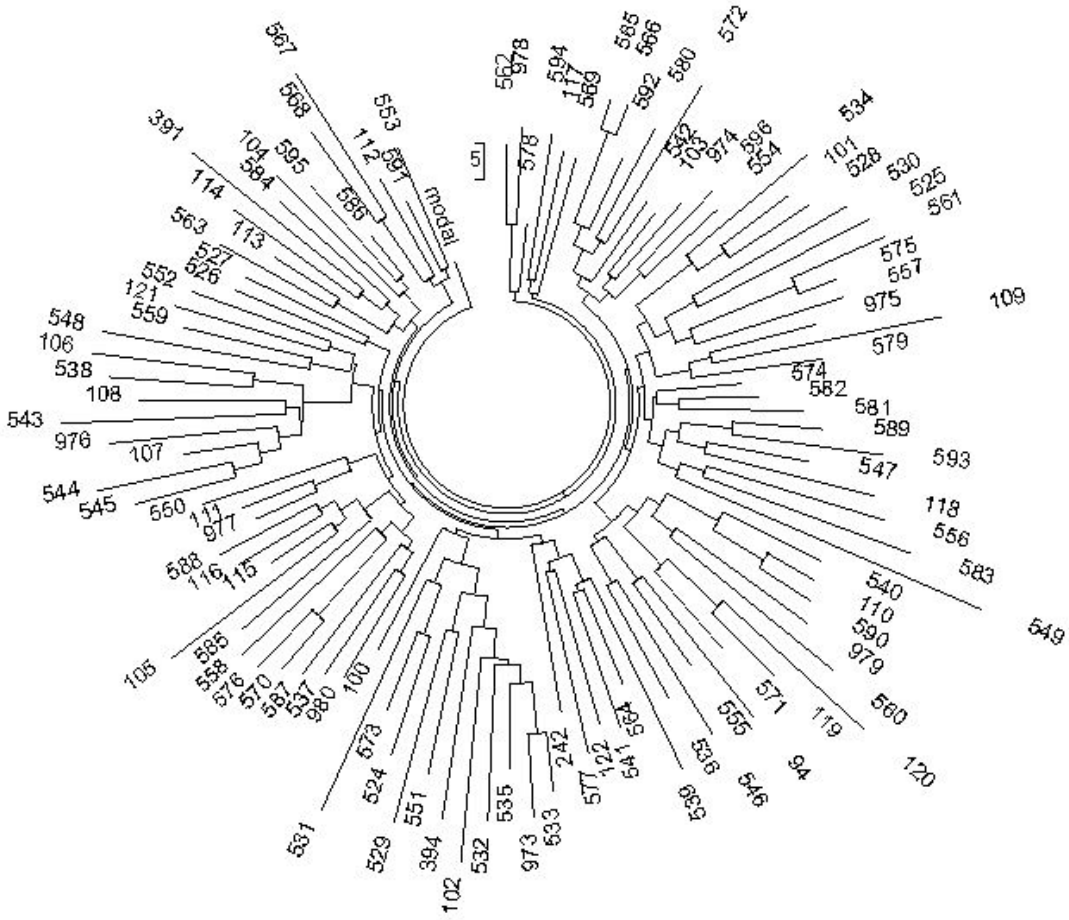


Fig. 2. 67 marker haplotype tree of 108 haplotypes of R1b1a2-L21 which form a distinct branch in Fig. 1 at 7 o'clock. The tree contains 24 haplotypes marked as L21 (numbers 94 and 100-122), 76 haplotypes marked as L21* (between 242 and 596), and 8 haplotypes marked as L21 (973-980). Those groups did not form any distinct sub-branches.**

The base haplotype shown above differs by 10 mutations from the L21 base haplotypes (shown above in bold). Surprisingly, it is seemingly not identified as yet subclade of L21. Its 108 haplotypes contain 798 mutations, which give 798/108/0.12 = 62 →66 generations, that is 1650±175 years to a common ancestor of the branch. It is a middle of the 1st millennium AD.

It is not the M222 base haplotype

13 **25** 14 **11** 11 **13** 12 12 12 13 **14 29 – 17** 9 10 11 11 25 15 **18** 30 15 **16 16** 17 –
11 **11** 19 **23 17 16** 18 17 **38 39** 12 12 – **11** 9 15 16 8 10 10 8 10 10 12 **21** 23 16 10 12 12
16 8 **12 25** 20 13 12 11 13 11 11 12 12

with a common ancestor of 1450±160 ybp (Klyosov, 2010a,b), and with 22 mutations (shown in bold) between the base haplotype of M222 and that of the unidentified subclade above. Those 22 mutations between two 67 marker haplotypes place their common ancestors by 5600 years apart, and put THEIR common ancestor at 4350 ybp. Apparently, it was the "original" L21 common ancestor before his descendants passed the population bottleneck.

It is not the L226 (Irish Type III) base haplotype

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

with a common ancestor of 1450±290 ybp (Klyosov, 2010b), and with 20 mutations (marked above) between the base haplotype of L226 and that of the unidentified subclade above. Those 20 mutations between two 67 marker haplotypes place their common ancestors by 5000 years apart, and put THEIR common ancestor at 4050 ybp. Again, it was apparently the "original" L21 common ancestor before his descendants passed the population bottleneck.

Despite L222 (R1b1a2a1- b4b) and L226 (R1b1a2a1-b4h) were ruled out, there are still at least seven subclades, downstream of L21 which could be considered for this new branch with a common ancestor of 1650±175 ybp. Those are M37 (-b4a), P66 (-b4c), L96 (-b4d), L144/L195 (--b4e), L159.2 (--b4f), L193 (-b4g),P314.2 (-b4i). Its assignment remains to be established.

Since the 1024 haplotype series of L21 contained 108 haplotypes of some younger, downstream branch, with a common ancestor of 1650±175 ybp, this admixture could distort the TSCA value for L21, making it higher or lower, depending on specifics of the admixture. However, since the admixture adds only about 10% of all the haplotypes, the distortion might be negligible. Let us examine it.

When the 108 haplotypes of the "young branch" were subtracted from the 1024 L21 haplotypes, the base haplotypes stayed the same, and 14,278 mutations were found from it in the remaining 916 haplotypes. It gave 14278/916/0.12 = 130 → 149 generations, that is 3725 years to a common ancestor, which is practically equal to 3750 years for the initial 1024 haplotypes. One can see that in this particular case the reduction of the number of haplotypes in the series (from 1024 to 916, by 10.6%) was only a little lower than the reduction of a number of mutations (from 16,056 to 14,278, by 11.1%), hence, there was practically no distortion in the TSCA.

L21 null mutation in a series of haplotypes

Null mutations happen in various markers, however, DYS425 is probably the most subjected to null mutations in the 67 marker panel. Null mutation is not, of course, a complete elimination of the marker altogether. It just shows that an allele in this marker cannot be determined by the same method (the same primer) as that successfully employed with other populations of haplotypes. A null mutation is inherited, that is transmitted onto haplotype of a descendant. Hence, it is a "marker" by itself.

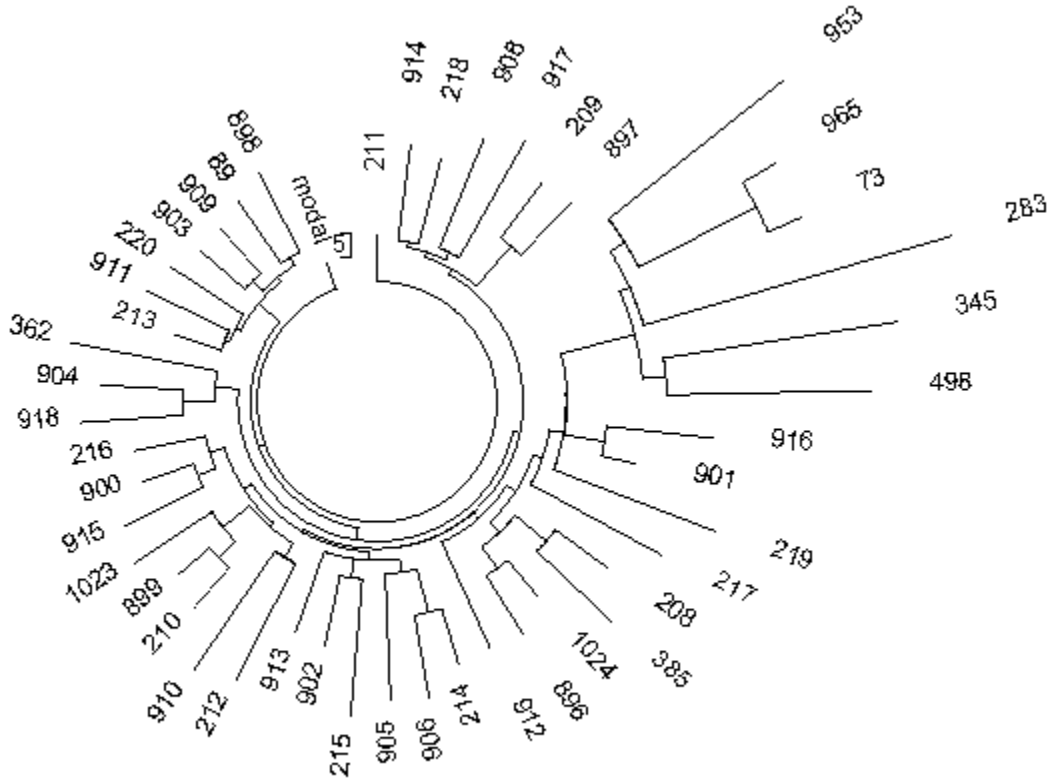


Fig. 3. 67 marker haplotype tree of 46 haplotypes of R1b1a2-L21 null mutation series. Haplotypes numbered between 73 and 220 (15 haplotypes total) belong to L21, 283-918 (27 haplotypes total) belong to L21*, haplotypes 953, 965, 1023 and 1024 belong to L21. There are 24 Ireland haplotypes, five Scotland, four England, one Wales, Spain and Sweden each, the rest are of unknown ancestry.**

In the considered dataset of 2299 haplotypes null mutations occur in the L21 subclade (46 null mutated haplotypes), U152 (seven haplotypes), SRY2627 (six haplotypes), L20 and P312 (three haplotypes in each), L193 and M222 (one haplotype in each).

Fig. 3 shows a haplotype tree of null mutated haplotypes of L21 subclade. The branch of six haplotypes on the right-hand side is obviously the most ancient one. Its base haplotype differs by 5 mutations from that of L21 parent subclade (deviations from L21 are marked below):

13 24 14 11 11 14 12 12 12 13 13 **30** – 17 9 10 11 11 25 15 19 29 15 15 **16** 17 –
11 11 19 23 **15** 15 18 **18** 36 38 12 12 – 11 9 15 16 8 10 10 8 10 10 **0** 23 23 16 10 12 12
15 8 12 22 20 13 12 11 13 11 11 12 12 (**L21 null DYS425, 3025±460 ybp**)

13 24 14 11 11 14 12 12 12 13 13 29 – 17 9 10 11 11 25 15 19 29 15 15 17 17 –
11 11 19 23 16 15 18 17 36 38 12 12 – 11 9 15 16 8 10 10 8 10 10 12 23 23 16 10 12 12
15 8 12 22 20 13 12 11 13 11 11 12 12 (**L21, 3750±380 ybp**)

All six haplotypes contain 78 mutations from the above base haplotype, which gives 78/6/0.12 = 108 → 121 generations, that is 3025±460 years from a common ancestor. Five mutations between L21 and L21-null base haplotypes results in 5/0.12 = 42 → 44 generations, that is 1100 "lateral" years between the common ancestors. It indicates that THEIR common ancestor lived approximately (1100+3750+3025)/2 = 3940 years ago. It is in a vicinity of 4050 ybp for the common ancestor for L21, which was described above as "it was apparently the "original" L21 common ancestor before his descendants passed the population bottleneck".

Therefore, it is very likely that the L21-null branch split from the initial L21 common ancestor at the very beginning (time-wise) of the subclade.

Let us verify it with the more recent branches in Fig. 3.

Immediately at the foot on the just described old branch there is a small branch of four haplotypes. Still, they provide us with as many as 268 markers (67x4). The base haplotype of the branch is as follows (mutations from the 6-haplotype old branch base haplotype are marked):

13 24 14 11 11 **15** 12 12 12 13 13 **29** – 17 9 10 11 11 25 15 19 **28** 15 15 **17** 17 –
11 11 19 23 15 15 18 **19** 36 38 12 12 – 11 9 15 16 8 10 10 8 10 **9** 0 **22** 23 16 10 12 12 **16**
8 12 22 20 13 12 11 13 11 11 12 12 (**L21 null DYS425, sub-branch, 1500±325 ybp**)

All four haplotypes contain 27 mutations from the above base haplotype, which gives 27/4/0.12 = 56 → 60 generations, that is 1500±325 years from a common ancestor. 7 mutations (most of them are fractional ones) between the two null base haplotypes result in 7/0.12 = 58 → 62 generations, that is 1550 "lateral" years between the common ancestors, and place THEIR common ancestor to approximately (1550+1500+3025)/2 = 3040 years ago. It is practically equal to the "age" of the older branch (3025 ybp). In other words, the "older' null branch split

off L21 lineage, and after more than a thousand years, namely 1500 ybp, it split the "younger" null branch.

There are two other "young" small branches on the right-hand side of the tree: one, 6-haplotype branch in the upper part (between haplotypes 914 and 997), with the base haplotype

13 24 14 11 11 15 12 12 **14** 13 13 29 – 17 9 10 11 11 25 15 19 28 15 15 17 17 –
11 11 19 23 **16** 15 18 **18** 36 37 12 12 – 11 9 15 16 8 10 10 8 10 9 0 22 23 16 10 12 12 16
8 12 22 20 13 12 11 13 11 11 12 12 (**L21 null DYS425, 1050±220 ybp**)

and another, 5-haplotype branch, in the lower part of the tree (between 208 and 912), with the base haplotype

13 24 **15** 11 11 15 12 12 12 13 13 29 – 17 9 10 11 11 25 15 19 28 15 15 17 17 –
11 11 19 23 **16** 15 18 19 36 38 12 12 – 11 9 15 16 8 10 10 8 10 9 0 22 23 16 10 12 12 16
8 12 22 20 13 12 11 13 11 11 12 12 (**L21 null DYS425, 1100±250 ybp**)

each has a (different) common ancestor, who both lived only about a thousand years ago: 29/6/0.12 = 40 → 42 generations, that is 1050±220 ybp, and 25/5/0.12 = 42 → 44 generations, that is 1100±250 ybp. They differ by only 4 and 2 mutations (850 and 425 "lateral" years) from the above null-mutated branch of 1500±325 ybp (marked in bold, some mutations are fractional), hence, they are derived from the 1500-ybp null-L21 common ancestor.

The last branch, stretching from the top to the bottom on the left-hand side of the tree in Fig. 3, consisting of 25 haplotypes, has the following base haplotype:

13 24 14 11 11 **15** 12 12 **13** 13 13 **29** – 17 9 10 11 11 25 15 19 **28** 15 15 **17** 17 –
11 11 19 23 **16** 15 18 18 36 **37** 12 12 – 11 9 15 16 8 10 10 8 10 **9** 0 **22** 23 16 10 12 12 **16**
8 12 22 20 13 12 11 13 11 11 12 12 (**L21 null DYS425, 1300±170 ybp**)

All the 25 haplotypes contain 148 mutations, which gives 148/25/0.12 = 49 → 52 generations, that is 1300±170 years from their common ancestor. It is the same null mutated haplotype of 1500±325 ybp common ancestor, within the margins of error of calculations.

We have already established that the 3025±460 ybp and the 1500±325 ybp base haplotypes differ by 7 mutations, which shows that the second one is a downstream from the first one.

Let us take a look at the presumed upstream (1500 ybp) and downstream, younger null mutated base haplotypes.

13 24 14 11 11 15 12 12 12 13 13 29 – 17 9 10 11 11 25 15 19 28 15 15 17 17 –
11 11 19 23 **15** 15 18 19 36 38 12 12 – 11 9 15 16 8 10 10 8 10 9 0 22 23 16 10 12 12
16 8 12 22 20 13 12 11 13 11 11 12 12 (**L21 null DYS425, 1500±325 ybp**)

13 24 14 11 11 15 12 12 **13** 13 13 29 – 17 9 10 11 11 25 15 19 28 15 15 17 17 –
11 11 19 23 16 15 18 **18** 36 **37** 12 12 – 11 9 15 16 8 10 10 8 10 9 0 22 23 16 10 12 12
16 8 12 22 20 13 12 11 13 11 11 12 12 (**L21 null DYS425, 1300±170 ybp**)

13 24 **15** 11 11 15 12 12 12 13 13 29 – 17 9 10 11 11 25 15 19 28 15 15 17 17 –
11 11 19 23 16 15 18 19 36 38 12 12 – 11 9 15 16 8 10 10 8 10 9 0 22 23 16 10 12 12
16 8 12 22 20 13 12 11 13 11 11 12 12 (**L21 null DYS425, 1100±250 ybp**)

13 24 14 11 11 15 12 12 **14** 13 13 29 – 17 9 10 11 11 25 15 19 28 15 15 17 17 –
11 11 19 23 16 15 18 **18** 36 37 12 12 – 11 9 15 16 8 10 10 8 10 9 0 22 23 16 10 12 12
16 8 12 22 20 13 12 11 13 11 11 12 12 (**L21 null DYS425, 1050±220 ybp**)

The above four base haplotypes contain 9 mutations, that is 9/4/0.12 = 19 generations, or 475 years plus 1200 years (an average ybp for all four base haplotypes), which gives 1675±240 to THEIR common ancestor. This is the 1500 ybp common ancestor within the margin of error.

The above can be summarized as the oldest null mutated L21-DYS425=0 base haplotype split from L21 3025±460 ybp, then the null mutated 1500±325 descendant split, and the latter has derived a series of null mutated lineages with common ancestors of 1300±170, 1100±250 and 1050±220 ybp. It seems that all of them were split in parallel, not consecutively.

Subclades other than L21

R1b1a2-P312* and subclades (M65, M153, L21-M222, L21-L144, L21-159.2, L21-L193, L21-L226, L21-P314.2)

Fig. 4 shows the 67 marker haplotype tree of R1b1a2-P312 and its subclades. Its base haplotype is as follows

13 24 14 11 11 14 12 12 12 13 13 29 – 17 9 10 11 11 25 15 19 29 15 15 16 17 –
11 11 19 23 16 15 18 17 37 38 12 12 – 11 9 15 16 8 10 10 8 10 10 12 23 23 16 10 12 12
15 8 12 22 20 13 12 11 13 11 11 12 12 **(base haplotype for the tree in Fig. 4)**

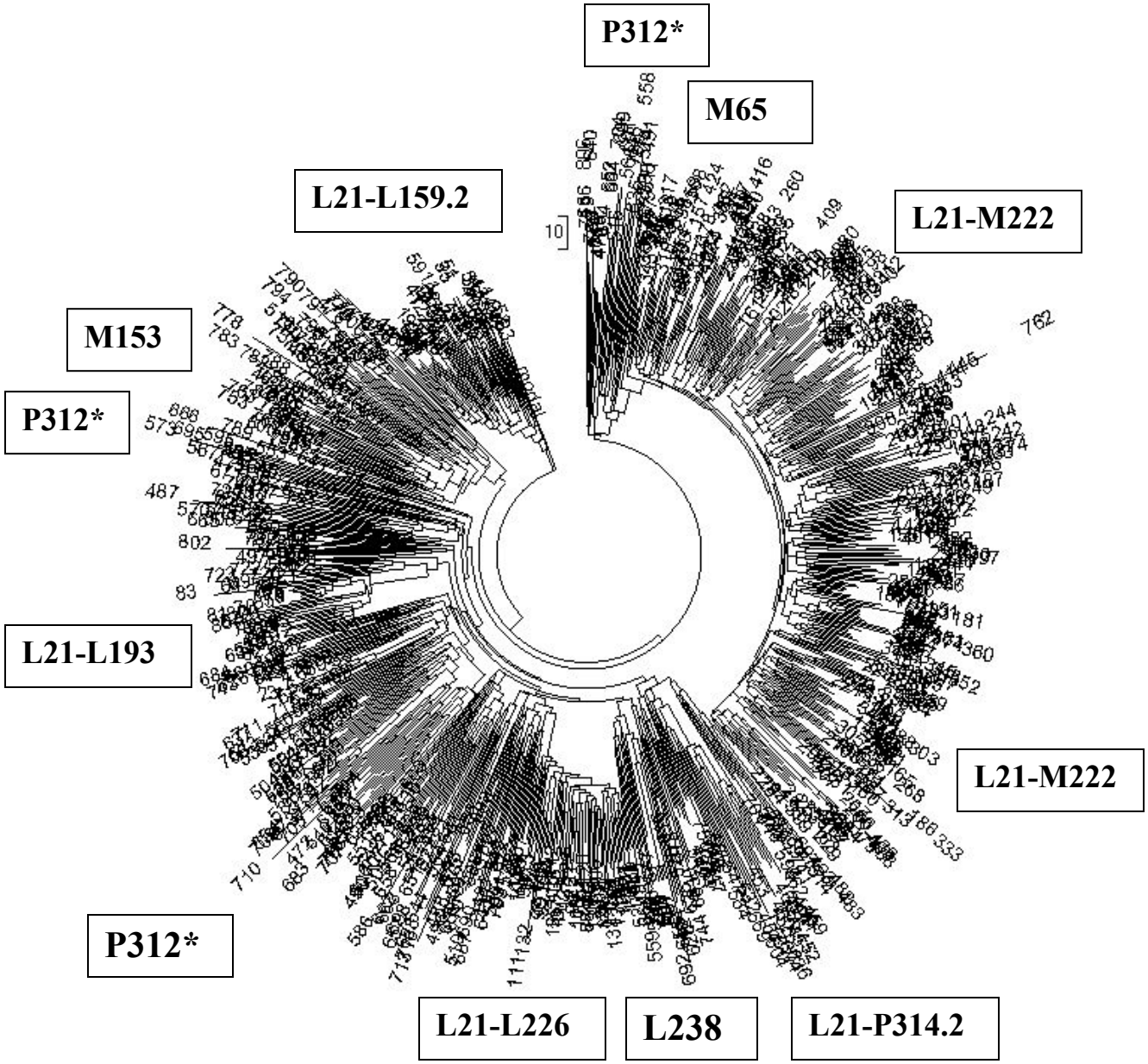


Fig. 4. 67 marker haplotype tree of 808 haplotypes of R1b1a2-P312 with some subclades. The haplotypes were provided by Michael Walsh, administrator of R1b1a2-L21 site http://www.familytreedna.com/public/R-L21/default.aspx?section=yresults**)**

This base haplotype is very close to that of P312 (see above), and differs by only fractions of mutations. The first 25 alleles contain 5252 mutations from the base haplotype, which gives 5252/808/0.046 = 141 → 164 generations, that is 4100±415 years from a common ancestor.

The whole 808 haplotype dataset contains 39 of 12 marker base haplotypes

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

This gives ln(808/39)/0.022 = 138 → 161 generations, that is approximately 4025 years from a common ancestor. Since the both methods, the linear and the logarithmic, gave very close TSCAs (4100 and 4025 ybp), it shows that the whole tree was derived from one common ancestor, of the P312 subclade.

P312-M65 subclade

There is only one M65 haplotype (from Wales) in the whole dataset:

13 24 14 11 11 14 12 12 12 13 **14** 29 – 17 9 10 11 11 25 15 19 **30** 15 15 16 **18** –
11 11 19 23 **15** 15 **17** 17 **38 41** 12 12 – 11 9 15 16 8 10 10 8 10 10 12 23 **25** 16 10 12 12
15 8 **11** 22 20 13 12 11 13 11 11 12 12 **(R1b1a2-P312-M65, ~ 1800 ybp)**

Since it is a single haplotype, its mutations (marked above) compared to the base haplotype of the whole tree (see above) may be not statistically justified. Let us take it as a semi-quantitative comparison. 10 mutations between two 67 marker haplotypes correspond to about 2275 years of the mutational evolution. This places a (tentative) common ancestor for the M65 subclade to the beginning of AD, approximately 1800 ybp.

P312-M153 subclade

There are only five haplotypes of this subclade on the tree, with 29 mutations from the base haplotype

13 24 14 11 11 14 12 12 12 13 13 29 – 17 9 10 11 11 25 **14 18** 29 15 15 **17** 17 –
10 11 19 23 16 15 18 17 **35** 38 12 12 – 11 9 15 16 8 10 10 8 10 10 12 23 23 16 10 12 12
15 8 12 22 20 13 12 11 13 11 11 12 12 **(R1b1a2-M153, ~ 3640 ybp)**

and with 4 mutations from the base haplotype of P312. This places a common ancestor of these five M153 haplotypes at 29/5/0.12 = 48 → 51 generations, that is 1275±270 ybp, and with 850 years of a mutational distance from the P312 common ancestor. These figures do not fit, since the P312 common ancestor lived around 4100 ybp. It seems that those five haplotypes are related to each other, hence, a close time to their common ancestor. Indeed, an analysis of 37 haplotypes of the M153 subclade (Klyosov, 2010c) showed that a common ancestor of those haplotypes lived 2325±340 ybp. With this figure a common ancestor of M153 and P312 lived (2325+4100+850)/2 = 3640 ybp, which is the P312 itself within a margin of error of calculations.

L21-M222

312 haplotypes of this subclade take almost the whole right-hand side of the tree in Fig. 4. A detailed analysis of this subclade was performed in (Klyosov, 2010b), and the following base haplotype was determined:

13 25 14 11 11 13 12 12 12 13 14 29 – 17 9 10 11 11 25 15 18 30 15 16 16 17 –
11 11 19 23 17 16 18 17 38 39 12 12 – 11 9 15 16 8 10 10 8 10 10 12 21 23 16 10 12 12
16 8 12 25 20 13 12 11 13 11 11 12 12 **(R1b1a2-L21-M222, 1450±160 ybp)**

Analysis of 257 haplotypes of this subclade showed that a common ancestor of this subclade lived **1450±160 ybp**. 312 of M222 haplotypes in this dataset (of 2299 haplotypes) contained 98 of 12 marker base haplotypes and 25 of 25 marker base haplotypes. This gives ln(312/98)/0.022 = 53 → 56 generations, that is **1400 ybp** with 12 marker haplotypes, and ln(312/25)/0.046 = 55 → 58 generations, that is **1450 ybp** with 25 marker haplotypes. These figures are practically identical

L21-L144

There are only 7 haplotypes of this subclade among all 2299 haplotypes of R1b1a2-P312 the dataset. Six of them form a rather tight branch (see Fig. 5), with the base haplotype (17-step deviations from the parent L21 base haplotype are marked in bold [some deviations are fractional])

13 24 14 **10** 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 15 **17** **18** 36 **40** 12 12 – 11 **10** 15 16 8 10 10 8 10 10 12 **16** 23 16 10 12 12
15 8 12 22 20 **12** **10** 11 13 11 11 12 12 (**L21-L144, branch, 925±200 ybp**)

and one haplotype (number 1 on the tree) very distant from the pack (28 mutations compared to the above base haplotype are marked in bold):

13 24 14 **11** 11 **15** 12 12 **11** 13 13 29 – 17 9 **9** 11 11 25 15 19 **29** 15 15 **15** 17 –
11 **10** 19 23 **17** 15 **20** **17** **39** 40 12 **13** – 11 **9** 15 16 8 10 10 8 10 10 12 **23** 23 16 10 **13** 12
16 8 12 22 20 **13** **12** 11 13 11 11 12 12 (**L21-L144, single haplotype**)

It is of interest that the single L144 haplotype and the tight 6-haplotype branch are equidistant from the parent L21 base haplotype (17 mutations in the both cases, which corresponds to 4150 "lateral" years) but have a huge amount of 28 mutations between them.

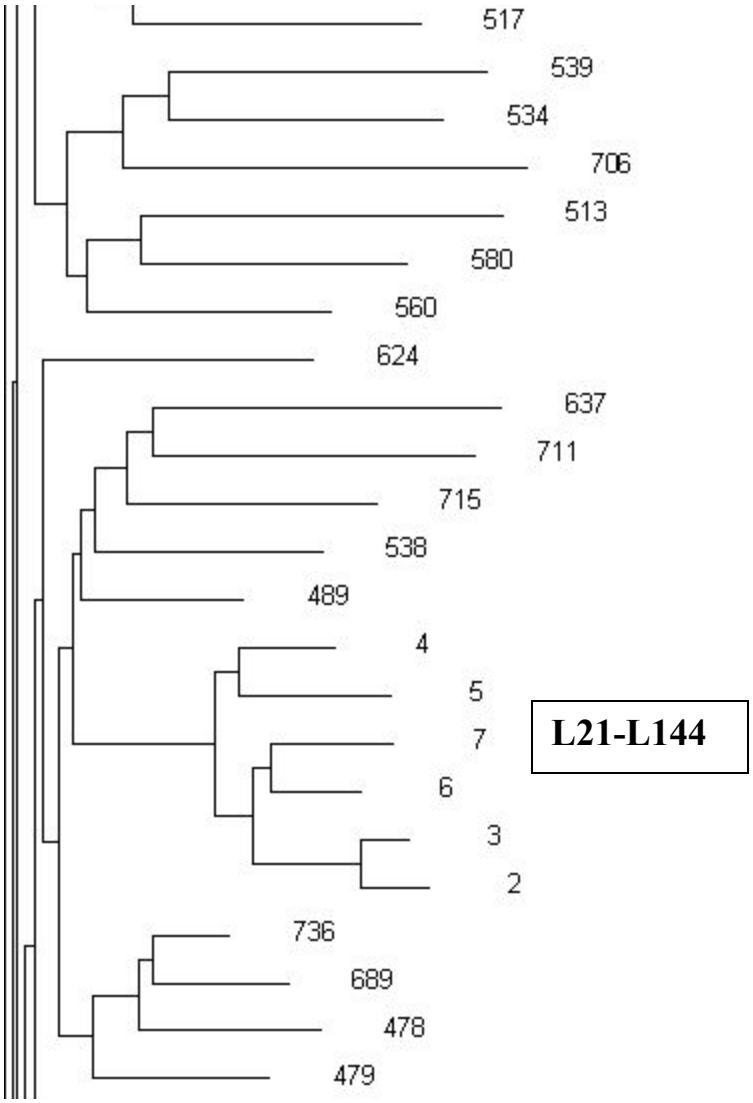


Fig. 5. A fragment of the 67 marker haplotype tree of 808 haplotypes of R1b1a2-P312 with some subclades. Haplotypes 1-7 belong to the subclade L21-L144. Six of them are rather closely related to each other, haplotype 1 is remote on the tree. All other haplotypes in this fragment (and all with numbers above 472) belong to subclade P312*.

Let us figure out why so. The 6-haplotype pack has 26 mutations from their base haplotype (see above) which gives 29/6/0.12 = 36 → 37 generations, that is 925±200 years from their common ancestor. This places a common ancestor of L21 and the tight pack of L144 to approximately 4400 ybp. This somewhat "deeper" than the time when L21 supposedly arose, however, a possible population bottleneck in the L21 subclade and/or margin of error of the calculations might be factors. Still, it suggests that L144 subclade is a direct

descendant of L21. The single L144 haplotype with its 17 mutations from the base of L21 places THEIR common ancestors at 3950 ybp, which clearly is L21 itself. 28 mutations between the two L144 places their common ancestor at 4200 ybp, which is again the L21 subclade itself.

In other words, the actual "age" of L21-L144 is close to 4000 years, and not to 925 years, found for just six (obviously closely related) representatives of the L144 subclade.

L21-159.2

A haplotype tree of 56 haplotypes of this subclade is shown in Fig. 6.

One can notice a separate branch of 16 haplotypes in the upper right-hand side of the tree. Still, the tree is derived from one common ancestor.

The upper 16-haplotypes branch has the following base haplotype:

13 24 14 11 11 14 12 12 12 **14** 13 30 – 17 9 10 11 11 25 15 **18 30** 15 15 17 17 –
11 11 19 23 16 15 18 **18 40 40 11** 12 – 11 9 15 16 8 10 10 8 10 10 12 23 23 **17** 10 12 12
15 8 12 22 20 **14** 12 11 13 11 11 12 12 (**L21-L159.2, sub-branch1350±190 ybp**)

It differs by 13 mutations (marked above) from its parent L21 base haplotype, and has 97 mutations in the whole branch. It gives 97/16/0.12 = 51 → 54 generations, that is 1350±190 years from its common ancestor. 13 mutations from L21 (which corresponds to 3025 "lateral" years between the two ancestor, and places THEIR common ancestor to (3025+3750+1350)/2 = 4060 ybp. It is the L21 common ancestor within the margin of error.

The rest of the tree (40 haplotypes) has the following base haplotype:

13 24 14 11 11 14 12 12 12 14 13 30 – 17 9 10 11 11 25 15 18 30 15 15 **16** 17 –
11 11 19 23 16 15 18 18 **38.5** 40 11 12 – 11 9 15 16 8 10 10 8 10 10 12 23 23 17 10 12 12
15 8 12 22 20 14 12 11 13 11 11 12 12 (**L21-L159.2, 1775±200 ybp**)

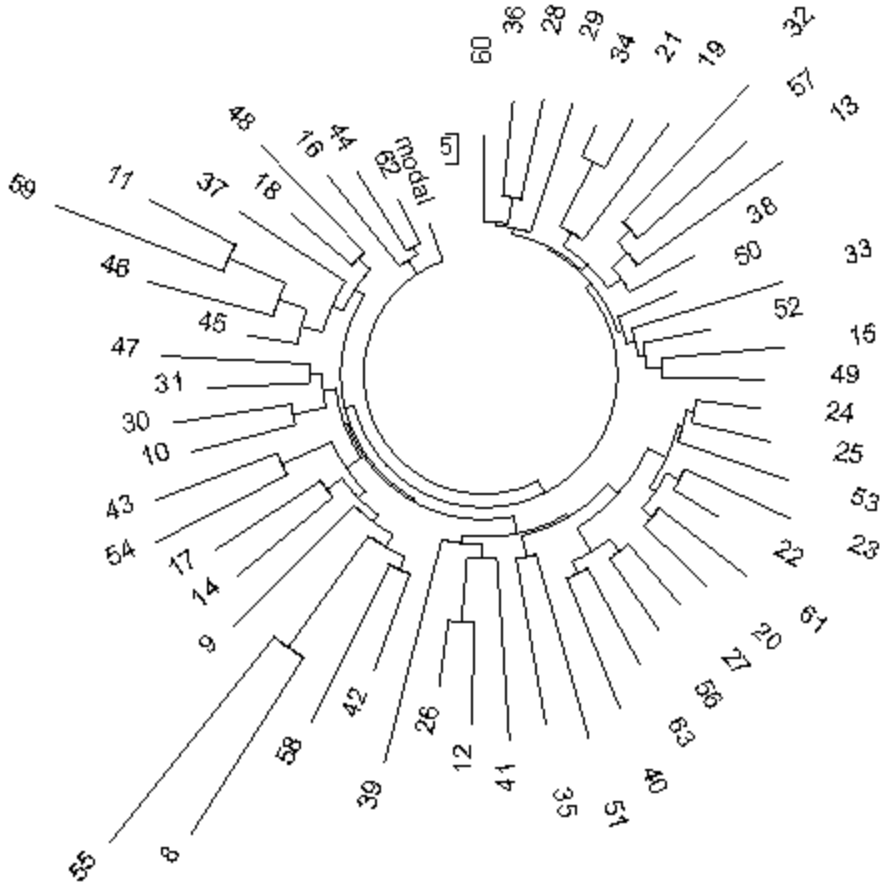


Fig. 6. 67 marker haplotype tree of 56 haplotypes of R1b1a2-L159.2 subclade.

The two base haplotypes differ by only 1.5 mutations, that is by 12 generations, or 300 years between their common ancestors. All 40 haplotypes contain 319 mutations from the above base haplotype, which gives 319/40/0.12 = 66 → 71 generations, that is 1775±200 years to a common ancestor. A common ancestor of the both parts of the tree lived approximately (1350+1775+300)/2 = 1710 ybp. This is the older base haplotype within the margin of error.

L21-L193

A haplotype tree of 24 haplotypes of this subclade is shown in Fig. 7.

The whole tree has the following base haplotype (11 deviations from L21 are marked in bold; some of them are larger than 1 mutation):

13 24 14 11 11 14 12 12 12 13 13 29 – 17 9 10 11 11 25 15 19 29 15 15 17 17 –
11 11 19 23 16 **16** 18 17 **38 40** 12 12 – 11 9 15 16 8 10 10 8 **11** 10 12 23 23 16 10 12 12
14 8 12 22 20 13 **13** 11 13 **12 12** 12 12 (**L21-L193, 1275±170 ybp**)

One can notice that the first 25 markers in L193 do not have any mutation from L21; however, the rest of the 67 marker panel adds 10 mutations.

All 24 haplotypes contain 138 mutations from the above base haplotype, which gives 138/24/0.12 = 48 → 51 generations, that is 1275±170 years to a common ancestor.

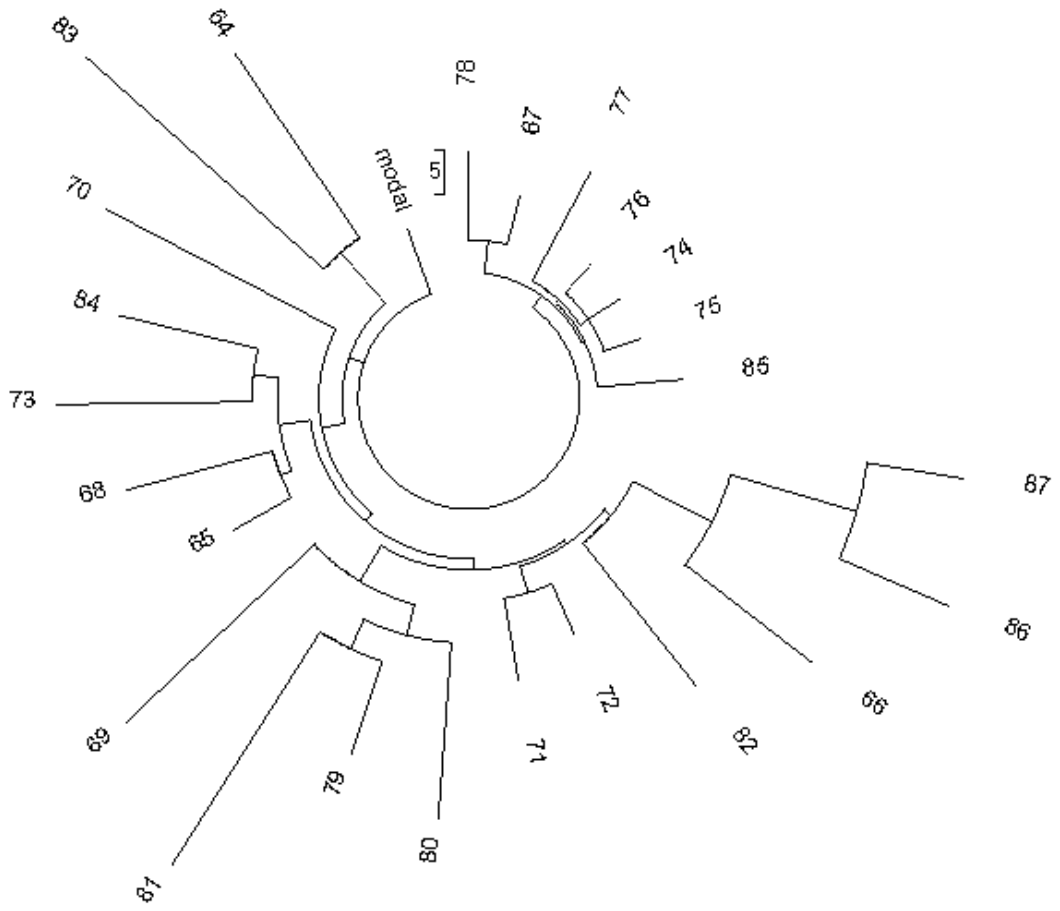


Fig. 7. 67 marker haplotype tree of 24 haplotypes of R1b1a2-L193 subclade

11 mutations between L193 and L21 base haplotype are equivalent to 2550 years between their common ancestors. This places THEIR common ancestor at approximately (2550+3750+1275)/2 = 3800 ybp. This is the L21 base haplotype (3775 ybp) within the margin of error.

L21-L226 ("Irish III" subclade)

This subclade has been studies in (Klyosov, 2010a,b,d,e), and the following base haplotype was identified.

13 24 14 11 11 14 12 12 11 13 13 29 – 17 8 9 11 11 25 15 19 29 13 13 15 17 –
11 11 19 23 15 15 18 17 36 38 12 12 – 11 9 15 16 8 10 10 8 10 10 12 23 23 15 10 12 12
15 8 12 22 20 13 12 11 13 11 11 12 12 (**L21-L226, 1500±170 ybp**)

A timespan to a common ancestor was determined as 1450±290 years (13 haplotypes), 1325±225 years (22 haplotypes), 1425±225 years (27 haplotypes), 1275±150 years (44 haplotypes), 1175±135 years (146 haplotypes).

Fig. 8 shows a haplotype tree for 50 haplotypes of subclade L226

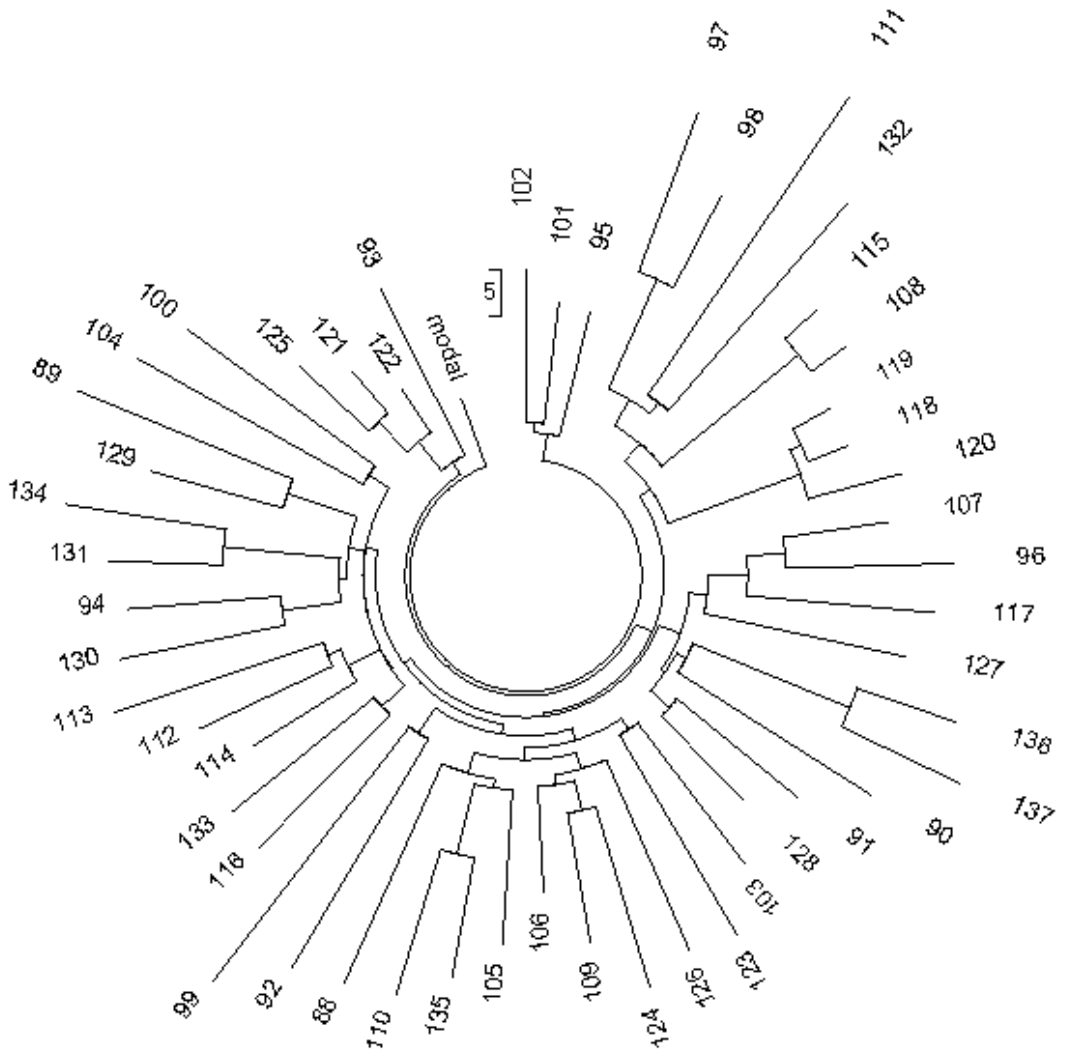


Fig. 8. 67 marker haplotype tree of 50 haplotypes of R1b1a2-L226 subclade

The tree has exactly the same 67 marker base haplotype as shown above, and has 335 mutations. It gives 335/50/0.12 = 56 → 60 generations, that is 1500±170 years from a common ancestor. It practically coincides with the earlier data, shown above, within margin of error. A slightly higher last figure is more correct, since it was calculated with an improved and recalibrated mutation rate constant for 67 marker haplotypes of 0.12 mutation/haplotype/generation compared with the reconsidered value of 0.145 mutation/haplotype/generation (Klyosov and Rozhanskii, 2010).

L21-P314.2

A haplotype tree of this subclade is shown in Fig. 9.

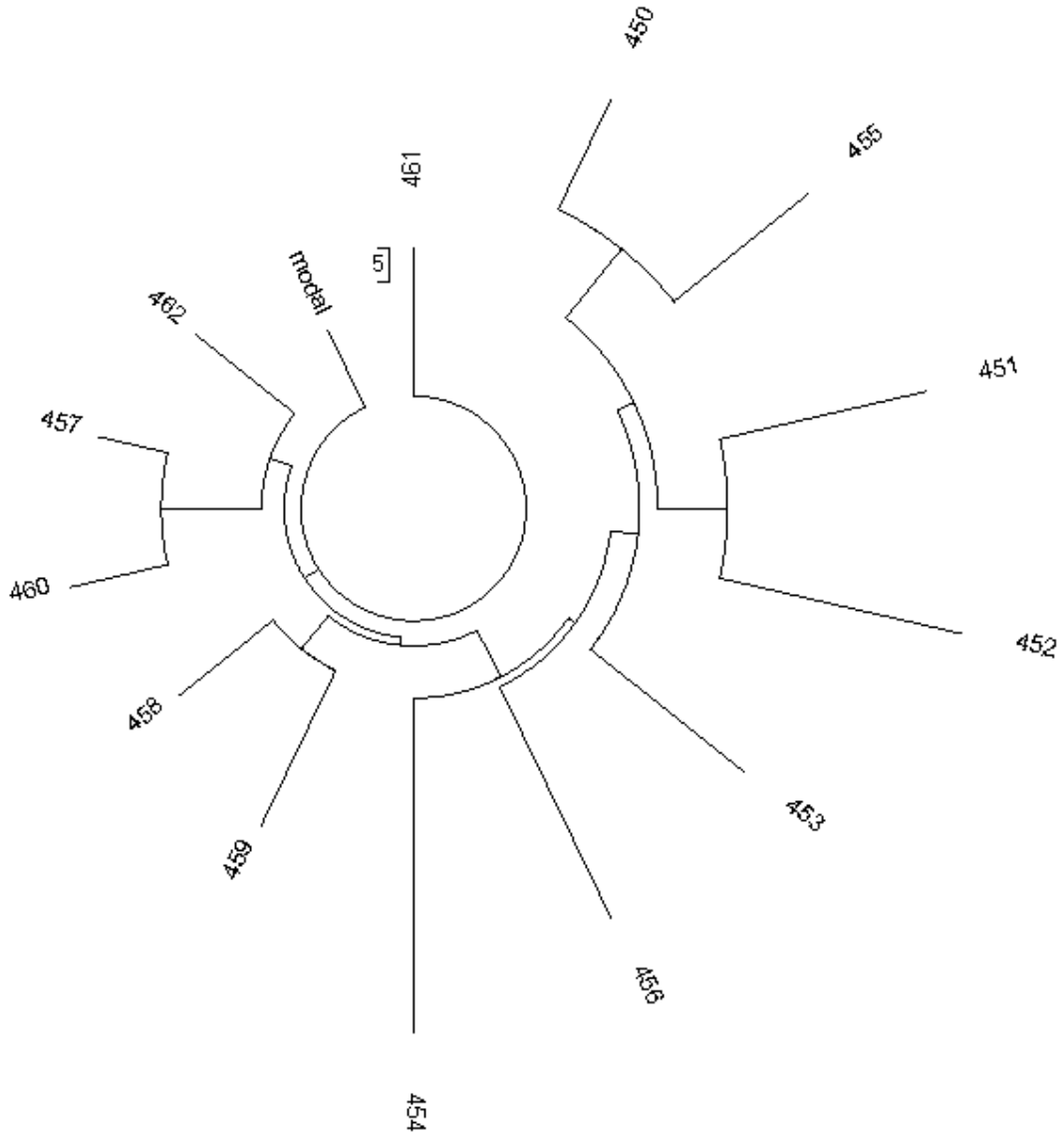


Fig. 9. 67 marker haplotype tree of 13 haplotypes of R1b1a2-L21-P314.2 subclade

Its base haplotype is as follows, 12 deviations from the L21 base haplotypes are marked in bold:

13 **23** 14 11 11 14 12 **13 13** 13 13 **28** – 17 9 10 11 11 **26** 15 19 29 15 15 **16** 17 –
10 11 19 23 16 15 18 **18 37** 38 12 12 – 11 9 **16** 16 8 10 10 8 **11** 10 12 23 23 **17** 10 12 12
15 8 12 22 20 13 12 11 13 11 11 12 12 (**P314.2, 2225±300 ybp**)

All 13 haplotypes contain 126 mutations from the above base haplotype, which gives 126/13/0.12 = 81 → 89 generations, that is 2225±300 years from a common ancestor of those 13 haplotypes. 12 mutations from the parent L21 base haplotype translate to 2775 years between both of their common ancestors, which results in approximately (2775+2225+3750)/2 = 4375 years; this is an ancestor of L21 subclade within margin of error.

One can notice that while calculated directly, using branches/subclades, an ancestor of the L21 subclade lived 3750±380 ybp, however, when calculated from downstream subclades, the "age" of the common ancestor of L21 is as follows:

3940 ybp, calculated with L21 null mutation haplotypes
4050 ybp, with L21-L226 subclade
4060 ybp, with L21-L159.2 subclade
4375 ybp, with L21-L314.2 subclade
4400 ybp, with L21-L144 subclade

It might reflect margins of error, however, it might point at an "older" age of the L21 subclade rather then determined directly from L21* haplotypes (see below).

R1b1a2-P312

Subclade P312 was studies in (Klyosov, 2009b, 2010a,f) and found that its base haplotype as follows

13 24 14 11 11 14 12 12 12 13 13 29 – 17 9 10 11 11 25 15 19 29 15 15 17 17 –
11 11 19 23 15 15 18 17 36 38 12 12 – 11 9 15 16 8 10 10 8 10 10 12 23 23 16 10 12 12
15 8 12 22 20 13 12 11 13 11 11 12 12 **(R1b1a2-P312, 3950±400 ybp)**

in which DYS456 is fluctuating between 15 and 16 depending on a dataset. With 464 P312 haplotypes (Klyosov, 2010a) it was close to 15; in this work, with 337 haplotypes, it is 15.72, that is close to 16. Apparently some local branches which always present in any haplotype dataset cause these (slight) variations.

The P312 haplotype tree is shown in Fig. 10, and its base haplotype for the whole tree is shown below.

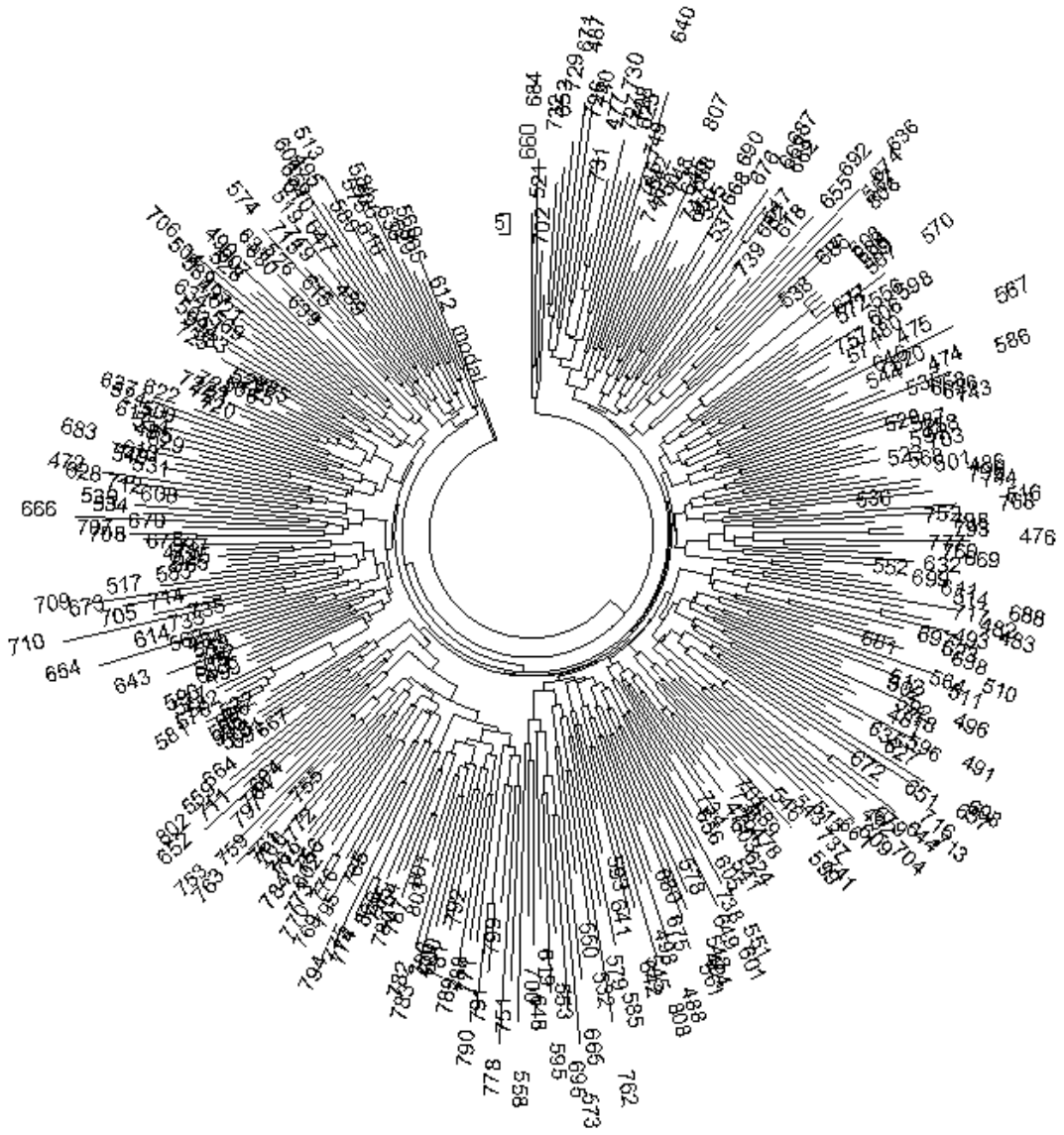


Fig. 10. 67 marker haplotype tree of 337 haplotypes of R1b1a2-P312 subclade. Six haplotypes in the dataset are marked as P312, they are scattered around the tree. 329 haplotypes are marked as P312*, and two (807 and 808) as P312. Three haplotypes have DYS425=0, they do not form a separate branch.**

13 24 14 11 11 14 12 12 12 13 13 29 – 17 9 10 11 11 25 15 19 29 15 15 17 17 –
11 11 19 23 16 15 18 17 36 38 12 12 – 11 9 15 16 8 10 10 8 10 10 12 23 23 16 10 12 12
15 8 12 22 20 13 12 11 13 11 11 12 12 **(R1b1a2-P312, variant, 3525±360 ybp)**

All 337 haplotypes contain 4956 mutations from the above base haplotype, which gives 4956/337/0.12 = 123 → 141 generations, that is 3525±360 years from a

common ancestor. The first 37 markers contain 3663 mutations, which gives 3663/337/0.09 = 121 → 138 generations, that is 3450±350 ybp; the first 25 markers contain 1981 mutations, which gives 1981/337/0.046 = 128 → 147 generations, that is 3675±380 ybp. One can see that those are practically the same figures (3525, 3450, 3675 ybp) within margin of error.

However, various datasets of P312 haplotypes can produce slightly different data, albeit within margins of error. For example, a 273 haplotype P312 dataset (October 2010) having exactly the same base haplotype as shown above for the 337 haplotype dataset, contained 1804 mutations in the first 25 markers, which gave 1804/273/0.046 = 144 → 168 generations, that is 4200±430 years to a common ancestor. That is why to consider only principal figures in TSCAs and ignoring margins of error might be misleading.

Apparently, the most accurate calculation of a timespan to a common ancestor of the P312 subclade was performed by a colleague of mine, Dr. Igor L. Rozhanskii. Using the logarithmic method, he subdivided P312 haplotypes in the 273 haplotype tree into several branches (they formed some yet unidentified subclades of P312) which showed the same TSCA by the logarithmic and the linear method. All those branches resulted in a common ancestor for P312 who lived 4350±700 ybp. Three "post-Iberian" downstream subclades, L21, U152 and SRY2627 gave 4000±500 ybp for their common ancestors. Finally, a common ancestor of the P312 and U106 subclades gave a common ancestor who lived 4800±700 ybp. This is the same figure which was shown above for a common ancestor of presumably L11 subclade in Europe, which also fits the time for the beginning of the Bell Beakers.

P-312 null mutation haplotypes

All 2299 67 marker haplotypes of R1b1a2-P312 and its subclade contained 67 of DYS425=0 haplotypes. Since that null mutation can be traced along a number of subclade branches starting from P312 down to (parallel)

- P312-U152-L2-L20,
- P312-L21-M222,
- P312-L21-M193, and
- P312-L176.2-SRY2627

some information can be obtained on P312 directions of migrations and their timing. Three null mutated P312* haplotypes were from Ireland, five of their downstream U152* were from Portugal, Hungary and Germany, two yet more downstream U152-L2-L20 were from Germany. This may provide a hint of a movement of descendant haplotypes from Iberia to North-East, to Germany.

Fifteen of P312-L21 null mutated haplotypes were all from Ireland and England, 28 of P312-L21* were from Ireland, England and Scotland, and four P312-L21** were from Spain, Ireland and Sweden. Both branches, L21-M222 and L21-L193, were represented with one null mutated haplotype each, both from Ireland. Finally, six null mutated P312-L176.2-SRY2627 were all from the Isles, namely from Ireland and England.

Fig. 11 shows the respective haplotype tree. It contains two principal branches, one at the bottom, which is obviously the older one, consisting of 25 haplotypes, and another a younger one, of 37 haplotypes.

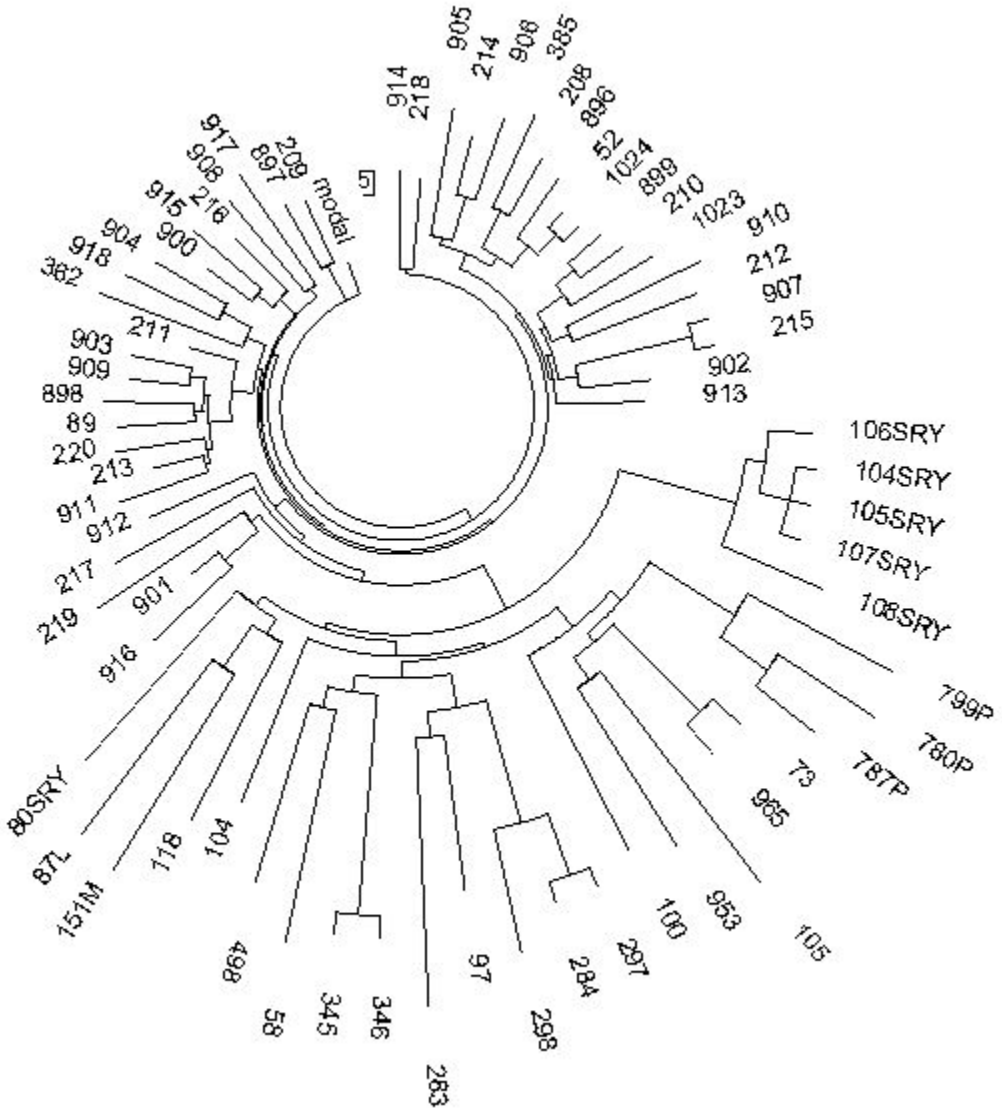


Fig. 11. 67 marker haplotype tree of null mutation (DYS425=0) 67 haplotypes (284 and 297 are identical, as well as 345 and 346) of R1b1a2-P312 and its subclades.

The older branch has the following base haplotype:

13 24 14 11 11 14 12 12 12 13 13 29 – 17 9 10 11 11 25 15 19 29 15 15 17 17 –
11 11 19 23 15 15 18 17 36 38 12 12 – 11 9 15 16 8 10 10 8 10 10 0 23 23 16 10 12 12 15
8 12 22 20 13 12 11 13 11 11 12 12 (**P312 null, sub-branch, 3575±400 ybp**)

It is exactly the R1b1a2-P312, 3950±400 ybp base haplotype, shown above, only with DYS425=0. The branch contains 374 mutations in all 25 haplotypes, which gives 374/25/0.12 = 125 → 143 generations, that is 3575±400 years from a common ancestor. It means that null mutated P312 haplotype which later became a common ancestor of this haplotype arose very close in time to its parent P312 ancestral haplotype.

The younger branch (in the upper part of the tree in Fig. 11) has the following base haplotype:

13 24 14 11 11 **15** 12 12 **13** 13 13 29 – 17 9 10 11 11 25 15 19 **28** 15 15 17 17 –
11 11 19 23 **16** 15 18 **18** 36 **37** 12 12 – 11 9 15 16 8 10 10 8 10 **9** 0 **22** 23 16 10 12 12 **16**
8 12 22 20 13 12 11 13 11 11 12 12

It deviates from the parent R1b1a2-P312, 3950±400 ybp base haplotype, by 8 mutations (marked above), besides the DYS425=0 mutation. All 37 haplotypes contain 234 mutations from the above base haplotypes, which gives 234/37/0.12 = 53 → 56 generations, that is 1400±170 years to a common ancestor. 8 mutations separates their common ancestors by 1800 years, which places THEIR common ancestor at (1800+1400+3950)/2 = 3575 ybp, which is exactly the common ancestor for the older branch. Therefore, the older null-mutated haplotypes is the parent one of the younger null mutated branch.

P312-U152

Subclade U152 was studied in (Klyosov, 2010a) as a 84 haplotype 67 marker dataset, and it was found that its base haplotype was as follows:

13 24 14 11 11 14 12 12 12 13 13 29 – 17 9 10 11 11 25 15 19 29 15 15 17 17 –
11 11 19 23 15 15 18 17 36 **39** 12 12 – 11 9 15 16 8 10 10 8 10 10 12 23 23 16 10 12 12
15 8 12 22 20 13 12 11 13 11 11 12 12 (**U152, 4125±450 ybp, 84 haplotypes)**

and its common ancestor lived 4125±450 ybp. Its base haplotype differs by only one mutation (marked above) compared to its ancestral P312 base haplotype.

In the same study 302 of 25 marker U152 were considered. They contained 1968 mutations, which gave 1968/302/0.046 = 142 → 166 generations, that is 4150±425 ybp.

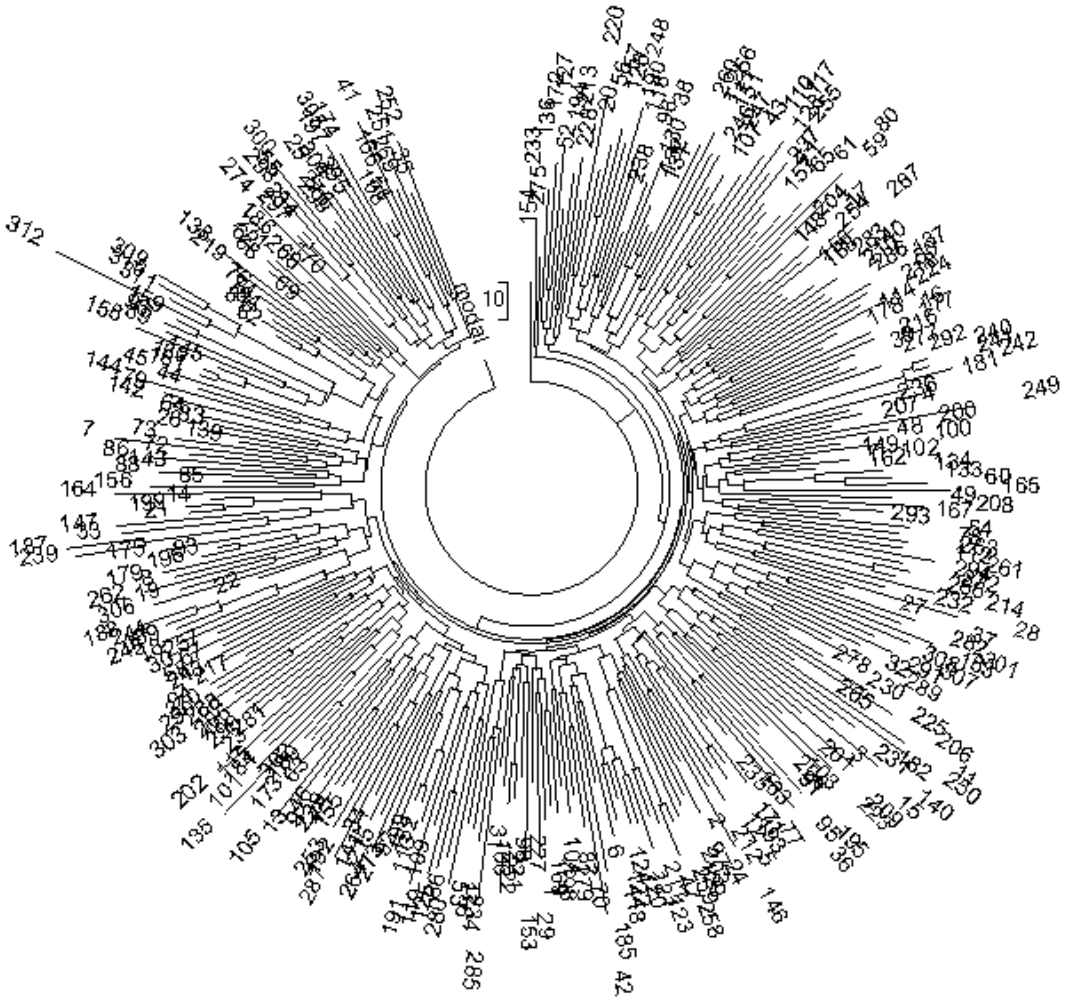


Fig. 12. 67 marker haplotype tree of 312 haplotypes of R1b1a2-P312-U152 subclade and downstream subclades. Subclade U152, haplotypes 1-162; subclade U152-L2, haplotypes 163-275; subclade U152-L20, haplotypes 276-308; subclade U152-L4, haplotypes 309-312.

A tree of 312 U152 haplotypes from the 2299 haplotype dataset is shown in Fig. 12. Its base haplotype is

13 24 14 11 11 14 12 12 12 13 13 29 – 17 9 10 11 11 25 15 19 29 15 15 17 17 –
11 11 19 23 15 15 18 17 36 **38** 12 12 – 11 9 15 16 8 10 10 8 10 10 12 23 23 16 10 12 12
15 8 12 22 20 13 12 11 13 11 11 12 12 **(U152, 3800±380 ybp, 312 haplotypes)**

that is exactly as that in P312 parent subclade. The CDYb allele, shown above in bold, equals to 38.02 as an average in all 312 alleles.

All 312 haplotypes have 4950 mutations from the above base haplotype, which gives 4950/312/0.12 = 132 → 152 generations, that is 3800±380 years to a common ancestor.

All 312 haplotypes (U152 and downstream subclades) contain 17 base haplotypes in the 12 marker format. It gives [ln(312/17)]/0.022 = 132 → 152 generations, that is 3800 years to a common ancestor. The identity of the results of the linear and logarithmic methods points at the one common ancestor for all the 312 haplotypes in the dataset.

162 haplotypes of only U152 subclade contain 9 base 12 marker haplotypes, which gives [ln(162/9)]/0.022 = 131 → 150 generations, that is 3750 years to a common ancestor. It means that removal of downstream subclades (L2, L20, and L4) from the 312 haplotype dataset, as well as removal of 8 base haplotypes which belong to downstream subclades does not change the calculated timespan to a common ancestor of U152 subclade.

U152-L2

113 haplotypes of this subclade have the following base haplotype:

13 24 14 11 11 14 12 12 12 13 13 29 – 17 9 10 11 11 25 15 19 29 15 15 **16** 17 –
11 11 19 23 **16** 15 18 17 36 38 12 12 – 11 9 15 16 8 10 10 8 10 10 12 23 23 16 10 12 12
15 8 12 22 20 13 12 11 13 11 11 12 12 (**L2, 4025±410 ybp, 113 haplotypes)**

All haplotypes contain 1867 mutations from the base haplotype, which gives 1867/113/0.12 = 138 → 161 generations, that is 4025±410 years from a common ancestor. The 1.5 mutations, separating the L2 and U152 base haplotypes (shown in bold; they all are fractional mutations), are equivalent to 300 years between their common ancestors, and place THEIR common ancestor to approximately (300+4125+4025)/2 = 4200 ybp. It is the U152 subclade itself, within margin of error.

In a total tree of 808 haplotypes of R1b1a2-P312 subclade, or 312 haplotypes of P312-U152 subclade subclade, L2 haplotypes are scattered around the whole tree. It means that the L2 subclade does not have a distinct branch. It was expected, however, since haplotypes P312, U152 and L2 subclades were practically identical, and their common ancestors (of the present-day haplotypes of these three subclades) lived practically in the same time. Therefore, except the SNP itself, nothing makes those haplotypes different at visual inspection or in their position on a haplotype tree.

The data obtained in this work fits well with a 95 haplotype 25 marker dataset. All haplotypes contained 631 mutations, which gave 631/95/0.046 = 144 → 168 generations, that is 4200±450 years from a common ancestor (Klyosov, 2010a).

U152-L2-L20

Haplotypes of this subclade were considered in (Klyosov, 2010a), however, only 15 of 37 marker haplotypes were available that time (end of 2009). They contained 175 mutations, which gave 175/15/0.09 = 130 → 149 generations, that is 3725±470 ybp.

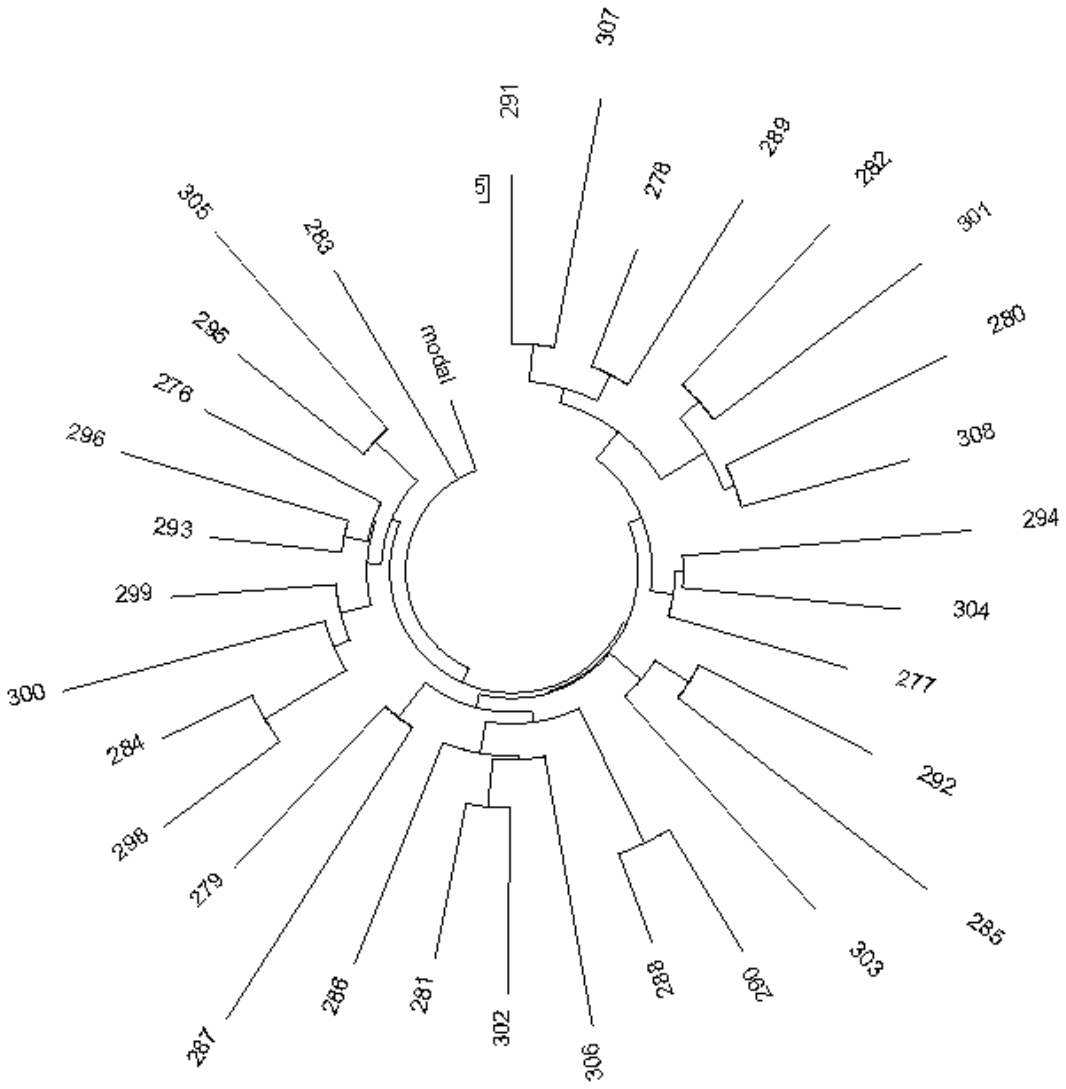


Fig. 13. 67 marker haplotype tree of 32 haplotypes of R1b1a2-P312-U152-L20 subclade.

An updated haplotype tree of 32 haplotypes in the 67 marker format is shown in Fig. 13. It results in the following base haplotype (deviations from that of the parent L2 base haplotype are marked):

13 24 14 11 11 14 12 12 12 13 13 29 – 17 9 10 11 11 25 15 19 29 15 15 16/**17** 17 –
10 11 19 23 **15** 15 18 17 36 **37** 12 12 – 11 9 15 16 8 10 10 8 10 10 12 23 23 16 10 12 12
15 8 12 22 20 13 12 11 13 11 11 12 12 **(L2-L20, 3650±400 ybp)**

All haplotypes contain 489 mutations from the base haplotype, which gives 489/32/0.12 = 127 → 146 generations, that is 3650±400 years from a common ancestor. What it looks like 4 mutations between the L20 and L2 base haplotypes, is in fact 1.74 mutations on average, since all of them are fractional ones. It is equivalent to only 360 years between their common ancestors, and places THEIR common ancestor to approximately (360+3650+4025)/2 = 4000 ybp. It is the U152-L2 subclade itself, within margin of error.

U152-L2-L20-L4

The whole dataset of 2299 haplotypes includes only four 67 marker haplotypes of this subclade, which form a distinct branch on the tree (Fig. 14).

Most (or all) of these four are Ashkenazi from Latvia, Poland and Ukraine. They have a base haplotypes as follows (19 mutations from the parent base L20 haplotype are marked; recLOH in DYS385a=14 was counted as one mutation):

13 24 14 11 **14** 14 12 12 12 13 13 29 – 17 9 10 11 **12** 25 15 19 **28** 15 **16 17** 17 –
12 12 19 23 15 15 18 **20** 36 **36 13** 12 – 11 9 15 16 8 10 10 8 10 **11** 12 23 **24** 16 10 12 12
15 8 **13** 22 20 **12** 12 11 13 11 11 **14** 12 **(L2-L20-L4, 1275±290 ybp)**

These four haplotypes contain 23 mutations from the base haplotype above, which gives 23/4/0.12 = 48 → 51 generations, that is 1275±290 years from a common ancestor.

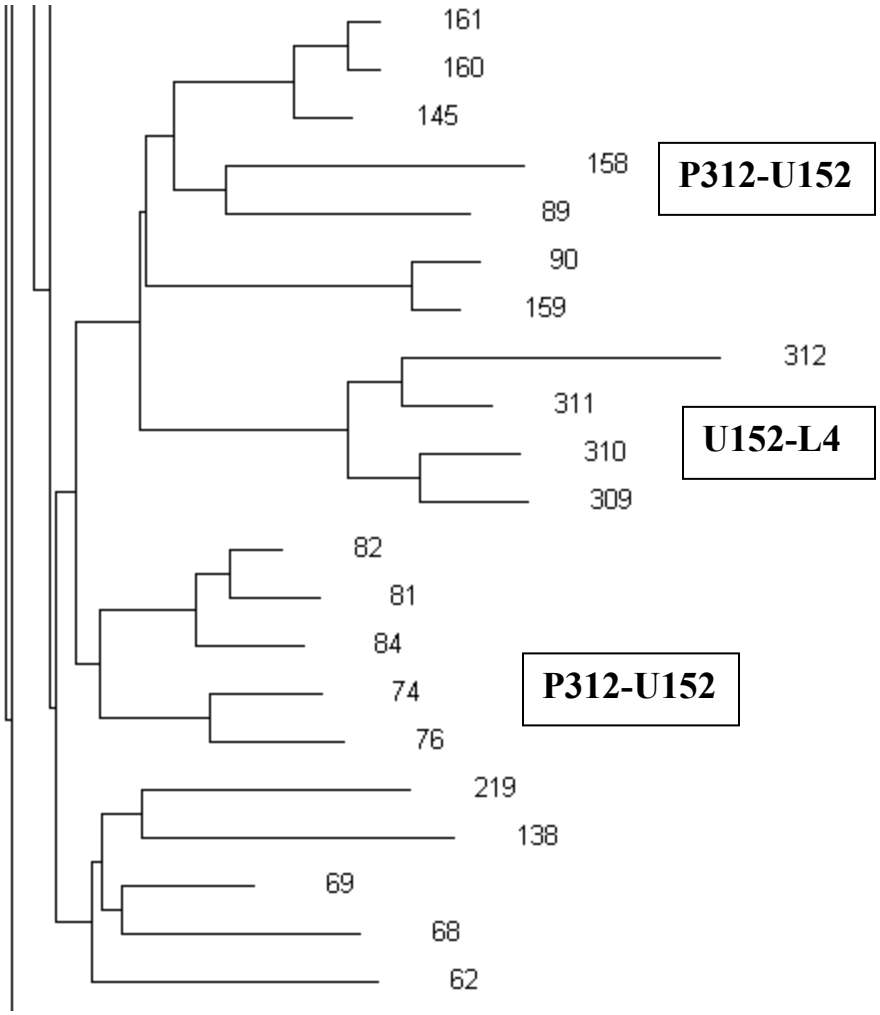


Fig. 14. A fragment of the 67 marker haplotype tree of 312 haplotypes of R1b1a2-P312-U152 with some subclades. Haplotypes 309-312 belong to subclade U152-L2-L20-L4, and form a separate branch (center). They are flanked by U152 haplotypes.

19 mutations between the L20 and L4 base haplotypes are equivalent to 4675 years between their common ancestors, and place THEIR common ancestor to approximately 4800 ybp. It is much "older" than that of the parent subclade, and can be explained by a poor statistics of the four-haplotype dataset.

P312-L176.2

There are only five haplotypes of this subclade among all the 2299 haplotypes of P312 dataset (see Fig. 15, haplotypes number 1 through 5). Among these five two haplotypes are from England, one from Ireland, one from France, and one from

Germany. The base haplotype is as follows (10 fractional deviations/mutations compared to the base P312 haplotypes are marked; they amount to 6.6 mutations):

13 24 14 11 11 14 12 12 12 13 13 **30** – 17 9 **9** 11 11 25 15 19 29 15 15 17 17 –
11 11 19 23 **16 14 19 16 37** 38 **13** 12 – 11 9 15 16 8 10 10 8 10 10 12 23 23 16 10 12 12
15 8 12 22 **21 14** 12 11 13 11 11 12 12 **(R1b1a2-P312-L176.2, 3675±560 ybp)**

Said five L176.2 haplotypes contain 77 mutations from the above base haplotype, which gives 77/5/0.12 = 128 → 147 generations, that is 3675±560 to their common ancestor.

6.6 mutations between P312 and L176.2 base haplotypes separate their common ancestors by 1450 years, and place THEIR common ancestor approximately at 4500 ybp. This is P312 itself within margin of error.

P312-L176.2-SRY2627

A tree of 155 haplotypes, which contains 141 haplotypes of SRY2627, five haplotypes of its upstream L176.2, and nine of its downstream L165 haplotype is shown in Fig. 15. All 141 haplotypes contain 1896 mutations from the base haplotype

13 24 14 11 11 14 12 12 12 13 13 **29** – 17 9 **10** 11 11 25 15 19 29 15 15 17 **18** –
11 11 19 23 16 **15 18 17** 37 38 **12** 12 – 11 9 15 16 8 10 10 8 10 10 12 23 23 16 10 12 **10**
16 8 12 22 **20 13** 12 11 13 11 11 12 12 (**P312-L176.2-SRY2627, 3150±320 ybp**)

which gives 1896/141/0.12 = 112 → 126 generations, that is 3150±320 years to a common ancestor.

Since most of the mutations between L176.2 and SRY2627 are fractional ones, there are not 12 of them, as marked above, but 8.2 mutations, which separate their common ancestors by 1825 years, and place THEIR common ancestor to approximately 4300 years before present. It is too high for that for L176.2, with its common ancestor of 3675±560 ybp, however, it might be within margin of error.

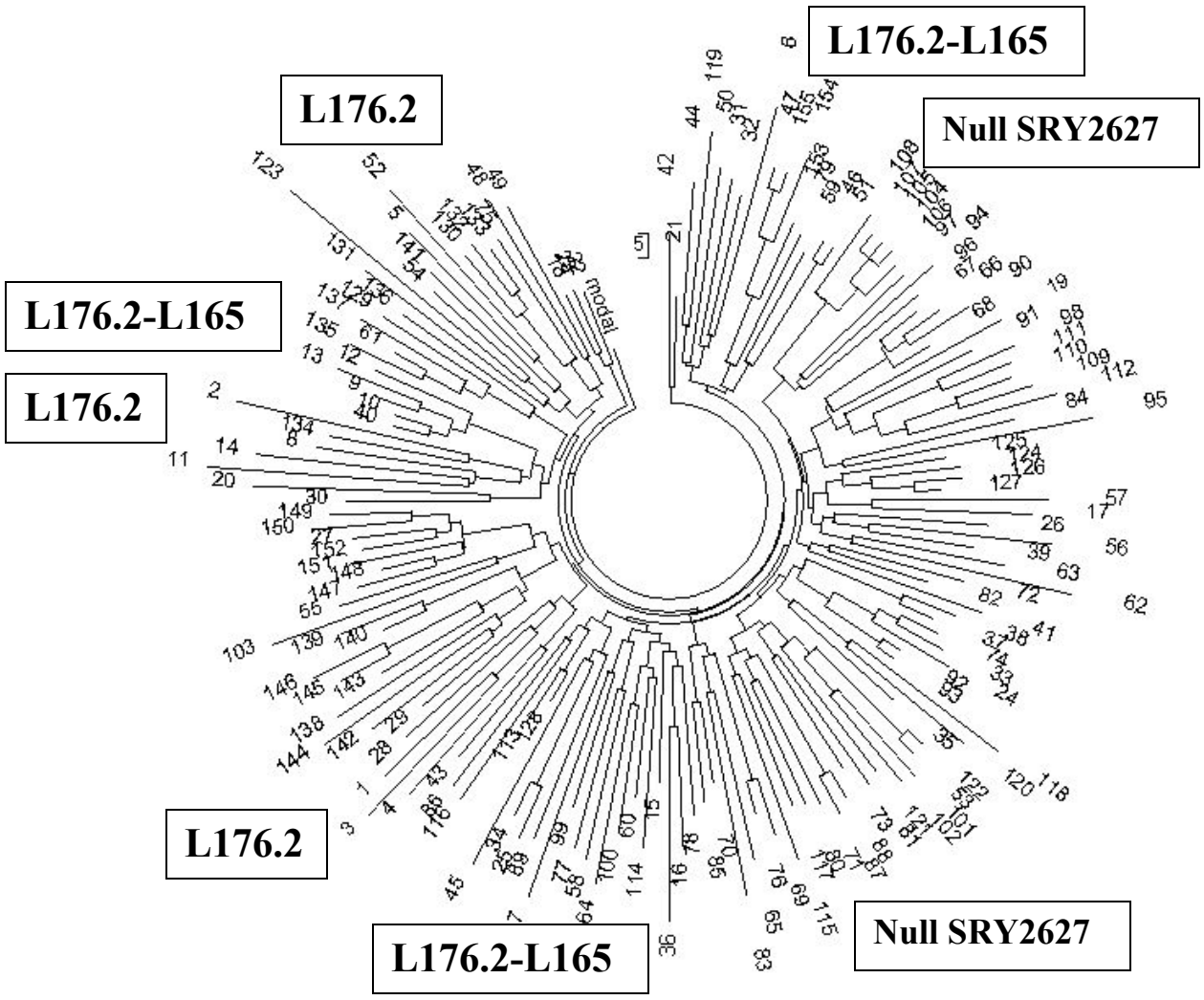


Fig. 15. 67 marker haplotype tree of 155 haplotypes of R1b1a2-P312-L176.2 (haplotypes 1-5) and its subclades L176.2-L165 (haplotypes 6-14) and L176.2-SRY2627 (haplotypes 15-155). There is a branch of null mutation SRY2627 haplotypes on the upper-right hand side (haplotypes number 104-108), and an isolated null mutation haplotype (number 80) on the lower right-hand side.

One can notice that the mutational difference between SRY2627 and its presumably parent L176.2 base haplotype is larger (12 "visible" mutations and 8.2 actual ones, equivalent to 1825 years between their common ancestors) compared the "grandfather" P312 base haplotype (6 "visible" mutations and 4.7 actual, which is equivalent to 1000 years between their common ancestors). On the other hand, the "ladder"of the subclades and their datings seems to be right: P312 (**3950 ybp**) → L176.2 (**3675 ybp**) → SRY2627 (**3150 ybp**). In other words, the L176.2 branch deviates (by their haplotypes) from both P312 and SRY2627, albeit should be between them. Something might be missing in the phylogeny of these

subclades. The answer what is missing is given below, after the section on the L165 subclade.

P312-L176.2-SRY2627 null mutation

There are only six haplotypes in this group among the 2299 haplotypes of the P312 dataset (five of them are located in the upper right-hand side in Fig. 15, haplotypes 104-108). Those five haplotypes have the following base haplotype

13 24 14 11 11 14 12 12 **11** 13 13 **30** – 17 9 10 11 11 25 15 19 **30** 15 15 **15 17** –
11 11 19 23 **15** 15 18 **18 38 39** 12 12 – 11 9 15 16 8 10 10 8 10 10 **0 22** 23 16 10 12 10 **19**
8 12 22 **21** 13 12 11 13 11 11 12 12 (**SRY2627, null, 325±120 ybp**)

in which mutations from the SRY2627 base haplotype are marked. Those five haplotypes have only 8 mutations from the above base haplotype, which gives 8/5/0.12 = 13 generations, or 325±120 from their common ancestor. These individuals are likely rather close relatives (from Ireland and England). A comparison with the remote null mutation haplotype (number 80)

13 24 14 11 11 **15** 12 12 **12 14** 13 31 – **19** 9 10 11 11 25 15 19 30 **14** 15 **17 18** –
11 11 19 23 **16 14** 18 **17 37 38** 12 12 – 11 9 15 16 8 10 10 8 10 10 0 **23** 23 16 10 12 10 **15**
8 **11 23 20** 13 12 11 13 11 11 12 12 (SRY2627, null, singular haplotype #80)

shows 21 mutations between them, which translates to 5300 years between their common ancestors, and places THEIR common ancestor to approximately (5300+325)/2 = 2800 ybp. This fits rather well with the "age" of SRY2627 of 3150±320 ybp.

P312-L176.2-L165

This is a subclade, "parallel" to SRY2627 subclade, both with the upstream P312-L176.2 subclade. There were only 9 haplotypes among the 2299 haplotype P312 dataset, and they were spread around the tree (Fig. 15) in three different branches. This is an indication that the subclade is old, maybe almost as old as the parent P312 subclade.

With haplotypes, spread around a tree in such a non-symmetrical manner (one branch of 7 haplotypes, and two single haplotypes in opposite position around the tree) it would be inaccurate to consider all nine haplotypes as randomly distributed and coming as such from one common ancestor. This common ancestor might be a phantom one, since the largest branch (with 7 haplotypes) would "pool" mutations to itself, resulting in a distorted base haplotype, hence,

distorted number of mutations. Instead, the system of 9 haplotypes should be treated as three independent "branches", two of them are single haplotypes.

The 7-haplotype branch has the following base haplotype

13 **25** 14 11 11 14 12 12 12 13 13 29 – 17 9 10 11 11 25 15 19 **30 14** 15 **15 16** –
11 11 19 23 **16** 15 18 17 **37** 38 12 12 – 11 9 15 16 8 10 10 8 10 10 12 23 23 16 10 12 12
15 8 12 22 20 **14** 12 11 13 11 11 12 12 **(L165, 7-hapl. branch, 2325±360 ybp)**

with 71 mutations from it. It gives 71/7/0.12 = 85 → 93 generations, that is 2325±360 years from its common ancestor. The apparent 8 mutations between L153 and its parent P312 base haplotypes (marked above) is fact amount to 8.72 mutations (two of them are 1.43 and 1.29 on average). It translates to 1925 years between their common ancestors, and places THEIR common ancestor to (1925+2325+3950)/2 = 4100 ybp. This is indeed a common ancestor of the P312 subclade within margin of error.

However, this is only one sub-branch of L165 on the haplotype tree (Fig. 15). When we add two singular haplotypes (numbers 6 and 7 on the tree, Fig. 15)

13 25 14 11 11 14 12 12 **13** 13 13 29 – **18** 9 10 11 11 25 15 19 **31** 14 15 **17 17** –
11 **10** 19 23 **17** 15 **19 16 38** 38 12 12 – 11 9 15 16 8 10 10 8 10 10 **8** 23 **24** 16 10 12 12
15 8 **13** 22 20 14 **13** 11 13 11 11 12 12 (L165, Scotland, singular haplotype #6)

13 **24** 14 **10 14** 14 12 12 **13 14** 13 30 – **18** 9 10 11 11 **24** 15 19 29 **15** 15 **16 17** –
11 11 19 23 16 15 **17 18 38 39 11** 12 – 11 9 15 16 8 **11** 10 8 10 10 12 23 **24** 16 10 12 12
15 8 12 22 20 **13** 12 11 **14** 11 11 12 12 (L165, England, singular haplotype #7)

they three have 29 mutations from their presumed base haplotype. This places THEIR common ancestor to approximately 3000 ybp. His (base) haplotype was apparently (minimized on mutations between all the three base haplotypes)

13 **25** 14 11 11 14 12 12 **13** 13 13 29 – **18** 9 10 11 11 25 15 19 **30 14** 15 **16** 17 –
11 11 19 23 **16** 15 18 17 **38** 38 12 12 – 11 9 15 16 8 10 10 8 10 10 12 23 **24** 16 10 12 12
15 8 12 22 20 **14** 12 11 13 11 11 12 12 (**L165, ~ 3000 ybp**)

As it was expected, this figure, 3000 ybp, obtained by comparing haplotypes around the tree, is (slightly in this particular case) higher than calculated from all 9 haplotypes. They have 105 mutations, which gives 105/9/0.12 = 97 → 108 generations, that is a (slightly lower) value of 2700±380 years from a common ancestor.

What might be missing in the phylogeny of P312?

As it was mentioned above, and not once, calculations in the L176.2 subclade along with its downstream subclades SRY2627 and L165 typically produce TSCA values higher than it is expected. It seems that a root of these subclades is missing, since the P312 subclade, which supposed to be their root, has a lower TSCA (around 4000 ybp) that it should have had. It is not a surprise, though, since it was already calculated above that P312 along with U106 have a common ancestor who lived ~ 4800 ybp. Now let us check if P312 itself could arose that time, and not around 4000 ybp.

When we consider all three base haplotypes which hint that their common ancestor could have lived earlier than 3700-4100 ybp, we find that all of them have 20 mutations (marked below) from a deduced base haplotype.

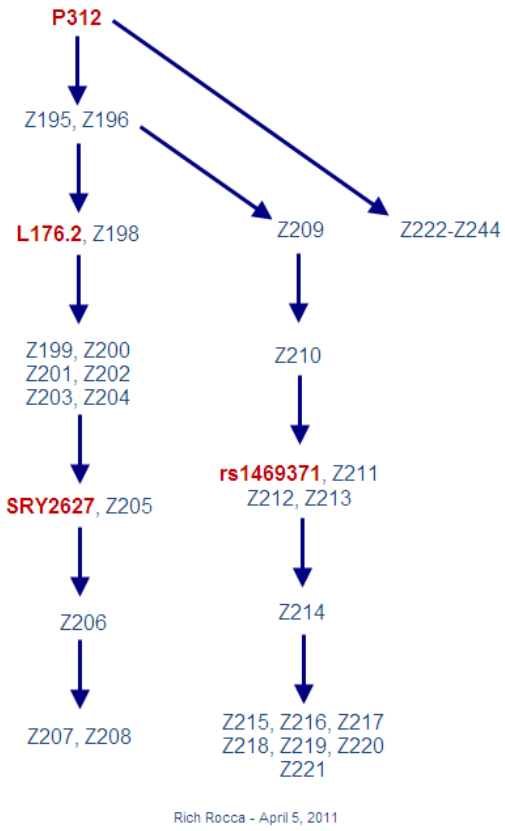
13 24 14 11 11 14 12 12 12 13 13 **30** – 17 9 **9** 11 11 25 15 19 29 15 15 17 17 –
11 11 19 23 16 **14 19 16** 37 38 **13** 12 – 11 9 15 16 8 10 10 8 10 10 12 23 23 16 10 12 12
15 8 12 22 **21** 14 12 11 13 11 11 12 12 (**R1b1a2-P312-L176.2, 3675±560 ybp**)

13 24 14 11 11 14 12 12 12 13 13 29 – 17 9 10 11 11 25 15 19 29 15 15 17 **18** –
11 11 19 23 16 15 18 17 37 38 12 12 – 11 9 15 16 8 10 10 8 10 10 12 23 23 16 10 12 **10**
16 8 12 22 20 **13** 12 11 13 11 11 12 12 **(P312-L176.2-SRY2627, 3150±320 ybp**)

13 **25** 14 11 11 14 12 12 **13** 13 13 29 – **18** 9 10 11 11 25 15 19 **30 14** 15 **16** 17 –
11 11 19 23 16 15 18 17 **38** 38 12 12 – 11 9 15 16 8 10 10 8 10 10 12 23 **24** 16 10 12 12
15 8 12 22 20 14 12 11 13 11 11 12 12 (**L165, ~ 3000 ybp**)

20 mutations in three base haplotypes translate into 20/3/0.12 = 1500 years plus their average "age", which is 3275 years, which gives 4775 years from a common ancestor of all the three subclades. It supposed to be either the common ancestor of the P312 subclade himself, or an immediate P312 downstream, unidentified as yet. Again, it is the beginning of Bell Beaker times in Iberia.

Some details to these data can be found by considering the 1000 Genomes Project (http://dna-forums.org/index.php?/topic/14907-two-potentially-large-p312-snps/). Rich Rocca has reported that considering samples from Iberia and former Spanish colonies in Colombia, Mexico and Peru he was able to identify several new candidate SNPs (all indexed starting with Z) downstream of P312 and upstream of L176.2 (see chart below). All these new candidate SNPs were found to be negative in P310, U152, L21 and U106 (Rich Rocca, ibid.).



From: **http://dna-forums.org/index.php?/topic/14907-two-potentially-large-p312-snps/**

In fact, all these new findings are unrelated to TSCA calculations in this paper, and serve – in this context - no more than an illustration of a complexity of phylogeny of R1b1ba2, though, as well as any other haplogroup.

P312-L238

This subclade contains only two haplotypes among all 2299 haplotypes in the dataset, one from England and one from Sweden. They are very similar with each other, and sit next to each other on the haplotype tree on the same small branch (Fig. 4):

13 24 14 11 11 13 12 12 11 13 13 29 – 16 9 10 11 11 25 15 19 29 15 15 17 17 –
11 11 19 23 17 15 20 18 37 37 12 12 – 11 9 15 16 8 10 10 8 10 10 12 23 23 16 10 12 12
17 8 12 21 20 15 12 11 13 11 11 12 12 **(P312-L238, singular, #463)**

13 24 14 11 11 13 12 12 11 13 13 29 – **17** 9 10 11 11 25 15 19 **28** 15 15 17 17 –
11 **12** 19 23 **15** 15 **19** **17** 37 **38** 12 12 – 11 9 15 16 8 10 10 8 **11** 10 12 23 23 16 10 12 12
17 8 12 21 20 15 12 11 13 11 11 12 12 **(P312-L238, singular, #464)**

Nine mutations between them place their common ancestor to approximately 1000 ybp.

Geographical distribution of R1b1a2-P312 subclades

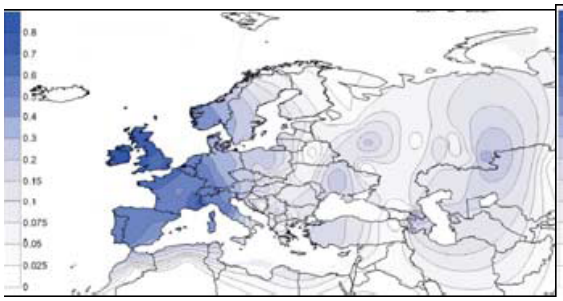
The main challenge in getting anything significant from a regional distribution of P312 and/or U106 subclades and timespans to their common ancestors is in the fact that all of them are often alike, and if not, they might reflect more recent movements, mixing, migrations, etc. The Gallic wars of Julius Caesar, for example, has changed the population landscape a great deal more than two millennia ago. There were other major population-related event such as black plague of the 14th century, almost permanent wars in Europe since those times (and before), recent world wars, etc. As a result, the geographical spread of P312 and U106 in Europe is far-flung and uneven, and hardly reflects their distribution 4000-3000 years before present.

Let us see what can be observed from geography of R1b1a2-P312 distribution, if anything. It would be helpful if we put forward a general hypothesis based on several observations, and see if data on regional distributions of P312 and its subclades are compatible with that hypothesis.

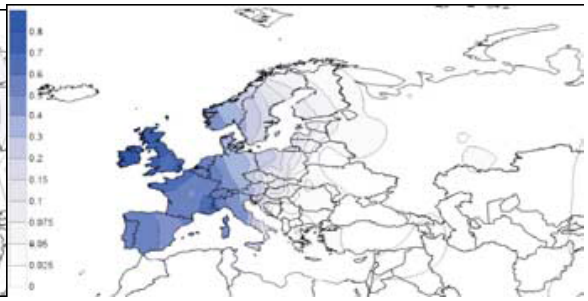
This hypothesis was in fact advanced earlier and briefly described in the beginning of this paper (Klyosov, 2008-2011, see references there). According to it, an upstream subclade of P312 and U106 (likely R1b1a2-L11/L151/P310/P311, along with L51/M412) has arrived to Iberia ~ 4800 years before present, split P312 and U106, which moved up North to continental Europe as Bell Beakers, and spread over Europe between 4500 and 3200 ybp. It seems that both P312 and/or U106 went through a severe population bottleneck in the 3rd millennium BC, which (the bottleneck) had lasted almost a thousand years. Therefore common ancestor of current bearers of P312 and U106 can be detected only around 3700-4100 ybp. Only superposition of the their respective base haplotypes along with their TSCAs made possible to calculate that their common ancestor (presumably R1b1a2-L11, or P312 itself) lived 4800 ybp, at the very beginning of Bell Beaker movements.

This part of the hypothesis is supported by archaeological data, according to which the oldest artifacts related to the Bell Beakers were found in the Pyrenees and dated by 2900-2500 BC (Muller et al, 2001).

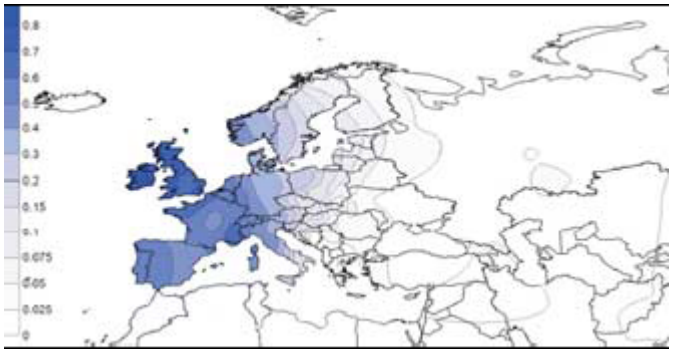
The following maps, published by Myres et al (2010), provide additional support to this hypothesis. One can see that all upstream (of P312 and U106) subclades – **M269 → L51 → L11** have the highest density in Iberia.



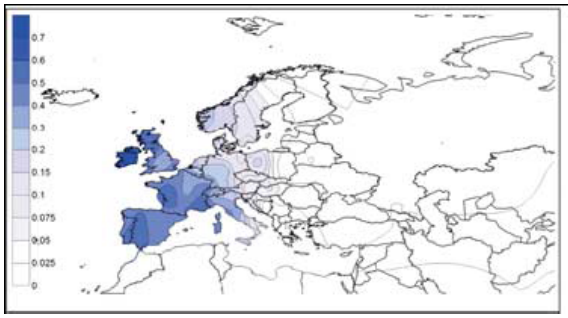
The current distribution of the **M269** subclade (with downstream subclades) in Europe (Myres et al, 2010)



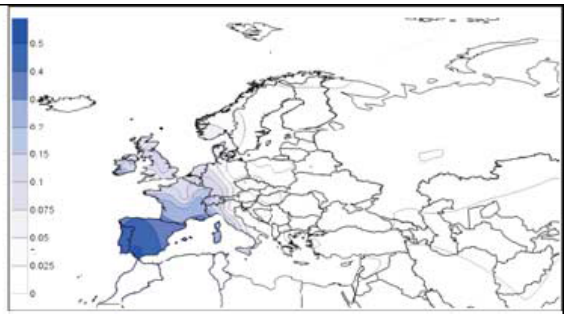
The current distribution of the **L51/ M412** subclade (with downstream subclades) in Europe (Myres et al, 2010)



The current distribution of the **L11** subclade (with subclades) in Europe (Myres et al, 2010)



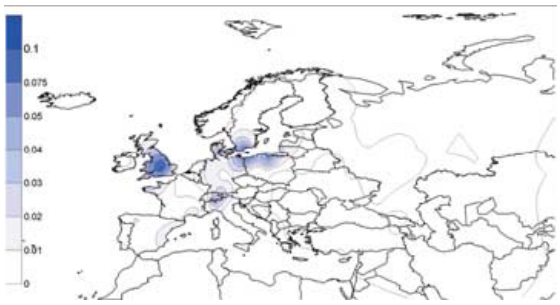
The current distribution of the **P312** (with subclades) in Europe (Myres et al, 2010)



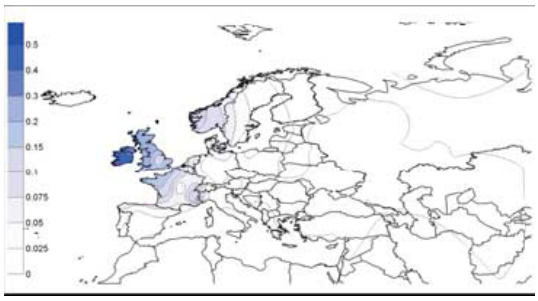
The current distribution of the **P312*** subclade in Europe (Myres et al, 2010)

The most significant data here is that the P312* subclade, separated from its downstream subclades, still has the highest presence in the Pyrenees. Subclade M153, which is the immediate subclade of P312, is presented only in Spain among all 2299 haplotypes in the dataset. It has a common ancestor of ~ 3640 ybp. As many as 37 haplotypes of this M153 subclade were found among 750 R1b haplotypes in the Pyrenees (Adams et al, 2009), and the Basques had it as 15 haplotypes from their 37, the highest amount among all other R1b subclades (Klyosov, 2010c).

The Bell Beakers apparently moved in two principal directions – to the continental Europe, and directly to British Isles, apparently by sea. The last direction is supported by the following two maps:

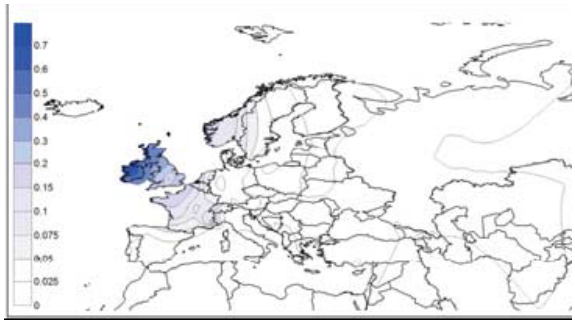


The current distribution of the **L11** (xU106, P312), that is **L11*** subclade In Europe (Myres et al, 2010)



The current distribution of the **L21** (xM222) subclade in Europe (Myres et al, 2010)

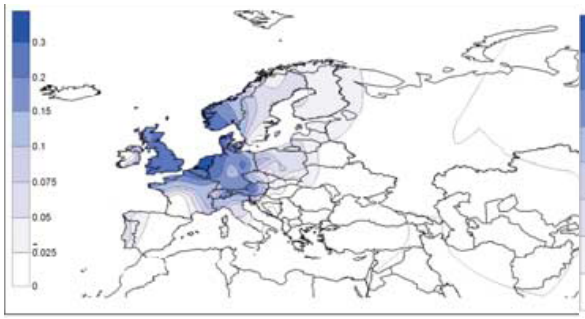
One can see that the L11* subclade, the ancient and upstream one with respect to P312, is presented on the Isles, when its downstream subclades are removed. Otherwise it has a weak presence on the continent. In the same manner, L21, an immediate downstream subclade of P312, has the highest presence in Ireland, when its downstream M222 is removed. On the other hand, bearers of L21 subclade could reach the Isles by land, from the Pyrenees via South-Western part of Europe (now France) and to the Isles:



The current distribution of the **L21** subclade in Europe (Myres et al, 2010)
Subclades P312, U106, and their immediate subclades U152, L21, and L176.2

considered in this study) all coalesce to their common ancestors who all lived between 4100 and 3700 ybp (see the chart in the beginning of this article). Besides, common ancestors of the immediate subclades of U152, namely L2 and L20, also lived within the same time span (4025 and 3650 ybp on average, respectively). P312 null mutation subclade (3575 ybp) and U152 null mutation subclade (3525 ybp) also lived at the same time or close to it.

It seems that U106 and U152 either moved out as Bell Beakers from the Pyrenees, or arose in the continental Europe, after the Beakers left the peninsula. They both practically do not show their presence in Iberia nowadays (see the maps below).



The current distribution of the **U106** subclade (with subclades) in Europe (Myres et al, 2010)

The current distribution of the **U152** subclade in Europe (Myres et al, 2010)

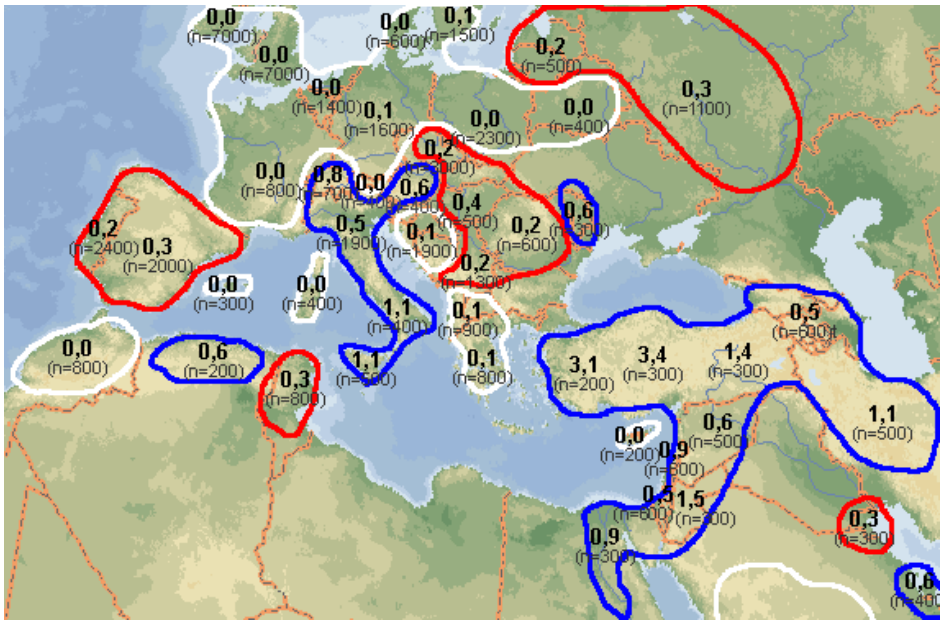
It seems that all other regions in Europe, such as nowadays Scandinavia, North-Western Europe, Germany, Poland, Central Europe have been populated by R1b1a2-P312, U106 and their subclades along the Bell Beakers migration route, between 4500 and 3200 ybp. However, since the migrants carried their haplotypes across Europe without principal population bottlenecks, TSCAs are just about the same across Europe, namely between 4500 and 3200 ybp.

Archaeological data in general supports this pattern of DNA genealogy. According to them (see map below, from www.buildinghistory.org/distantpast/bellbeaker.shtml, with a reference to Muller et al, 2001), the oldest Bell Beaker sites were found in the Pyrenees (4900-4500 ybp) and nearby (France - North Italy), and datings of Bell Beaker sites shifts to slightly later times in Central Europe, Germany and Scandinavia.

In principle, archaeology chronological data show that the development of Bell Beakers took place from the west (more specifically from the southwest) toward the east and northeast. The same data follows from DNA genealogy, as it is shown in this article.

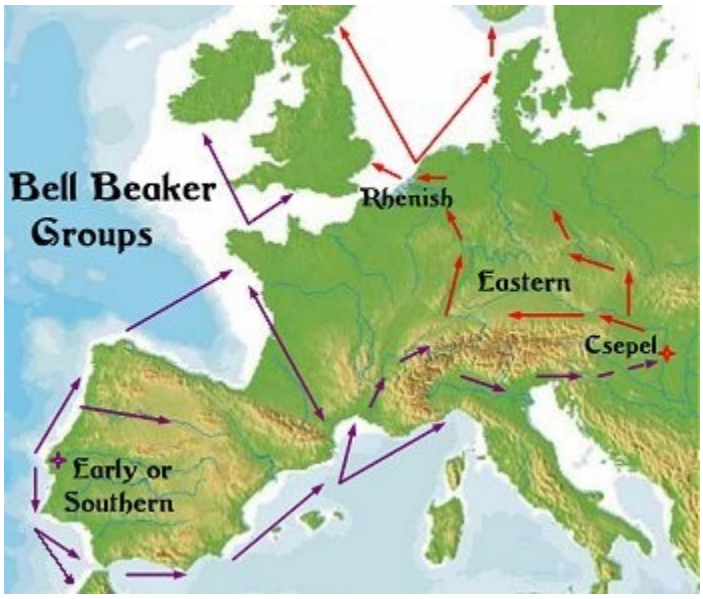
The South-Eastern part of Europe, including the Balkans, was a principal destination of yet another migration wave of R1b1a2, but largely of L23 subclade, with a particular feature of DYS393=12. This allele is typical for the Russians, R1b1a2 inhabitants of the Caucasus (Armenians, Dagestanis and others), Anatolia, Middle East, Asia Minor. This is apparently an immediate source of the L23 subclade in Europe. There is almost none of that subclade in Iberia, it apparently did not survive the migration route from Middle East across North Africa into Pyrenees (see above). Indeed, the map above does not show Bell Beaker sites in South-East Europe and Asia Minor, and East of Germany (except in Northern Poland in later times). Some traces of Bell Beakers on the Balkans are attributed to Bronze Age, significantly later times.

The following map, taken from the World Families Forums (R-L277+)
http://www.worldfamilies.net/forum/index.php?topic=9562.25
shows a possible route of R1b1a2-M269-L23-L277 across Asia Minor to Mediterranean and the Balkans.



One can see that the highest amount of that subclade is in Asia Minor (up to 3.4%), in South Italy up to 1.1%, on the Balkans in the North Italy up to 0.4-0.6-0.8%, and close to zero in Central Europe, in the Isles, in Northern Europe.

The next map, from the same source as quoted above, shows some alleged ways how R1b1a2-P312 could have reached the Isles by sea.



The above statement that before arrival to the Iberian peninsula R1b1a2 bearers went across North African shore from Middle East to the Atlantic, is supported

by archaeological data that the area with Bell Beaker sites takes also parts of northern Africa in Algeria and Morocco
(http://www.novelguide.com/a/discover/aneu_01/aneu_01_00103.html)

Now, let us take a look at haplotypes of the R312-L21 subclade across Europe. Much of the data arranged by the regions and considered below were provided by Mark Jost. He expressed a concern that the 67 marker panel includes fast mutated markers as well as markers subjected to recLOH mutations, and suggested to employ a 49 marker panel, in which 18 markers are removed (see above, the Methodology). Since it is a wide-spread (and not justified, as it is shown below) concern, this work also shows that is does not matter which haplotype panel to employ. A panel should be properly calibrated, this is the only one thing that matters.

Ireland, L21

Let us start with the most representative population in this dataset, Ireland, which contains 153 haplotypes. We will first compare 49 and 67 marker panels (Fig. 16 and 17). In both of them the haplotype tree looks rather symmetrical and "smooth", with a good likelihood of being derived from one common ancestor. Local small branches always present in any haplotype tree, and they should not cause any concern unless they form a separate distinct branch.

The 49 marker L21 Irish 153 haplotype dataset has the following base haplotype:

13 24 14 11 12 12 12 13 13 – 17 11 11 25 15 19 29 – 11 11 16 15 18 17 12 12 – 11 9 8 10 10 8 10 10 16 10 12 12 15 8 12 22 20 13 12 11 13 11 11 12 12
(**Ireland, L21, 49-marker panel, 3650±380 ybp**)

In fact, it is exactly the same as that of the "overall" L21 1024-haplotype dataset, considered above, except said 18 markers are eliminated.

All 153 haplotypes contained 1550 mutations, which gives 1550/153/0.08 = 127 → 146 generations, that is 3650±380 years to a common ancestor. 0.08 here is the mutation rate constant measured in mutations per haplotype per generation (25 years, see explanations above). The "overall" L21 dataset of 1024 of 67 marker haplotypes has a common ancestor who lived 3750±380 ybp. They are practically the same figures, with a difference of only 2.7% between them, well within the margin of error.

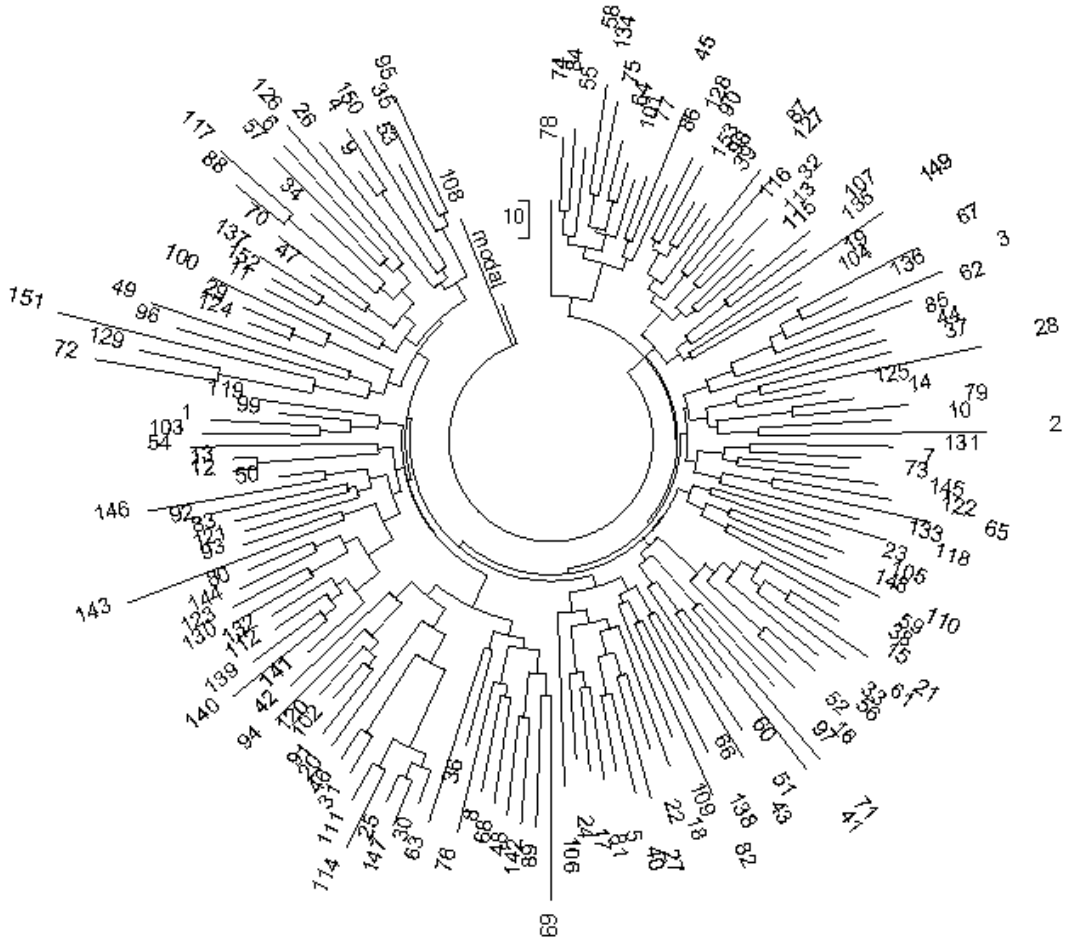


Fig. 16. 49 marker haplotype tree of 153 Irish R1b1a2-L21 haplotypes

In the 67 marker format the Irish 153 haplotype L21 dataset reveals the following base haplotype:

13 24 14 11 11 14 12 12 12 13 13 29 – 17 9 10 11 11 25 15 19 29 15 15 17 17 –
11 11 19 23 16 15 18 17 36 38 12 12 – 11 9 15 16 8 10 10 8 10 10 12 23 23 16 10 12 12
15 8 12 22 20 13 12 11 13 11 11 12 12
(**Ireland, L21, 67-marker panel, 3575±365 ybp**)

It is exactly as that deduced from 1024 L21 haplotypes, with a common ancestor of 3750±380 ybp (see above).

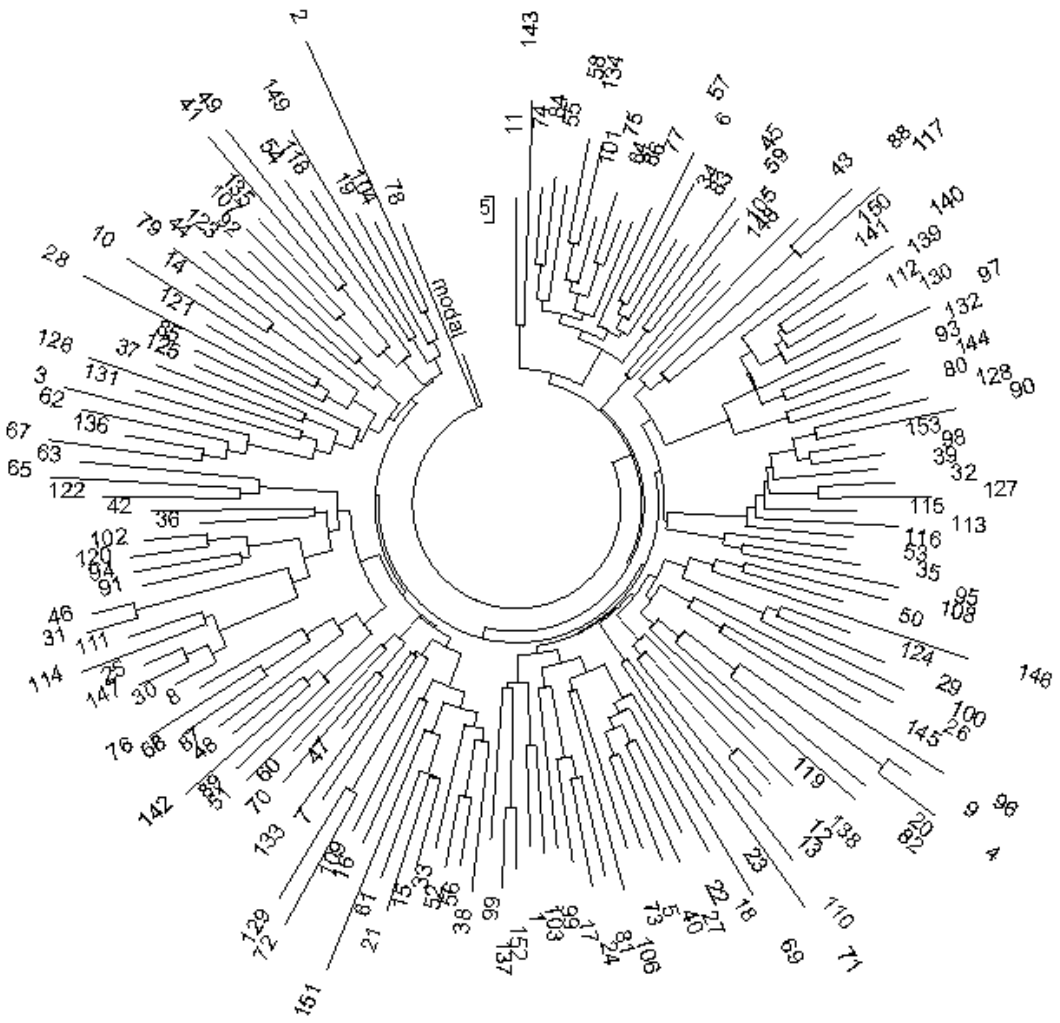


Fig. 17. 67 marker haplotype tree of 153 Irish R1b1a2-L21 haplotypes

The missing 18 ("fast and palindromic") markers added in the 67 marker panel (compared to those in the 49 marker haplotypes) 746 mutations, making it 2296 mutations. It gives 2296/153/0.12 = 125 → 143 generations, that is 3575±365 years to a common ancestor. It is again practically the same figure as that for 49 marker haplotypes and for the "overall" L21 dataset of 1024 haplotypes. An average between all the three figures results in 3658±88 ybp, that is 2.4% variation. In other words, ancient bearers of the L21 subclade "brought a timespan to their common ancestor" to Ireland without any noticeable change.

It should be noted that the 153-haplotype series contains 8 of 12-marker base haplotypes, which gives [ln(153/8)]/0.022 = 134 → 155 generations, that is ~ 3875 years to a common ancestor. One can see that the logarithmic method gave

results within margin of error with that by the linear method, 3575±365 ybp. It shows that the dataset is "smooth" indeed, and is derived from one common ancestor.

England, L21

Another example – England. The L21 dataset contains 74 haplotypes (Fig. 18). In the 49 marker format it has the following base haplotype:

13 24 14 11 12 12 12 13 13 – 17 11 11 25 15 19 **30** – 11 11 16 15 18 17 12 12 – 11 9 8 10 10 8 10 10 16 10 12 12 15 8 12 22 20 13 12 11 13 11 11 12 12
(England, L21, 49-marker panel, 3950±420 ybp)

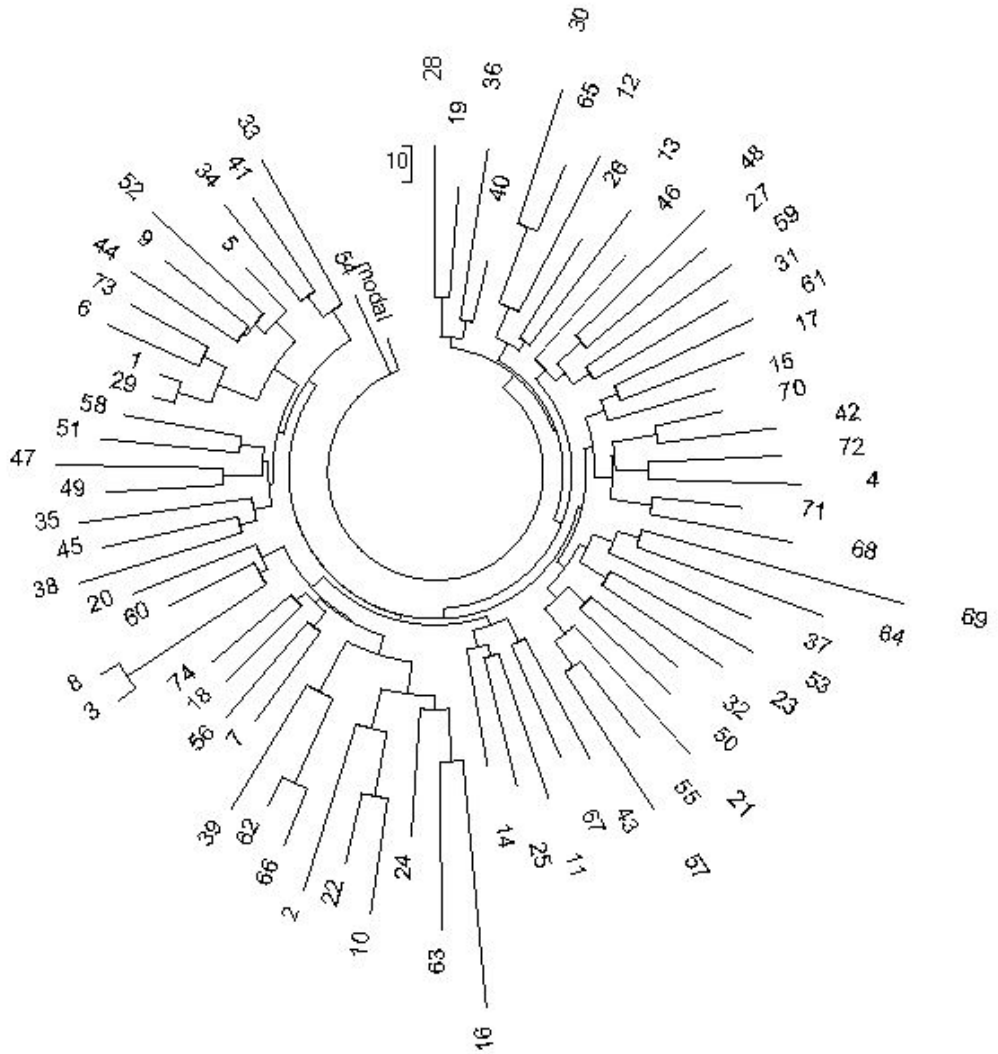


Fig. 18. 49 marker haplotype tree of 74 English R1b1a2-L21 haplotypes

The only difference in DYS449 (marked) is actually not a real difference: it is 29.42 in the Irish haplotype, hence, rounded to 29, and it is 29.62 in the English haplotype, therefore, rounded to 30. In fact, the difference is 0.2 mutations. In all 1024 L21 haplotypes the average allele there equals to 29.50.

All 74 of 49 marker haplotypes contain 805 mutations from the above haplotype, which gives 805/74/0.08 = 136 → 158 generations, that is 3950±420 years to a common ancestor.

The 67 marker dataset adds 424 mutations, making it 1229/74/0.12 = 138 → 161 generations, that is 4025±420 ybp. The difference between 49- and 67-marker panels is less than 2%.

Fig. 19. 49 marker haplotype tree of 78 Scotland R1b1a2-L21 haplotypes

Scotland, L21

The L21 Scottish dataset contains 78 haplotypes (Fig. 19). In the 49 marker format it has the following base haplotype:

13 24 14 **10** 12 12 12 13 13 – 17 11 11 25 15 19 **30** – 11 11 16 15 18 17 12 12 – 11 9 8 10 10 8 10 10 16 10 12 12 15 8 12 22 20 13 12 11 13 11 11 12 12
(**Scotland, L21, 49-marker panel, 3325±350 ybp)**

Again, the deviations (marked above) are minimal ones between the Irish, English and Scottish base haplotypes. In DYS391 the respective averaged alleles are 10.69, 10.58 and 10.40. In DYS449 they are 29.42, 29.62, and 29.73, respectively. There might be some trend, however, its analysis is beyond the scope of this article. As it was noticed above, in all 1024 L21 haplotypes the average DYS449 allele is 29.50.

All 78 haplotypes contain 730 mutations from the above base haplotype. It gives 730/78/0.08 = 117 → 133 generations, that is 3325±350 years to a common ancestor. The 18 markers in the 67 marker panel add 372 mutations, which gives 1102/78/0.12 = 118 → 134 generations, that is 3350±350 years to a common ancestor. The difference between 49- and 67-marker panel is less than 1%.

One can see that on some reason a timespan to a common ancestor of the L21 subclade is consistently lower in Scotland compared to England and Ireland. The difference is rather small and on the verge (or within) of the margin of error, hence, there is no reason to discuss it seriously, particularly in absence of any supportive data, such as archaeological ones.

Wales, L21

There are 38 haplotypes available. The base haplotype is identical as that in England, with DYS449=29.74, and DYS391=10.79:

13 24 14 11 12 12 12 13 13 – 17 11 11 25 15 19 **30** – 11 11 16 15 18 17 12 12 – 11 9 8 10 10 8 10 10 16 10 12 12 15 8 12 22 20 13 12 11 13 11 11 12 12
(**Wales, L21, 49-marker panel, 3700±410 ybp)**

391 mutations in the dataset give 391/38/0.08 = 129 →148 generations, that is 3700±410 years to a common ancestor. It is essentially the same as those in England (3950±420 ybp**)** and Ireland (3650±380 ybp).

France, L21

The base haplotype of a dataset of 20 haplotypes is identical as that in Ireland, with DYS449=29.30, and DYS391=10.65:

13 24 14 11 12 12 12 13 13 – 17 11 11 25 15 19 29 – 11 11 16 15 18 17 12 12 – 11 9 8 10 10 8 10 10 16 10 12 12 15 8 12 22 20 13 12 11 13 11 11 12 12
(**France, L21, 49-marker panel, 4100±490 ybp**)

225 mutations in the dataset give 225/20/0.08 = 141 →164 generations, that is 4100±490 years to a common ancestor. It is tempting to say that it "older" compared to those in the Isles and points at a route from France to the Isles, however, it is in fact within the margin of error.

Germany, L21

The base haplotype of a dataset of 14 haplotypes is identical as those in Ireland and France, and generally as that of the L21 subclade, except an allele at DYS534 (marked below). While it is on average 15.33 for all 2299 haplotypes (P312), 15.19 in 1024 L21 haplotypes, 15.32 in England, 15.25 in France, 15.13 in Wales, it is 15.64 in Germany.

13 24 14 11 12 12 12 13 13 – 17 11 11 25 15 19 29 – 11 11 16 15 18 17 12 12 – 11 9 8 10 10 8 10 10 16 10 12 12 **15/16** 8 12 22 20 13 12 11 13 11 11 12 12
(**Germany, L21, 49-marker panel, 3700±480 ybp**)

The noticeable deviation of this marker in Germany from those from Iberia, France, the Isles, might indicate the distant location of the region from South-Western Europe, where L21 is apparently arose around 4000 ybp.

145 mutations in the dataset give 145/14/0.08 = 129 →148 generations, that is 3700±480 years to a common ancestor.

Spain, L21

The base haplotype of a dataset of 11 haplotypes is identical as that in all 1024 L21 haplotypes, described above. DYS449 averaged allele among 11 haplotypes is 29.73, which is close to 29.50 in all 1024 L21 haplotypes, and practically to all regional populations considered in this article.

13 24 14 11 12 12 12 13 13 – 17 11 11 25 15 19 **30** – 11 11 16 15 18 17 12 12 – 11 9 8 10 10 8 10 10 16 10 12 12 15 8 12 22 20 13 12 11 13 11 11 12 12
(**Spain, L21, 49-marker panel, 3675±500 ybp**)

113 mutations in the dataset give 113/11/0.08 = 128 →147 generations, that is 3675±500 years to a common ancestor. It is again within margin of error with all other regional populations considered here.

One might say that the Spain L21 population (of 11 haplotypes) is "younger" compared to some other European population. It is not true. First, all of them are within margin of error. Second, statistics is becoming progressively poorer with reducing datasets being analyzed. Third, it was noted above that bearers of P312 haplotypes in the Pyrenees apparently went through a severe population bottleneck, which some of them moved as the Beakers up to the continental Europe. Therefore, a timespan to common ancestors could be shorter in the Pyrenees compared to their more lucky descendants in the continental Europe.

Spain + Portugal, P312*, 67 markers

When P312* subclade is considered in the67 marker format, it contains 20 haplotypes from Spain and Portugal, with exactly the same base haplotype as that of the total P312 dataset (see above):

13 24 14 11 11 14 12 12 12 13 13 29 – 17 9 10 11 11 25 15 19 29 15 15 17 17 –
11 11 19 23 15 15 18 17 36 38 12 12 – 11 9 15 16 8 10 10 8 10 10 12 23 23 16 10 12 12
15 8 12 22 20 13 12 11 13 11 11 12 12 **(Spain+Portugal, P312*, 3950±400 ybp)**

The base haplotype of a dataset of 11 haplotypes is identical as that in all 1024 L21 haplotypes, described above. DYS449 averaged allele among these 20 haplotypes is 29.00 (in the L21 series it was 29.50).

307 mutations in the dataset give 307/20/0.12 = 128 →147 generations, that is 3675±500 years to a common ancestor, exactly as it was with 11 49 marker haplotypes from Spain. However, five Portuguese haplotypes gave 84/5/0.12 = 140 → 163 generations, that is 4075±600 ybp.

Overall, dataset contained 32 of 67-marker haplotypes of the Iberian origin. They contained 209 mutations in the 25 marker format, 401 mutations in the 37 marker format, and 526 mutations in the 67 marker format. It gives:

209/32/0.046 = 142 → 166 generation, that is 4150±500 ybp

401/32/0.090 = 139 →162 generations, that is 4050±450 ybp

526/32/0.120 = 137 → 159 generations, that is 3975±430 ybp

These are practically the same timespans, within margin of error.

Other regions, L21 subclade, 49 marker haplotypes

The rest of regions in Europe with respect to their L21 haplotypes are represented by small number of haplotypes, hence, the data are not accurate and have wide margins of error. We will briefly describe them, however, any strong conclusions cannot be made from the data below.

Norway, 9 haplotypes, 82/9/0.08 = 114 → 129 generations, 3225±480 ybp.

Sweden, 6 haplotypes, 55/6/0.08 = 115 → 130 generations, 3250±550 ybp.

Luxemburg, 4 haplotypes, 37/4/0.08 = 116 → 132 generations, 3300±630 ybp.

Finland, 4 haplotypes, 19/4/0.08 = 59 → 63 generations, 1575±260 ybp.

Poland, P312* subclade, 67 marker haplotypes

The rest of regions in Europe with respect to their P312* haplotypes are represented by small number of haplotypes, hence, the data are not accurate and have wide margins of error. Here is an example of haplotypes from Poland, with an addition of some haplotypes from Lithuania and Ukraine.

Poland, 4 haplotypes in the P312* dataset, 50/4/0.12 = 104 → 116 generations, that is 2900±500 ybp

Poland/Lithuania/Ukraine, 7 haplotypes (the above 4 haplotypes plus 3 haplotypes from Lithuania and Ukraine, provided by Larry Mayka), 116/7/0.12 = 138 → 161 generations, 4025±550 ybp.

* * *

As a conclusion, consideration of R1b1a2-P312 haplotypes and downstream subclades and calculations of timespans to common ancestors of the respective populations resulted in a ladder of TSCAs associated with the phylogenetic (subclade) tree. The pattern obtained supports the hypothesis according to which an upstream subclade of P312 and U106 (likely R1b1a2-L11/L151/P310/P311, along with L51/M412) has arrived to Iberia ~ 4800 years before present, split P312 and U106, which moved up North to continental Europe as Bell Beakers, and spread over Europe between 4500 and 3200 ybp. It seems that both P312 and U106 went through a severe population bottleneck in the 3rd millennium BC, which had lasted almost a thousand years. Therefore a common ancestor of current bearers of P312 and U106 can be detected only around 3700-4100 ybp. Only superposition of the their respective base haplotypes along with their

TSCAs made possible to calculate that their common ancestor (presumably R1b1a2-L11 or P312 itself) lived 4800 ybp, at the very beginning of Bell Beaker movements. This part of the hypothesis is supported by archaeological data, according to which the oldest artifacts related to the Bell Beakers were found in the Pyrenees, and the Bell Beaker movement was in the direction from South-West to North and North-East of Europe and the Isles. It is of interest that archaeologists often claim that the theory of a "Beaker People" has been discarded, since they could not attribute "Beaker People" to any certain population in Europe. According to them, it was "Bell Beaker phenomenon", not "people". Now, with appearance of DNA genealogy, we can state that there were the "Beaker People". There were R1b1a2-P312 people and bearers of P312 downstream subclades.

Acknowledgements: The author is indebted to a Michael Walsh, Administrator of R1b1a2-L21 Project, Dennis Wright, Coordinator of the Irish Type III DNA Project, and Mark Jost, for providing extended haplotype datasets, and for valuable discussions.

References

Adamov, D. and Klyosov, A.A. (2008) Theoretical and practical evaluations of back mutations in haplotypes of Y chromosome. Proc. Russian Academy of DNA Genealogy (ISSN 1942-7484), 1, № 4, 631-645.

Adams, S.A., Bosch, E., Balaresque, P.L., Ballereau, S.J., Lee, A.C., Arroyo, E., López-Parra, A.M., Aler, M., Gisbert Grifo, M.S., Brion, M., Carracedo, A., Lavinha, J., Martínez-Jarreta, B., Quintana-Murci, L., Picornell, A., Ramon, M., Skorecki, K., Behar, D.M., Calafell , F., Jobling, M.A. (2008) The Genetic Legacy of Religious Diversity and Intolerance: Paternal Lineages of Christians, Jews, and Muslims in the Iberian Peninsula. Am. J. Human Gen. 83, 725-736.

Ballantyne, K.N., Goedbloed, M., Fang, R., Schaap, O., Lao, O., Wollstein, A., Choi, Y., van Duijn, K., Vermeulen, M., Brauer, S., Decorte, R., Poetsch, M., von Wurmb-Schwark, N., de Knijff, P., Labuda, D., Vezina, H., Knoblauch, H., Lessig, R., Roewer, L., Ploski, R., Dobosz, T., Henke, L., Henke, J., Furtado, M.R., Kayser, M. (2010) Mutability of Y-chromosomal microsatellites: rates, characteristic, molecular bases, and forensic implications. Am. J. Human Genet. 7, 341-353.

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

Klyosov, A.A. (2008) Mysteries of the "Western European" haplogroup R1b. Proc. Russian Academy of DNA Genealogy (ISSN 1942-7484), 1, № 4, 568-630.

Klyosov, A.A. (2009a). DNA Genealogy, mutation rates, and some historical evidences written in Y-chromosome. I. Basic principles and the method. J Genetic Genealogy 5, 186-216.

Klyosov, A.A. (2009b) DNA Genealogy, mutation rates, and some historical evidences written in Y-chromosome. II. Walking the map. J. Genetic Genealogy. 5, 217 – 256.

Klyosov, A.A. (2009c) Iberian haplotypes and analysis of populations of the Basques, Sephards, and other populations of Spain and Portugal. Proc. Russian Academy of DNA Genealogy (ISSN 1942-7484), 2, № 3, 390-421.

Klyosov, A.A. (2010a) Haplogroup R1b. Part 1. Proc. Russian Academy of DNA Genealogy (ISSN 1942-7484), 3, No.2, 249-299.

Klyosov, A.A. (2010b) Irish haplotypes and haplogroups. Proc. Russian Academy of DNA Genealogy (ISSN 1942-7484), 3, No. 6, 1029-1053.

Klyosov, A.A. (2010c) Subclade R1b1b2-M153 on the Iberian Peninsula and among the Basques. Proc. Russian Academy of DNA Genealogy (ISSN 1942-7484), 3, No. 6, 976-982.

Klyosov, A.A. (2010d) Subclade R1b1b2-L226 ("Irish III") – the latest update. Proc. Russian Academy of DNA Genealogy (ISSN 1942-7484), 3, No. 7, 1211-1213.

Klyosov, A.A. (2010e) Consideration of 37 Irish R1a1 haplotypes and 1036 R1b1 haplotypes in the same series of 67 marker haplotypes: a separation of the R1a1 subclade of the Russian Plain and R-M222, and timespans to their common ancestors. Proc. Russian Academy of DNA Genealogy (ISSN 1942-7484), 3, No. 3, 398-405.

Klyosov, A.A. (2010f) Haplotypes of P312* subclade: a history in letters. Proc. Russian Academy of DNA Genealogy (ISSN 1942-7484), 3, No. 7, 1165-1183.

Klyosov, A.A. and Rozhanskii, I. L. (2010) Reconsideration of an average mutation rate constant for 67 marker haplotypes – from 0.145 to 0.120 mutations per haplotype per generation. Proc. Russian Academy of DNA Genealogy (ISSN 1942-7484), 3, № 12, 2039-2058.

Klyosov, A.A. (2011a) Calculations of numerical values of mutation rate constants for the slowest 22 markers of the 67 marker panel of haplotypes. Proc. Russian Academy of DNA Genealogy (ISSN 1942-7484), 4, № 5, 968-987.

Klyosov, A.A. (2011b) DNA genealogy of the major haplogroups of the male half of mankind. Proc. Russian Academy of DNA Genealogy (ISSN 1942-7484), 4, № 5, 988-1014.

Muller, J., van Willigen, S. (2001) New radiocarbon evidence for European Bell Beakers and the consequences for the diffusion of the Bell Beaker phenomenon. In Bell Beakers Today: Pottery, People, Culture, Symbols in Prehistoric Europe. Proc. Int. Colloquium, Riva del Garda (Trento, Italy), ed. F. Nicolis, pp. 59-80.

Myres, N. M., Rootsi, S., Lin, A.A., Jarve, M., King, R.J., Kutuev, I., Cabrera, V.M., Khusnutdinova, E.K., Pshenichnov, A., Yunusnayev, B., Balanovsky, O., Balanovska, E., Rudan, P., Baldovic, M., Herrera, R.J., Chiaroni, J., Cristofaro, J. D., Villems, R., Kivisild, T., Underhill, P.A. (2010) A Major Y-chromosome haplogroup R1b Holocene era founder effect in Central and Western Europe. Eur. J. Human Genetics, advance on-line publication, 26 August 2010; doi: 10.1038/ejhg.2010.146

MacDonalds and Scottish R1b1a2 Haplotypes

(an updated version)

Anatole A. Klyosov

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

Introduction

This article updates an earlier paper [Klyosov, A.A. (2010) MacDonalds and Scottish R1b1b2 haplotypes. Proc. Russian Academy of DNA Genealogy, 3, No.10, 1696-1735], since after the publication the mutation rate constant for 67 marker haplotypes was decreased by 21% after its careful re-calibration [Klyosov, A.A. and Rozhanskii, I. L. (2010) Reconsideration of an average mutation rate constant for 67 marker haplotypes – from 0.145 to 0.120 mutations per haplotype per generation. Proc. Russian Academy of DNA Genealogy, 3, № 12, 2039-2058].

Clan Donald R1b1a2 haplogroups and lineages

The Clan Donald USA Project (http://dna-project.clan-donald-usa.org/tables.htm) after a recent consideration contained 657 R1b haplotypes, 288 of them were in the 67 marker format.

A haplotype tree of those 67 marker haplotypes is shown in Fig. 1. The numbering has been done in according with the listing of R1b haplotypes in the same order as they were presented by the end of August, 2010. The Project organizers employed a color code for designation of different subgroups of R1b on the list. A connection between the numbering and the color code is as follows:

Red	1-71
Pink	72-82
Brown	83-88
Maroon	89-100
Green	101-183
Dark Blue	184-193
Blue Green	194-196
Yellow Gray	197-204

Violet	205-210
Orange	211-215
Magenta	216-242
Pale Violet	243-246
Gray	247-251
Pale Green	252-260
Dark Blue – Green	261-265
Black	266-275
Pale Blue	276-278
Light Brown	279-285
Dark Green	286-288
Yellow Green	289-299
White	300-307
Yellow (unclassified)	308-657

Only 67 and 37 marker haplotypes have been considered and analyzed in this study. In principle, nothing prevents to repeat this work with 25- and 12-marker haplotypes, however, resolution of the haplotype tree would be not so good, since three-quarters and more of alleles in haplotypes will be lost.

The word “Scottish” in the title of this paper is conditional, since many of the listed haplotypes are not necessarily “Scottish”. However, many (and probably most) of them are. After all, names and territories in the field of DNA genealogy are all conditional, taking into account migrations of the past.

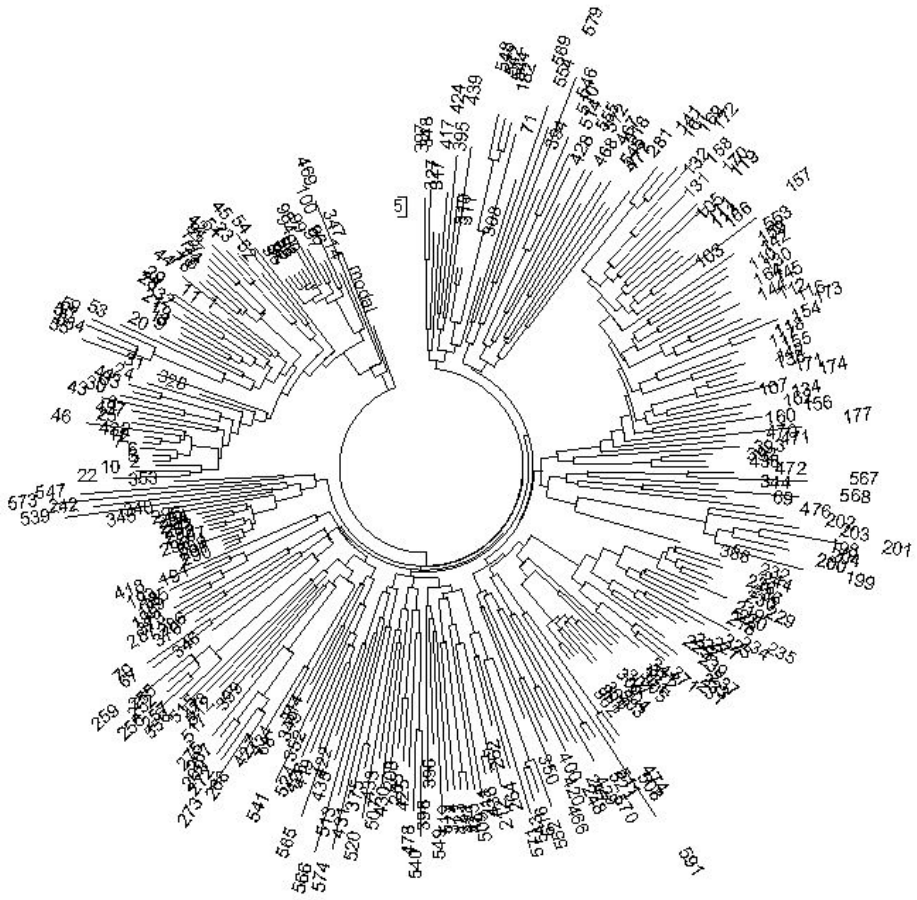


Fig. 1. A 67-marker haplotype tree for 288 haplotypes of R1b haplogroup (mainly subclades of R1b1b2) of the Clan Donald USA project http://dna-project.clan-donald-usa.org/tables.htm

The first wide branch of haplotypes in the upper left quadrant belongs mainly to RED subgroup, with its adjacent MAROON and PINK sub-sub-groups. We will consider here whether the last two are independent branches, or downstream ones, descending from the RED branch. Here as the branch looks on the linear tree (a fragment).

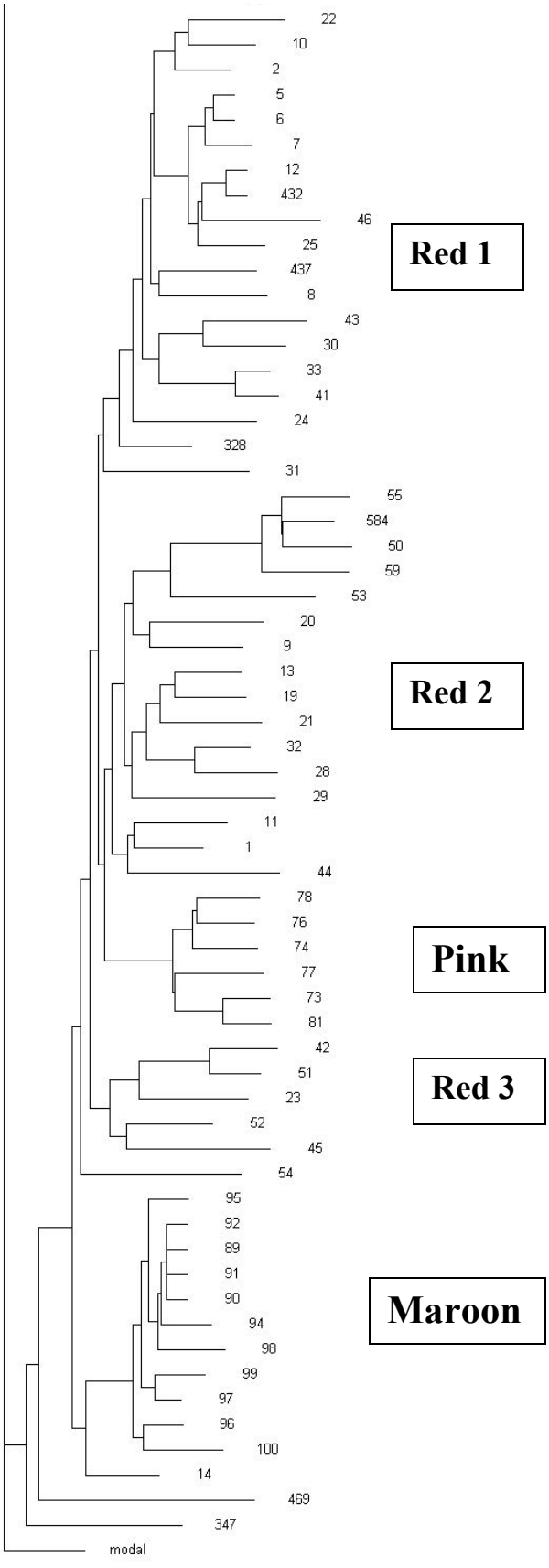

22
10
2
5
6
7
12
432
46
25
437
8
43
30
33
41
24
328
31
55
584
50
59
53
20
9
13
19
21
32
28
29
11
1
44
78
76
74
77
73
81
42
51
23
52
45
54
95
92
89
91
90
94
98
99
97
96
100
14
469
347
modal
Red 1
Red 2
Pink
Red 3
Maroon

Fig. 2. A fragment of the 67-marker linear haplotype tree, depicting the upper left branch of the haplotype tree, shown in Fig. 1. One can see that some "unclassified" haplotypes (numbers above 308) in fact are parts of "color coded" subgroups, such as haplotypes 328, 437, 584 belong to the RED subgroup, haplotypes 347 and 469 remain to be "unclassified" and do not join any subgroup in the tree, and haplotype 14 (RED) is in fact "unclassified", it does not belong to the RED subgroup, at least according to its haplotype structure.

RED

The Project organizers described the RED branch as follows:

Descendants of the kindred of the Dalriadic royal house in Scotland lie within in this group. However, only a fraction of people in the group are expected to be descendants of the royal house; the group is simply too large and diverse. Fergus, Angus, and Lorne, the sons of Erc, are descended per Irish and Scottish history from Cairbre Riada, king of Irish Dalriada. Cairbre Riada was descended from Conor II High King of Ireland and Sarad, daughter of Conn of the Hundred Battles. Irish history indicates that Cairbre had led his followers from Munster to Antrim. This kindred is traditionally considered to be of Erainn descent. Reverends Archibald and Angus MacDonald appear to have erred in placing Colla Uais in this line of descent. All our participants in this group who have been SNP tested have tested positive for the SNP marker L21/S145, making this group a separate Haplogroup within R1b.

According to the analysis in the preceding work (Klyosov, 2011) and earlier studies (Klyosov, 2010a,b,c), the base (ancestral) L21 haplotype is as follows:

13 24 14 11 11 14 12 12 12 13 13 29 – 17 9 10 11 11 25 15 19 29 15 15 17 17 –
11 11 19 23 16 15 18 17 36 38 12 12 – 11 9 15 16 8 10 10 8 10 10 12 23 23 16 10 12 12
15 8 12 22 20 13 12 11 13 11 11 12 12 (**L21, 3750±380 ybp**)

A common ancestor of this subclade lived 3750±380 years before present.

Let us see what the RED branch tells us in that regard. 18 haplotypes in the Red1 branch (67 marker haplotype 432, "unclassified", turned out to be identical with haplotype 12, and was removed from the count) contained collectively 102 mutations from the base haplotype (apparently ancestral)

13 24 14 **10** 11 14 12 12 12 13 13 **30** – **18** 9 10 11 11 25 15 19 **30** 15 15 17 17 –
11 **12** 19 **24** **15** 15 18 17 36 **37** 12 12 – **12** 9 15 16 8 10 10 8 10 10 12 **22** 23 16 10 12 12
15 8 **11** 22 20 13 12 11 13 11 11 12 12

It had 11 mutations (marked in bold) compared to the base L21 haplotype (see above), which corresponds to 2550 years between their common ancestors (actually, to the present time and back to another common ancestor). We will need this figure later.

Such an approach with a haplotype tree building allows us to move two "unclassified" haplogroups, 328 and 437 (see legend to Fig. 2) to the RED branch. They belong there. Bearers of these two haplotypes will probably be delighted.

102 mutations in 18 of 67 marker haplotypes correspond to 102/18/0.12 = 47 generations (25 year generations according to the calibration) without a correction for back mutations (Klyosov, 2009a), or 50 generations with the correction, that is 1250±175 years to a common ancestor for the branch Red1 (Figs. 1 and 2). This is the mid of the 8th century AD plus-minus a century or two.

All mutations in the Red1 base haplotype (above) are rather common, except one, in a pair YCAIIa,b = 19-24. Commonly in R1b1a2 haplotypes it is 19-23. This "24" is certainly a signature for the RED branch, for all Red1, Red2 and Red3 sub-branches, as it will be shown below. It already shows their tight relationships.

Let us move to the Red2 sub-branch. 16 haplotypes in the Red2 branch (including "unclassified" haplotype 584, which turned out to be a member of a tight family of four haplotypes, see Fig. 2) contained collectively 115 mutations from the base haplotype

13 24 14 **10** 11 14 12 12 12 13 13 **30** – **18** 9 10 11 11 25 15 19 **30** 15 15 17 17 – 11 **12** 19 **24** 16 15 18 17 **37** 38 12 12 – **12** 9 15 16 8 10 10 8 10 10 12 **22** 23 16 10 12 12 **14** 8 **11** 22 20 13 12 11 13 11 11 12 12

115 mutations in 16 of 67 marker haplotypes correspond to 115/16/0.12 = 60 generations without a correction for back mutations, or 64 generations with the correction, that is 1600±220 years to a common ancestor for the branch Red2 (Figs. 1 and 2). This is the 5th century AD plus-minus a couple of centuries.

The base Red2 haplotype has the same number of 11 mutations compared to the base L21 haplotype, which corresponds to 2550 years between their common ancestors. On the other hand, base haplotypes of Red1 and Red2 branches differ from each other by only 2.85 mutations on average (it looks like four mutations, when rounded up) in the third and the forth panels of the 67-marker haplotypes. This places their common ancestors by only 625 years apart. It means that a common ancestor of the Red1 and Red2 branches lived approximately (625+1250+1600)/2 = 1700 years before present. It seems that a common ancestor of the Red1 branch descended from the Red2 branch. We will examine it later.

The Red3 sub-branch contains only 6 haplotypes, which collectively contain 45 mutations from the base haplotype

13 24 14 **10** 11 14 12 12 12 13 13 **30** – **18** 9 10 11 11 25 15 19 **30** 15 15 **16** 17 –
11 **12** 19 **24** 16 15 18 17 **37** 38 12 12 – **12** 9 15 16 8 10 10 8 10 10 12 **22** 23 16 10 12 12
15 8 **11** 22 20 13 12 11 13 11 11 12 12

45 mutations in 6 of 67 marker haplotypes correspond to 45/6/0.12 = 63 generations without a correction for back mutations, or 67 generations with the correction, that is 1675±300 years to a common ancestor for the branch Red3 (Figs. 1 and 2). This is the 4th century AD plus-minus three centuries. Six haplotypes do not provide with a good statistics, hence, rather large margins of error.

One can see that common ancestors of the Red1, Red2 and Red3 branches lived within a few centuries, 1250±175, 1600±220, and 1675±300 years ago. They are equidistant from a common ancestor of L21 subclade (11 mutations in each of their base haplotypes). Furthermore, their base haplotypes (Red1, Red2 and Red3) differ from each other by only 5 mutations in 201 alleles (marked in bold):

13 24 14 10 11 14 12 12 12 13 13 30 – 18 9 10 11 11 25 15 19 30 15 15 17 17 –
11 12 19 24 **15** 15 18 17 **36** **37** 12 12 – 12 9 15 16 8 10 10 8 10 10 12 22 23 16 10 12 12
15 8 11 22 20 13 12 11 13 11 11 12 12

13 24 14 10 11 14 12 12 12 13 13 30 – 18 9 10 11 11 25 15 19 30 15 15 17 17 –
11 12 19 24 16 15 18 17 37 38 12 12 – 12 9 15 16 8 10 10 8 10 10 12 22 23 16 10 12 12
14 8 11 22 20 13 12 11 13 11 11 12 12

13 24 14 10 11 14 12 12 12 13 13 30 – 18 9 10 11 11 25 15 19 30 15 15 **16** 17 –
11 12 19 24 16 15 18 17 37 38 12 12 – 12 9 15 16 8 10 10 8 10 10 12 22 23 16 10 12 12
15 8 11 22 20 13 12 11 13 11 11 12 12

This places THEIR common ancestor at 14 generations (350 years) below their average "age" (1510±230 years bp), that is he lived 1860±250 years before present. Their common ancestor, that is the RED branch common ancestor had the following haplotype (base haplotype of the RED branch)

13 24 14 10 11 14 12 12 12 13 13 30 – 18 9 10 11 11 25 15 19 30 15 15 17 17 –
11 12 19 24 16 15 18 17 37 38 12 12 – 12 9 15 16 8 10 10 8 10 10 12 22 23 16 10 12 12
15 8 11 22 20 13 12 11 13 11 11 12 12

He lived in the 2nd century AD and its haplotype differed by 10 mutations from the base R1b1b2-L21 haplotype, shown below with the mutations marked:

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

Since the R1b1b2-L21 subclade' common ancestor lived 3750±380 years before present, those 10 mutations (equivalent to 2275 years between their common ancestors) in the RED base haplotype (with a common ancestor 1860±250 before present) place THEIR common ancestor at approximately (2275+1860+3750)/2 = 3940 years before present. This is the L21 common ancestor himself.

Therefore, a short story is that bearers of R1b1b2 haplogroup (subclades L51 and/or L11) who have arrived to Iberia around 4800 years before present (via the North-African route from the Middle East/Levant which began around 5500 years before present) and moved up to the European continent as the Bell Beakers and with their downstream L21 subclade (with a common ancestor of 3750±380 years bp), eventually got to the Isles and later split off the RED branch, with a common ancestor in the 2nd century AD. He well might have been an ancestor of *the kindred of the Dalriadic royal house in Scotland, of Fergus, Angus, and Lorne, the sons of Erc, who descended per Irish and Scottish history from Cairbre Riada, king of Irish Dalriada.*

PINK

There are only six 67 marker haplotypes of the PINK subgroup on the tree (Fig. 2). They are clearly part of the RED subgroup, with the PINK base haplotype

13 24 14 10 11 **15** 12 12 12 13 13 29 – 18 9 10 11 11 **24** 15 19 **32** 15 15 17 17 –
11 12 19 24 **15** 15 18 17 **38** **39** 12 12 – 12 9 15 16 8 10 10 8 10 10 12 22 23 16 10 12 12
16 8 11 22 20 13 12 11 13 11 11 12 12

with the same characteristic pair of 19-24 in YCAII, the same as that is the PINK subgroup. All 6 haplotypes have 19 mutations from the above base haplotype, which gives 19/6/0.12 = 26 → 27 generations, that is 675±170 years to their common ancestor. The above PINK base haplotype has 8 mutations (marked in bold) from the overall base haplotype of the RED subgroup, which translates to 1800 years between their common ancestors (8/0.12 = 67 → 72 generations), and places their with the RED subgroup common ancestor at (675+1800+1860)/2 = 2200±300 years before present. Indeed, PINK subgroup is the downstream branch of the RED subgroup (with a common ancestor of 1860±250 years before present, which are the same figures within the margin of error).

MAROON

The Project organizers describe this subgroup as follows: "*The group is a subset of the red 'Scots' group, consisting mostly of Alexanders*". Indeed, 11 of 12 group members are Alexanders, and one who is not (and who has only 37-marker haplotype) has the same haplotype structure, and even without any mutations from the base haplotype

13 24 14 10 11 14 12 12 12 13 13 30 – 18 9 10 11 11 25 15 19 30 15 **16** 17 17 –
11 12 19 **23** 16 15 **17** 17 37 38 12 12 – 12 9 15 16 8 10 10 8 10 10 12 **23** 23 16 10 12 12
15 8 11 22 20 13 12 11 13 11 11 12 12

It has only four mutations from the base RED haplotype (with a common ancestor of 1860±250 years bp). As one can see, it has a 19-23 pair in YCAII, unlike 19-24 in the RED branch. Apparently, we see a result of a back mutation. All 11 of 67 marker haplotypes have 15 mutations from the above base haplotype, which gives 15/11/0.12 = 11.4±3.2 generations from a common ancestor (if to be excessively precise), that is 285±80 years before present.

There is one more way of calculations, taking into account a number of base haplotypes in the dataset (that is identical ones in the dataset). In this case there are four identical haplotypes in the 67 marker set of 11 haplotypes, and seven identical ones in the 37 marker set of 12 haplotypes. At such a small amount of haplotypes in the datasets the method is rather imprecise, and serves just an illustration of the approach. In this case we have [ln(11/4)]/0.12 = 8.4 generations from a common ancestor, and [ln(12/7)]/0.09 = 6 generations. Those are approximate figures, however, they show that the sub-branch is indeed rather young. 285±80 years from a common ancestor of the MAROON branch is the most reliable figure here.

The haplotype tree (Fig. 2) suggests that the MAROON branch descended from the Red3 branch (with a common ancestor of 1675±300 years ago). Since there are five mutations between their base haplotypes (marked in bold in the Red3 base haplotype below)

13 24 14 10 11 14 12 12 12 13 13 30 – 18 9 10 11 11 25 15 19 30 15 **15** **16** 17 –
11 12 19 **24** 16 15 **18** 17 37 38 12 12 – 12 9 15 16 8 10 10 8 10 10 12 **22** 23 16 10 12 12
15 8 11 22 20 13 12 11 13 11 11 12 12

their common ancestors are separated by 5/0.12 = 42 → 44 generations, that is 1100 years, and THEIR (joint) common ancestor lived approximately (1100+1675+285)/2 = 1530 years before present. Indeed, this is the Red3 branch common ancestor (1675±300 ybp), within the margin or error.

BROWN

The BROWN subgroup haplotypes are available in 37 marker format only. Their base haplotype is

13 24 14 10 11 **15** 12 12 12 13 13 30 – **20** 9 10 11 11 25 15 19 30 15 15 **16** 17 –
11 12 19 24 **15 14** 18 **16 38 39** 12 12

and all six haplotypes in the dataset have 14 mutations from it (marked in bold are 9 mutations from the base RED haplotype). We see again the RED subgroup "signature" 19-24 pair in YCAII markers. The number of mutation gives 14/6/0.09 = 26 generations, or 27 generations with a correction for back mutations, that is 675±190 years from a common ancestor.

The closest to the BROWN base haplotype is the both the RED one, with 9 mutations between them in the 37 markers, and the PINK one, shown below (mutations between BROWN and PINK base haplotypes are marked in bold), also with 9 mutations between them:

13 24 14 10 11 15 12 12 12 13 13 **29** – **18** 9 10 11 11 **24** 15 19 **32** 15 15 **17** 17 –
11 12 19 24 15 **15** 18 **17** 38 39 12 12

"Closest" in this context means that the both common ancestors (BROWN and RED, or BROWN and PINK) descended from the same common ancestor, who lived earlier than them, or one of them was the common ancestor himself for the both branches. Let us see which case is more justified.

Nine mutations between two 37 marker haplotypes translate to 9/0.09 = 100 → 111 generations (with a correction for back mutations), that is 2775 years between them.

-- Since the RED common ancestor lived 1860±250 years before present, the RED and BROWN common ancestor lived approximately (1860+2775+675)/2 = 2655 years ago.

-- Since the PINK common ancestor lived 675±170 years before present, the PINK and BROWN common ancestor lived approximately (675+2775+675)/2 = 2060 years ago, at the break between BC and AD.

Clearly, in neither case the BROWN common ancestor has descended from the RED or the PINK one. 9 mutations between two 37 marker haplotypes (111 generations) are too many to be considered "closely related". So, the conclusion of the Clan Donald site "*The brown group is a subset of the red 'Scots' group*" is

highly doubtful. It is a rather recent branch (675±190 years from their common ancestor), but it likely passed through a population bottleneck before that, and in fact has its roots in much more ancient times (2500-2000 years bp), some time after the L21 formation (3750±380 years bp) and before the RED group formation (1860±250 years bp). That is where the 19-24 pair came from.

GREEN

The Clan Donald site describes this subgroup as follows:

The group is centered geographically in northwestern Ireland. It is always referred to as the 'Irish' or 'Niall' group. Clan Donalds with this genetic signature may be descendants of Colla Uais, descendants of the O'Neill, or O'Donnell kindreds with whom we regularly served in Irish military operations, the O'Cahans from the Dowry of fighting men provided to Angus Og upon his marriage to the O'Cahan's daughter or descendants of Colla Menn from whom Gillebride sought assistants to recover his lands from the Norse. All participants in this group who have been SNP tested have tested positive for the marker M222.

According to our data (Klyosov, 2011, 2010a) the base haplotype for R-M222 subclade is

13 25 14 11 11 13 12 12 12 13 14 29 – 17 9 10 11 11 25 15 18 30 15 16 16 17
11 11 19 23 17 16 18 17 38 39 12 12 – 11 9 15 16 8 10 10 8 10 10 12 21 23 16 10 12 12
16 8 12 25 20 13 12 11 13 11 11 12 12

with a common ancestor who lived 1450±160 years before present. This subclade embraces about 25% of the Irish R1b1b2 population (Klyosov, 2010a).

Let us see whether the GREEN subgroup matches this base haplogroup and a timespan to the common ancestor. A fragment of the Donald Clan haplotype tree (Fig. 1) encompassing the GREEN branch is shown in Fig. 3. It consists of many "mini-lineages", as in any real haplotype tree, and we will analyze the branch as a whole.

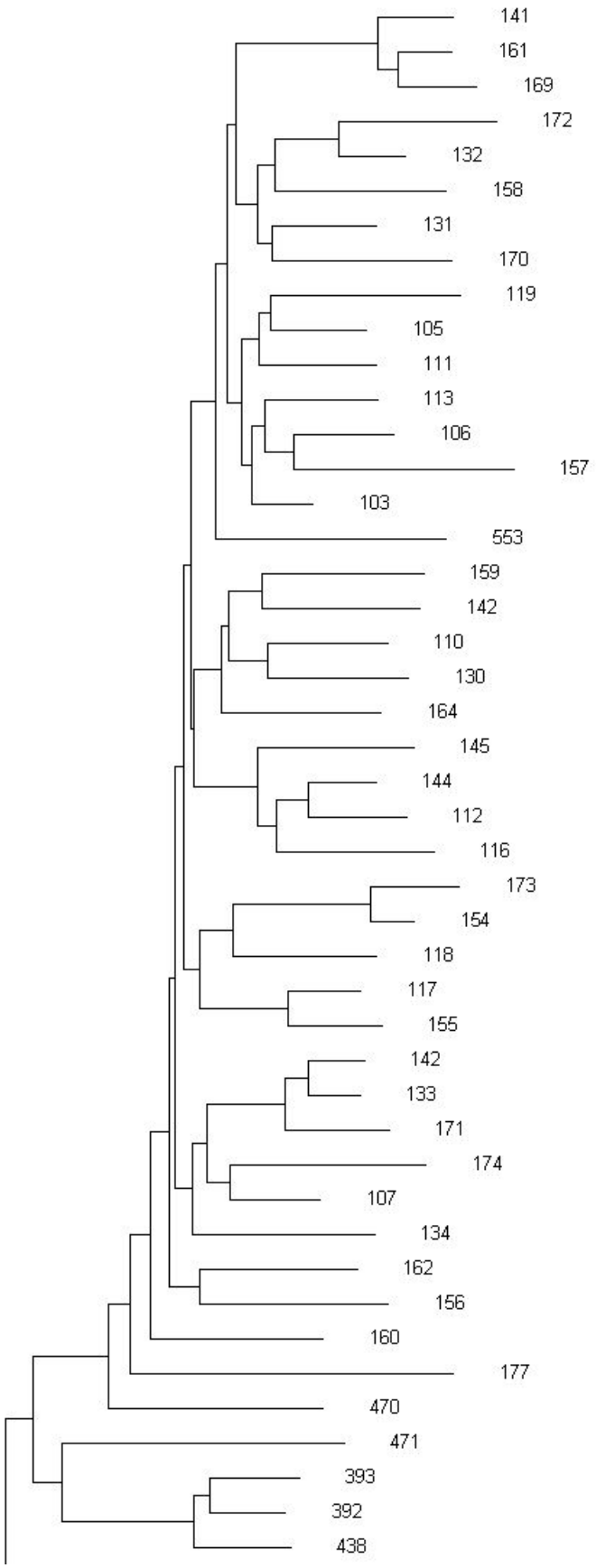

141
161
169
172
132
158
131
170
119
105
111
113
106
157
103
553
159
142
110
130
164
145
144
112
116
173
154
118
117
155
142
133
171
174
107
134
162
156
160
177
470
471
393
392
438

Fig. 3. A fragment of the 67-marker linear haplotype tree, depicting the right-hand side (between 1 and 3 o'clock) wide GREEN branch of the haplotype tree, shown in Fig. 1 (between haplotypes 101 and 183). One can see that "unclassified" haplotype 553 in fact included into this subgroup. The GREEN branch is very clean and does not include other color-coded haplotypes. Only haplotype 182, which was assigned by the GREEN branch by the Clan Donald organizers, landed in the "unclassified" branch in the very upper part of the right-hand side of the tree (Fig. 1)

In fact, relative positions of these mini-lineages (mini-subgroups, with a few haplotypes in each) do not bear any particular significance. This is illustrated with the same GREEN branch, being a part of a smaller haplotype tree, which includes only haplotypes from 1 through 204 (Fig. 4).

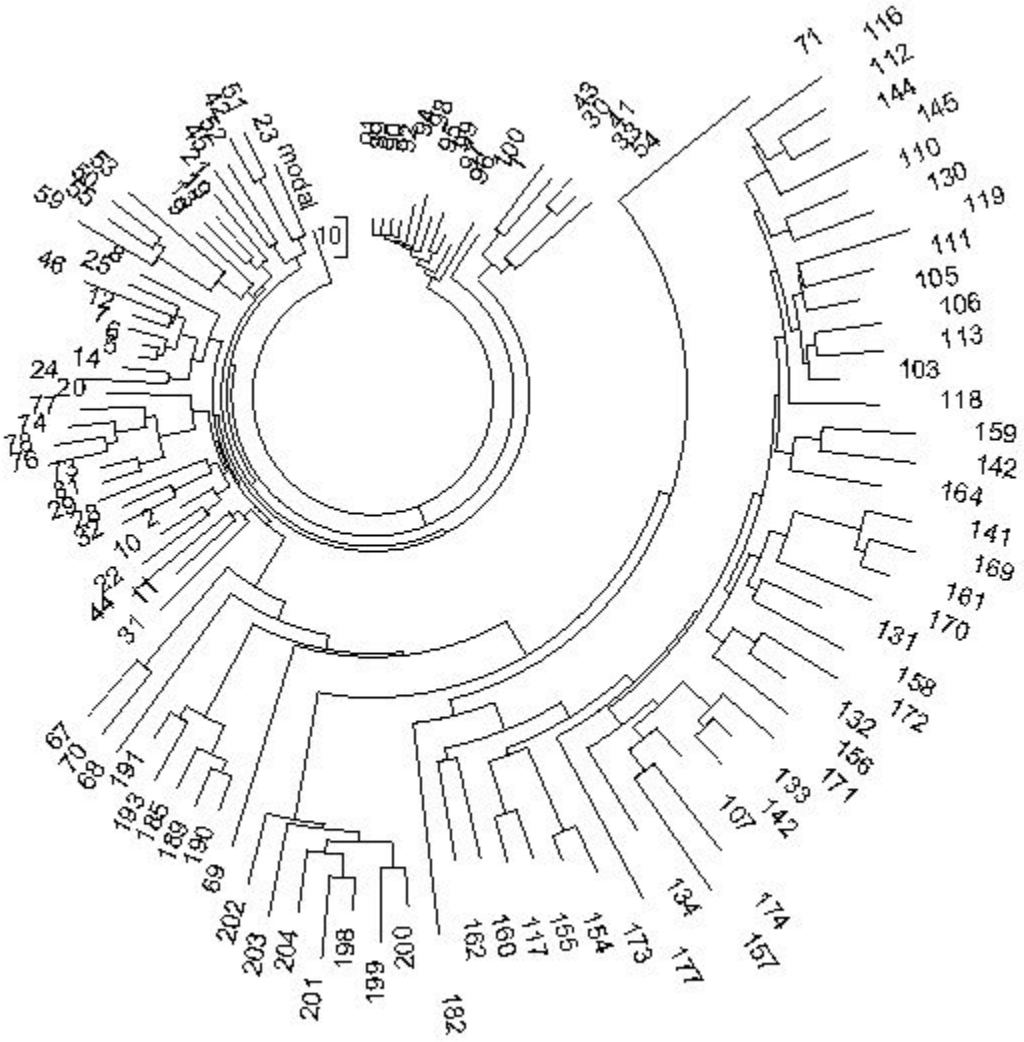


Fig. 4. A 67-marker haplotype tree for first 112 haplotypes in the tree in Fig. 1 (from haplotypes 1 through 204). The GREEN branch is on the right.

One can see that haplotypes on the tree in Fig. 3 and 4 are the same, however, their relative positions vary.

All 40 haplotypes of the GREEN branch contain 331 mutations from the base haplotype of the branch

13 25 14 11 11 13 12 12 12 13 14 29 – 17 9 10 11 11 25 15 18 30 15 16 16 17
11 11 19 23 17 16 18 17 38 39 12 12 – 11 9 15 16 8 10 10 8 10 10 12 21 23 16 10 12 12
16 8 12 25 20 13 12 11 13 11 11 12 12

This is exactly the M222 base haplotype, shown three pages up.

This number of mutations give 331/40/0.12 = 69 → 74 generations, that is 1850±210 years from a common ancestor. This is a figure similar with that of 1450±160 years bp for the M222 subclade, obtained earlier with 266 of 67 marker haplotypes (and calculated for the first 25 markers only) (Klyosov, 2010a,b), and with 1575±170 years bp calculated with 172 of 67-marker haplotypes (Klyosov, 2010a).

YELLOW GREEN

Below haplotype 22 from the Red1 subgroup in the haplotype tree (Fig. 1) at 9 o'clock, after a short branch of unclassified haplotypes (among which sits a lonely "magenta" haplotype 242, which actually does not belong to the MAGENTA subgroup), there is the YELLOW GREEN branch of 10 haplotypes. Nine of them belong to McConnell family. That is how the Clan Donald site describes the subgroup:

The group is thought to be from Leinster. It is easily distinguished by the 464X test, whose results turn out 15c-15c-17g-17g. One participant has tested positive for the marker L21.

In fact, 15-15-17-17 in DYS464a-d is very common among R1b1b2 haplotypes, and is characteristic to the "Atlantic Modal Haplotype", and also R-M269 and its subclades U106, P312, U152, L20, L21, "South Irish", "North Irish", "Scottish Borders" and also some unassigned branches (Klyosov, 2010a,b,c).

Since nine out of ten haplotypes in a dataset belong to the same surname, one can expect that their common ancestor lived rather recently. This was exactly what has happened. All ten haplotypes contain only 14 mutations, per 670 alleles, from the base haplotype

13 24 14 11 11 14 12 12 12 **14** 13 30 – 17 9 10 11 11 25 15 19 **30** 15 15 17 17 –
11 **10** 19 23 16 15 18 17 **39 41** 12 12 – 11 9 15 16 8 10 10 8 10 10 12 23 23 **16/17** 10 12
12 15 8 12 22 20 **14** 12 11 13 11 11 12 12

in which 10.7 mutations from the L21 base haplotype are marked in bold. Therefore, 14/10/0.12 = 11.7±3.3 generations, that is 290±80 years to a common ancestor. This is the first half of the 18th century, plus-minus about three generations. 10.7 mutations from the base L21 haplotype (3750±380 years before present) separate their common ancestors by 2450 years, which is not enough to have L21 as a parent subclade for the YELLOW GREEN subgroup (290±80 ybp). The latter descended from a common ancestor within the L21 subclade. Apparently, it is that "unclassified" branch next to the YELLOW GREEN subgroup with the base haplotype (9.5 mutations from the YELLOW GREEN base haplotypes are marked in bold; they separate the YELLOW GREEN and this unclassified base haplotype by 2150 years)

13 24 14 11 11 **15** 12 12 12 **13** 13 29 – 17 9 10 11 11 25 15 19 30 15 15 17 17 –
11 10 19 23 **15** 15 18 17 **36** 38 12 12 – 11 9 15 16 8 10 10 8 10 10 12 23 23 **16** 10 12 12
15 8 12 22 20 14 12 11 13 11 11 12 12

This unclassified branch is rather close to the L21 base haplotype (see below) and has only 4 mutations from it (marked) in all 67 markers. This separates their common ancestors by 850 years.

13 24 14 11 11 **14** 12 12 12 13 13 29 – 17 9 10 11 11 25 15 19 **29** 15 15 17 17 –
11 11 19 23 **16** 15 18 17 36 38 12 12 – 11 9 15 16 8 10 10 8 10 10 12 23 23 16 10 12 12
15 8 12 22 20 **13** 12 11 13 11 11 12 12 (**L21**)

All 7 haplotypes in the "unclassified" branch contain 96 mutations from its base haplotype (second to the above), which gives 96/7/0.12 = 114 → 129 generations, that is 3225±460 years from its common ancestor. This places a common ancestor of the YELLOW GREEN and the unclassified branch to (3225+260+2150)/2 = 2800±500 ybp, which is within the error margin the time when the common ancestor of the unclassified branch had lived (3225±460 ybp). A common ancestor of both L21 (3750±380 ybp) and the unclassified branch (2625±375 ybp) lived 3600±500 years before present, and he was likely (within the margin of error) the founder of L21 himself.

Conclusion: The YELLOW GREEN group descended from an "unclassified branch" (its base haplotype is shown above) which in turn belonged to the L21 subclade and descended from the L21 only 850 years after the L21 was established.

DARK BLUE

The subgroup contains 10 haplotypes, with their bearers having the same or similar surnames. Nine of those ten haplotypes are in the 37 marker format, only five of them were determined in the 67 marker format. Their base haplotype is as follows:

13 24 14 11 11 14 12 12 11 14 13 30 – 18 9 10 11 11 25 15 19 29 15 17 17 17 –
11 11 19 23 15 15 17 17 37 37 12 12 – 11 9 15 16 8 10 10 8 10 11 12 23 23 15 10 12 12
13 8 12 22 19 13 12 11 13 11 11 12 12

- All five of 67 marker haplotypes contain 14 mutations, which gives 14/5/0.12 = 23 generations, that is 575±165 years to their common ancestor.
- All nine of 37 marker haplotypes contain 12 mutations, and 12/9/0.09 = 15 generations, that is 375±115 years to their common ancestor.
- In the 25-marker format there were four base haplotype out of ten, hence, [ln(10/4)]/0.046 = 20 generations, that is 500 years to their common ancestor.
- In the 12-marker format there were seven base haplotype out of ten, hence, [ln(10/7)]/0.022 = 16 generations, that is 400 years to their common ancestor.

As one can see, all the variants of calculations by different methods and with different sets of haplotypes give practically the same figures within the margins of error. The average timespan to a common ancestor for the DARK BLUE subgroup is 460±90 years. It is the 16th century AD.

Since it is a rather recent branch, it is difficult to even suggest their subclade comparing their base haplotype with those of different subclades of R1b1a2. For example, it has as many as

- 28 mutations from the base M222 haplotype,
- 21 mutations from the L51 base haplotype,
- 19 mutations from the L226 base haplotype,
- 18 mutations from the “North Irish” base haplotype,
- 16 mutations from the “South Irish” base haplotype,
- 15 mutations from the L23 base haplotype,
- 14 mutations from the U152 base haplotype,
- 13 mutations from the base L21 haplotype.

One can see that it is the closest one to the L21 subclade.

PALE GREEN

The group is described in the Clan Donald as follows:

This group contains several persons known to descend from the line of the Glencoe chiefs. This line traditionally descends from Somerled. It originated Iain Og, son of Angus Og and brother of John first Lord of the Isles.

The subgroup contains 9 haplotypes, all McDonald or MacDonald, hence, we can expect to see a rather recent common ancestor of the family. Eight of them have their haplotypes determined in he 37 marker format, and only five were determined in the 67 marker format (see Fig. 5). Their base haplotype is as follows:

13 23 15 11 11 14 12 12 11 12 13 28 – 19 9 10 11 11 25 15 18 31 15 15 17 17 –
11 11 19 23 17 15 20 19 38 43 12 12 – 11 9 15 16 8 10 10 8 10 10 12 23 23 16 10 12 12 16 8 13 22 20 12 12 10 13 11 11 12 12

- All five of 67 marker haplotypes contain 13 mutations, which gives 13/5/0.12 = 22 generations, that is 550±160 years to their common ancestor.
- All eight of 37 marker haplotypes contain 11 mutations, and 11/8/0.09 = 15 generations, that is 375±120 years to their common ancestor.
- In the 25-marker format there were five base haplotype out of ten, hence, [ln(10/5)]/0.046 = 15 generations, that is 375 years to their common ancestor.
- In the 12-marker format there were no mutations in ten haplotypes which points out to a recent common ancestor, within a few centuries.

Again, all the variants of calculations by different methods and with different sets of haplotypes give practically the same figures within the margins of error. The average timespan to a common ancestor for the PALE GREEN subgroup is 430±100 years. It is the end of the 16th century AD, plus-minus four generations.

Since it is a rather recent branch, it is difficult to even suggest their subclade comparing their base haplotype with those of different subclades of R1b1a2. For example, it has as many as

- 33 mutations from the L226 base haplotype,
- 32 mutations from the L23 base haplotype,
- 31 mutations from the “South Irish” base haplotype,
- 30 mutations from the base M222 haplotype,
- 30 mutations from the L51 base haplotype,

- 26 mutations from the base L21 haplotype.
- 25 mutations from the "North Irish" base haplotype,
- 25 mutations from the U152 base haplotype.

In the case of the shortest distance – from the U152 base haplotype (4125±450 years bp) it amounts to 6550 years between the PALE GREEN and U152 common ancestors, which places THEIR common ancestor to 5500±800 years before present. In the case of the longest distance – from the L226 base haplotype (1325±225 years bp) it translates to 9400 years between the PALE GREEN and L226, which places THEIR common ancestor to 5600±800 years before present. It is apparently the same common ancestor, which is likely P312 or L11 subclades. Those are times when R1b1b2 were coming to Europe. There are the roots of the PALE GREEN subgroup. Then it is no surprise that its branch on the haplotype tree is surrounded by "unclassified" ancient branches (Fig. 5).

BLACK

The BLACK branch on the haplotype tree is surrounded by "unclassified" haplotypes, from which it was apparently split. There are six haplotypes in the 67 marker format in this group, two more are determined in the 37 marker format, and two more in the 25 marker format. There are only two surnames among the bearers of this subgroup, most of them (seven) are MacDonalds and McDonalds. Again, because of that it could have been expected that a common ancestor of the group lived rather recently. Indeed, there are only 25 mutations in six of 67 marker haplotypes, that is in 402 alleles, from the base haplotype

12 25 14 10 11 13 12 12 12 12 13 28 – 17 9 9/10 11 11 25 15 20 30 14 15 16 17 –
11 11 19 23 16 15 18 16 37 39 11 12 – 11 9 15 16 8 10 10 8 10 10 12 23 23 16 10 12 12
16 8 13 22 20 13 12 11 13 11 11 12 12

which results in 25/6/0.12 = 35 → 36 generations, that is 900±200 years to a common ancestor of the BLACK subgroup. Eight of 37 marker haplotypes contain 28 mutations, which gives 28/8/0.09 = 39 → 41 generations, that is 1025±220 years to a common ancestor. These figures are within the margin of error, which are rather large due to a small number of haplotypes. The logarithmic method would not be of help here, since in ten of 25-marker haplotypes only one still maintains its base structure, which formally gives [ln(10/1)]/0.046 = 50 → 53 generations, that is 1325 years to a common ancestor. Just one base haplotype more in that series, which statistically can easily happen, would have given [ln(10/2)]/0.046 = 35 → 36 generations, that is 900 years to a common ancestor. This simple example shows that with one of two base haplotype in a dataset the logarithmic method would give a margin of error of around 50-70%.

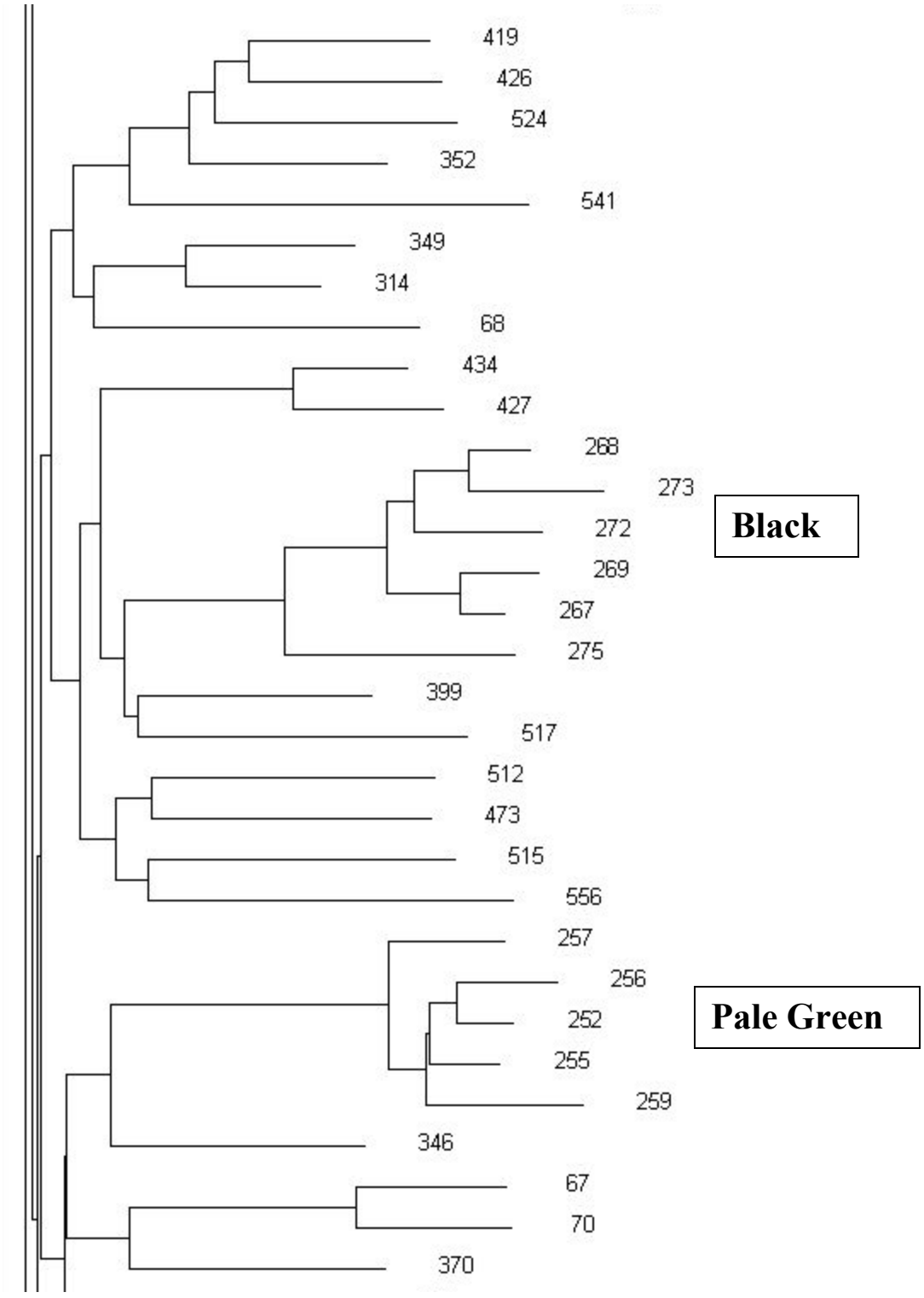


Fig. 5. A fragment of the 67-marker linear haplotype tree, depicting the left-hand side (between 7 and 8 o'clock) branches of the haplotype tree, shown in Fig. 1. "Unclassified" haplotypes have numbers above 308.

A haplotype tree for eight of 37 marker haplotypes of the BLACK subgroup (Fig. 6) shows that the dataset might be not a uniform one. It contains three clusters, or lineages, each with its "local" ancestral (base) haplotype. Their superposition gave that "base" haplotype for the BLACK subgroup. It might be a real one or it might be a "phantom" one. It can rather easily be examined and verified, if needed, using the above haplotype tree and comparing the lineages. It helps that the three lineages are of about the same "weight", so the "overall" base haplotype might be correct.

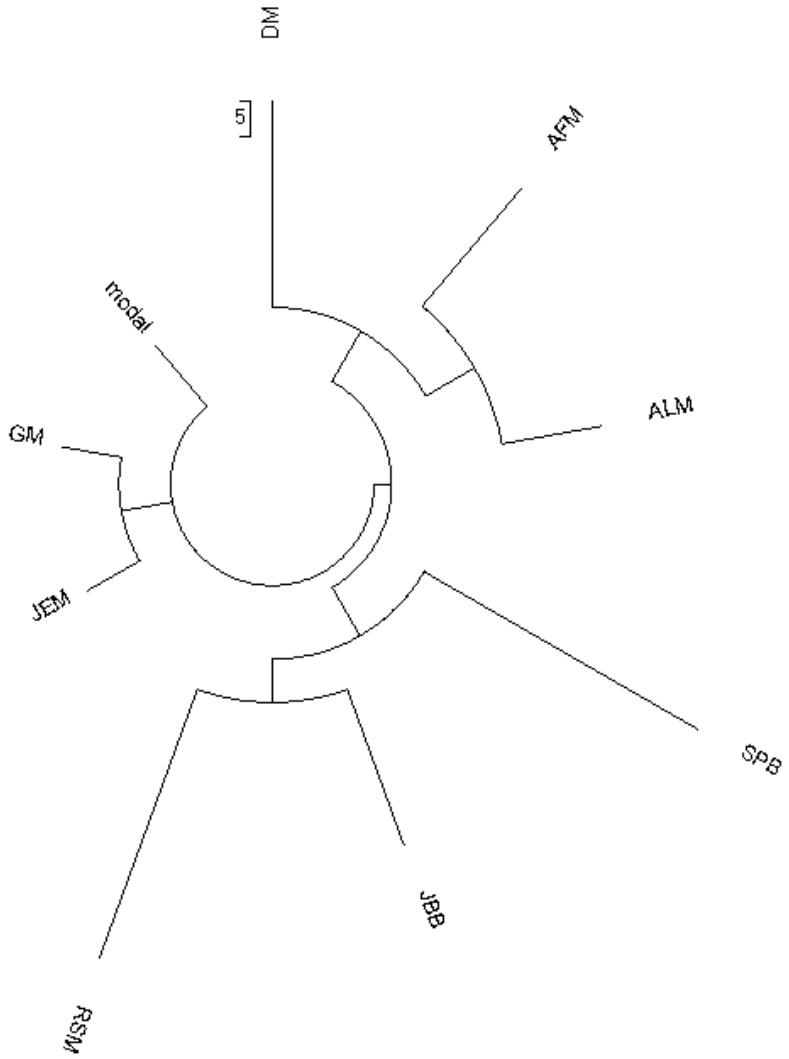


Fig. 6. A 37-marker haplotype tree for the BLACK subgroup.

A rather unusual DYS393 = 12 (the first allele from the left) is typical for R-L23 subclade (~ 6200 years to a common ancestor), or it might have appeared due to a simple, random mutation around a thousand years ago in a common ancestor of the BLACK subgroup (a relatively rare, but quite possible event). The above

by 16.5 mutations from the L23 base haplotype (shown below, mutations are marked):

12 **24** 14 **11** 11 **14** 12 12 12 **13** 13 29 – **16** 9 **10** 11 11 25 15 **19** 30 14 15 16 **18** –
11 11 19 23 **15** 15 **17** **17** 37 **37** **12** 12 – 11 9 15 16 8 10 10 8 10 **11** 12 23 23 16 10 12 12
15 8 **12** 22 20 13 12 11 13 11 11 12 12

Those 16.5 mutations are translated to 4025 years between common ancestors of the BLACK and L23 lineages. This is too little for the L23 subclade which is ~ 6200 year "old", which the BLACK is ~ 900 years "old". Therefore, the DYS393 = 12 was likely just an "ordinary", random mutation. The group might still belong to L23, but initiated from a more recent ancestor, "half-way" from a common ancestor of L23.

However, there are some additional indications that the BLACK subgroup does not belong to L23, and its parent branch has a more common DYS393 = 13. Fig. 5 shows that the BLACK sub-branch is only a half-branch of a wider one containing 14 haplotypes altogether, which have their common ancestor, upstream from the BLACK sub-lineage and belonging to an "unclassified" series of haplotypes. These 14 haplotypes have the following base haplotype:

13 24 14 10 11 14 12 12 12 13 13 29 – 17 9 10 11 11 25 15 19 30 14 15 16 17 –
11 11 19 23 16 15 17 17 37 39 12 12 – 11 9 15 16 8 10 10 8 10 10 12 23 23 16 10 12 12
16 8 13 22 20 13 12 11 13 11 11 12 12

All 14 haplotypes contain 200 mutations from the base haplotype, which gives 200/14/0.12 = 119 → 135 generations from a common ancestor, that is 3375±410 years to a common ancestor. This branch, half of which is "unclassified" is likely an ancestral one (see Fig. 5) to the BLACK subgroup (900±200 years to a common ancestor). Indeed, eight "unclassified" haplotypes surrounding the BLACK sub-branch have 109 mutations in their 67 marker haplotypes, which gives 109/8/0.12 = 114 → 129 generations, that is 3225±450 years from their common ancestor. This is the ancestral branch ("unclassified") for the BLACK subgroup (900±200ybp).

VIOLET, ORANGE, DARK BLUE-GREEN

These subgroups are all parts of "unclassified" branches of haplotypes, as the haplotype tree shows (Fig. 7). Most of those subgroups contain predominantly 25 and 37 marker haplotypes.

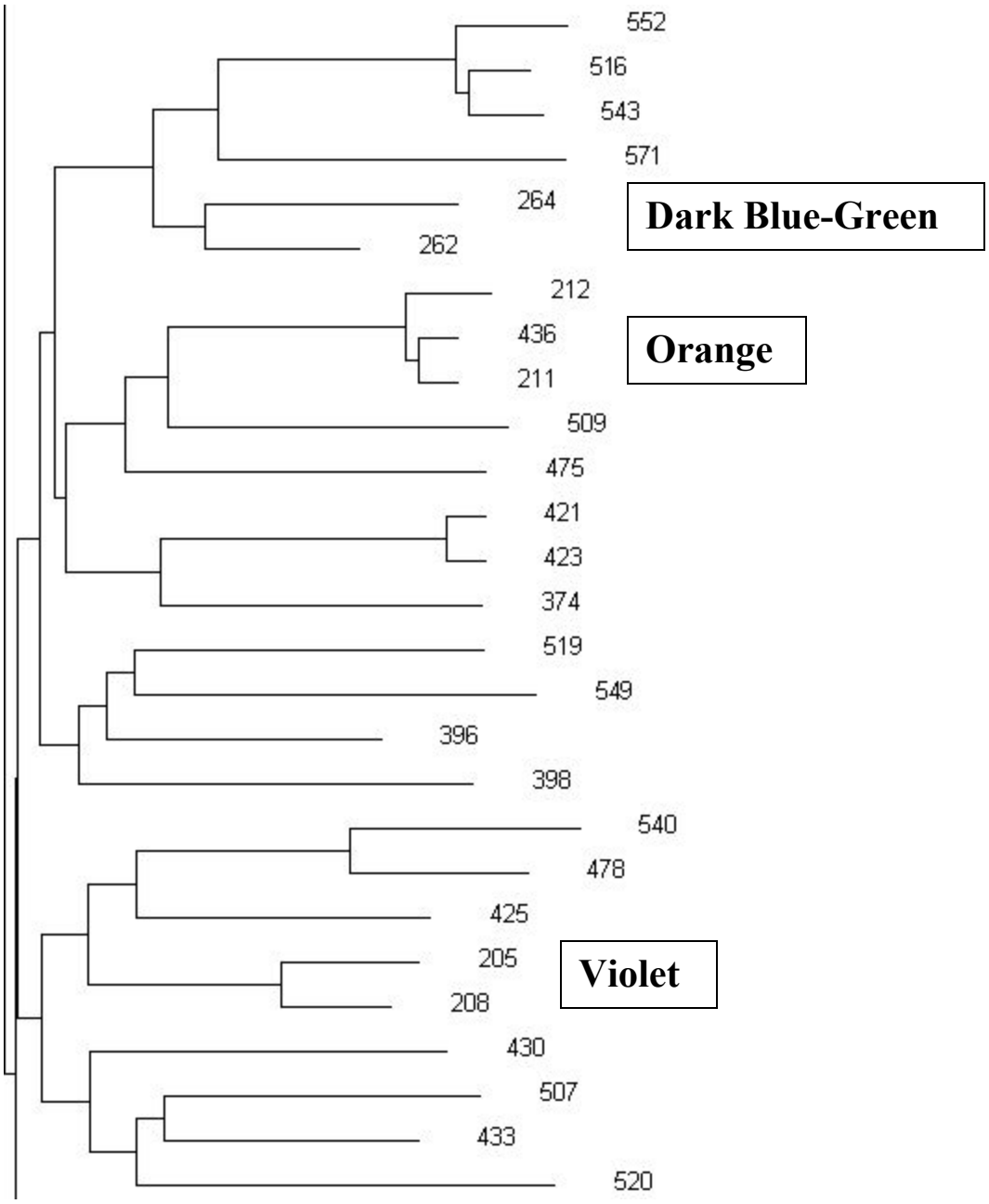


Fig. 7. A fragment of the 67-marker linear haplotype tree, depicting haplotypes at the bottom of Fig. 1. "Unclassified" haplotypes have numbers above 308.

The VIOLET subgroup has four mutations in six 25-marker haplotypes (14.5±7.4 generations), seven mutations in four 37-marker haplotypes (19.4±7.6 generations), four mutations in two 67-marker haplotypes (16.7±8.5 generations), and half of their haplotypes are identical (base) in 25-marker dataset (15.1±8.8 generations), on average 16.4±2.2 generations, that is 410±55 years from their

common ancestor, the beginning of the 17th century AD plus-minus a few generations.

The ORANGE subgroup (including "unclassified" haplotype 436, which in fact is a base haplotype for the ORANGE subgroup) has two mutations in six 25-marker haplotypes (7.2±5.2 generations), five mutations in six 37-marker haplotypes (9.3±4.5 generations), one mutation in three 67-marker haplotypes (2.8±2.8 generations), and four of their six haplotypes are identical in 25-marker dataset (8.8±4.5 generations), on average 175±80 years from their common ancestor, the middle of the 19th century AD plus-minus three generations.

The base haplotype of the ORANGE branch

13 24 14 10 11 14 12 12 **11** 13 13 **29** – **17** 9 10 11 11 **27** 15 19 30 **14** 15 17 **18** – **10 11** 19 **22** 16 15 18 17 **36 39 11** 12 – **11** 9 15 16 8 **11** 10 8 10 10 12 **23** 23 16 10 12 12 15 8 **12 23** 20 **14** 12 11 13 11 11 12 12

significantly differs from that for the RED branch (mutations are shown in bold), on as many as 20 mutations in 67 markers (which is equivalent to 5000 years between them). This places a common ancestor of the RED and ORANGE branches at (5000+1860+175)/2 = 3520 years before present, that is much earlier than the "age" of the RED branch itself (1860±250 years, see above). Understandably, a mutational difference between the ORANGE branch and Red1, Red2 and Red3 sub-branches, as "younger" compared with the RED common ancestor, is even larger – 22, 21 and 21 mutations, respectively. In other words, the ORANGE branch is a rather distant cousin to the RED branch.

Compared to the PINK group, the difference between it and the ORANGE group is even larger – 25 mutations (shown below in the base haplotype of the PINK branch), or 6550 years between their common ancestors.

13 24 14 10 11 **15** 12 12 **12** 13 13 29 – **18** 9 10 11 11 **24** 15 19 **32 15** 15 17 **17** – **11 12** 19 **24 15** 15 18 17 **38** 39 **12** 12 – **12** 9 15 16 8 **10** 10 8 10 10 12 **22** 23 16 10 12 12 **16** 8 **11 22** 20 **13** 12 11 13 11 11 12 12

This brings a common ancestor of the ORANGE and PINK subgroups to (6550+175+675)/2 = 3700 years before present. This is a common time for many R1b1a2 European subclades, such as P312, U106, L21, U152, L2, L20, , L176.2.

The DARK BLUE-GREEN subgroup has eight mutations in five 25-marker haplotypes (35 → 36 generations), 15 mutations in four 37-marker haplotypes (42 → 44 generations), eight mutations in two 67-marker haplotypes (33 → 34 generations), and three base haplotypes out of five in the 25-marker dataset (11 generations). Such a discrepancy comes mainly from one haplotype in the

dataset, which contributed the most into the mutations, and has the only different surname in the group. After it was removed, the pattern was as follows: two mutations in four 25-marker haplotypes (10.9±7.8 generations), 8 mutations in three 37-marker haplotypes (29.6±10.9 generations), eight mutations in two 67-marker haplotypes (33.3±12.2 generations), and three base haplotypes out of four in the 25-marker dataset (6.3±3.7 generations). Technically, most of the figures are within the same margin of error, and an average timespan to the common ancestor of the DARK BLUE-GREEN groups lived 20±13 generations back, that is 500±325 years ago, around the 16th century AD plus-minus three centuries.

GRAY

The branch of the GRAY subgroup has nine mutations in five 25-marker haplotypes (39 → 41 generations), 24 mutations in five 37-marker haplotypes (53 →56 generations), six mutations in two 67-marker haplotypes (25 → 26 generations), and only one base haplotype in five 25-marker haplotype dataset (35 → 36 generations), on average 40±13 generations, that is 1000±325 years from their common ancestor. It is the 11th century AD, plus-minus a few centuries.

WHITE

In this group all eight haplotypes were determined in the 67 marker format. They have a rather recent common ancestor with the MAGENTA group (see Fig. 8).

All the eight haplotypes have only 13 mutations from their base haplotype

13 **25** 14 11 11 14 12 12 **12** 13 13 29 – 17 9 10 11 11 **26** 15 19 28 **14** 15 **15** 17 –
11 10 19 23 **16** 15 18 **16** 37 37 12 12 – 11 9 15 16 8 10 10 8 10 10 12 23 23 16 10 12 12
15 8 12 22 20 13 12 11 13 11 11 12 12

which gives 13/8/0.12 = 14 generations, that is 350±100 years from a common ancestor of the WHITE group.

Eleven mutations marked here in bold are shown in comparison with the L20 subclade base haplotype in the first 25 markers. However, the same number of mutations the WHITE base haplotype also shows in comparison with the L21 base haplotype in the first 25 markers (the same number of mutations was in all the 67 markers):

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

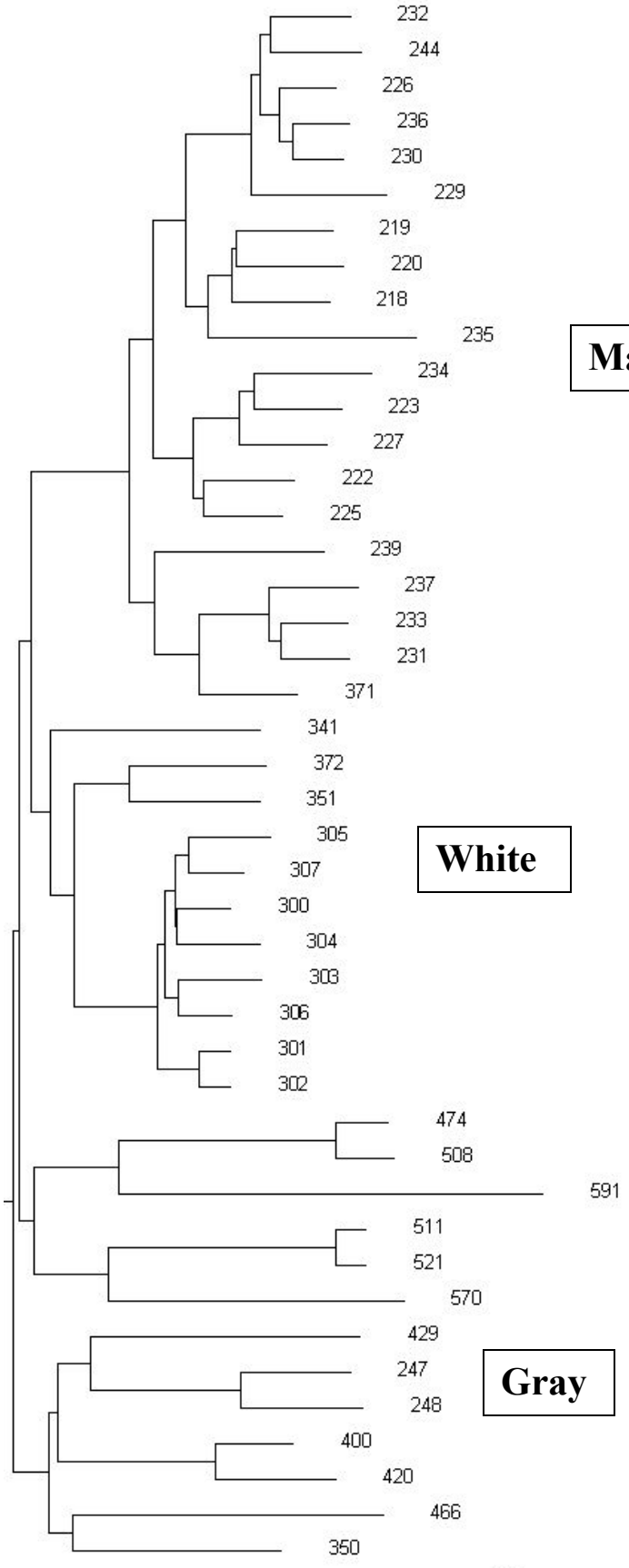

232
244
226
236
230
229
219
220
218
235
234
223
227
222
225
239
237
233
231
371
341
372
351
305
307
300
304
303
306
301
302
474
508
591
511
521
570
429
247
248
400
420
466
350
Magenta
White
Gray

Fig. 8. A fragment of the 67-marker linear haplotype tree, depicting haplotypes at the lower right-hand side of Fig. 1. "Unclassified" haplotypes are numbered above 308.

Therefore, a simple comparison of base haplotype cannot definitely resolve the assigning of the branch in this particular case (as in many other cases). However, in this particular case the WHITE group has descended from R-L21 subclade, as it will be shown in the next section.

MAGENTA

This subgroup was described in the Clan Donald as follows:

Signatures parallel to this group can be found among the McMahons of Fermanagh (one of the territories of ancient Oriel founded by the Collas who allegedly conquered Ulster around 330 AD). This group was one of the tribes from which Gillebride, Somerled's father, sought assistance against the norse.

It is not clear from the above what "signatures parallel to this group" is, and how McMahons of Ferrnanagh could be related to the MAGENTA subgroup. Fig. 8 shows that the MAGENTA group is made up of at least four lineages. Since all of them are of about the same "weight", a timespan to their common ancestor can be calculated using all 20 haplotypes of the group. More accurate calculation could have been done by considering each sub-branch separately, using the same approach as described in this study, and then considering all four base haplotypes for the separate lineages.

All 20 haplotypes of the MAGENTA branch contain 119 mutations from the base haplotype

13 **24 15** 11 11 **15** 12 12 **13** 13 13 29 – 17 9 10 11 11 **25** 15 19 28 **15** 15 **17** 17 –
11 **11** 19 23 16 15 18 **19 36** 37 12 12 – 11 9 15 16 8 10 10 8 10 **9 0 22** 23 16 10 12 12 **16**
8 12 22 20 13 12 11 13 11 11 12 12

which gives 119/20/0.12 = 50 → 53 generations, that is 1325±180 years from a common ancestor of MAGENTA group. Hence, the MAGENTA group common ancestor lived in the 7th century AD, plus-minus a couple of centuries.

In the above haplotype 17 mutations from the adjacent WHITE sub-branch are marked. 19 out of 20 haplotypes in the MAGENTA branch have null mutation in DYS425. It is remarkably close to a R-L21 null mutation (DYS425) branch described in the preceding paper (Klyosov, 2011) with a common ancestor of the

base haplotype with a common ancestor of 1500±325 ybp and the base haplotype (four deviations from the MAGENTA base haplotype are marked):

13 24 **14** 11 11 15 12 12 **12** 13 13 29 – 17 9 10 11 11 25 15 19 28 15 15 17 17 –
11 11 19 23 **15** 15 18 19 36 **38** 12 12 – 11 9 15 16 8 10 10 8 10 9 0 22 23 16 10 12 12 16
8 12 22 20 13 12 11 13 11 11 12 12 (**L21 null DYS425, sub-branch, 1500±325 ybp**)

Therefore, we have two sub-branches, WHITE (350±100 ybp) and MAGENTA (1325±180 years bp), with 17 mutations (equivalent to 4150 years) between their common ancestors. This brings THEIR common ancestor to approximately (350+1325+4150)/2 = 2900 years before present. It is very likely the R-L21 null mutation common ancestor himself (3025±460 years before present) [Klyosov, 2011].

YELLOW GRAY

Fig. 9 suggests that the YELLOW GRAY group descended from the adjacent "unclassified" branch (six haplotypes 472-476). Seven haplotypes of the YELLOW GRAY group contain 34 mutations from the base haplotype

13 24 14 11 11 14 12 12 12 13 13 29 – 16 9 9 11 11 25 15 19 31 15 16 16 17 –
11 11 19 19 17 15 19 17 38 40 13 12 – 11 9 16 17 8 10 10 8 10 10 12 23 23 16 10 12 12
14 8 13 24 22 13 12 11 13 11 12 12 12

which gives 34/7/0.12 = 40 → 42 generations, that is 1050±210 years from a common ancestor of the group.

Shown above is quite an unusual base haplotype compared with, say, the L21 base haplotype, typical for the Isle haplotypes

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

from which the YELLOW GRAY has as many as 23 mutations (marked), which translates to 5950 years difference between their common ancestors. This means that THEIR common ancestor lived approximately (5950+1050+3750)/2 = 5375 ybp. It might be that the YELLOW GRAY lineage arose before their ancestors had arrived to Europe. It could have been Asia Minor or the Middle East, e.g., Sumers.

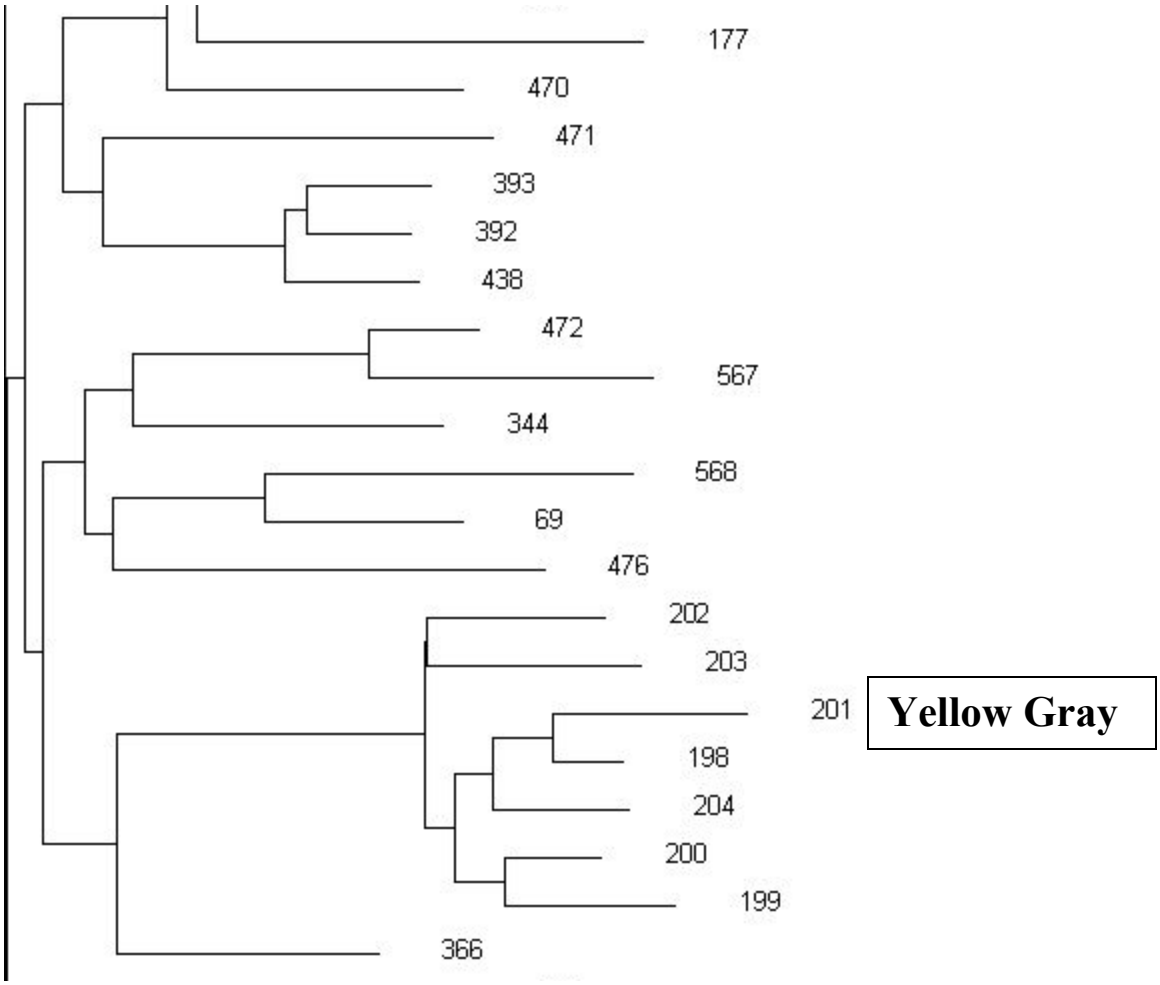


Fig. 9. A fragment of the 67-marker linear haplotype tree, depicting haplotypes at the right-hand side (at 3 o'clock) of Fig. 1. "Unclassified" haplotypes are numbered above 308.

The adjacent "unclassified" branch of six haplotypes has the base haplotype

13 24 14 11 11 14 12 12 12 13 13 29 – 17 9 10 11 11 **25/26** 15 **18** **28/29** 15 15 17 17 – 11 11 19 23 **16** 15 18 17 36 **37** 12 12 – 11 9 **15/16** 16 8 10 10 8 **11** 10 12 23 23 16 10 12 12 15 8 12 22 20 13 **12/13** 11 13 11 **12** 12 12

(7 mutations from the YELLOW GRAY base haplotype are marked, they are equivalent to 1550 years between the two common ancestors; some of those mutations them are fractional ones) with 78 mutations in 67 markers, which gives 78/6/0.12 = 108 →121 generations, that is 3025±460 years from a common ancestor. Therefore, a common ancestor of the YELLOW GRAY branch and the "unclassified" branch lived around (1550+3025+1050)/2 = 2800±400 years before present, which is likely the common ancestor of the "unclassified" branch itself.

PALE VIOLET, PALE BLUE and LIGHT BROWN

These three subgroups are small in the Clan Donald records and contain mainly 25- and 37-marker haplotypes.

Regarding the PALE VIOLET group, the Clan Donald site has suggested:

This group is probably a subset of the Magenta group, but we are not including them together since these people have not tested for DYS425. This group appears to all be descendants of Lt Brian McDonald, chief line of Leinster and Ulster in Ireland who emigrated to Brandywine Creek Delaware in the late 1600s.

Indeed, the base PALE VIOLET haplotype (shown below) is very close to that of MAGENTA, having from it only three mutations (marked)

13 24 15 11 11 15 12 12 13 13 13 29 – **18** 9 10 11 11 25 **14** 19 28 15 15 17 17 –
11 11 19 23 16 15 18 19 36 37 12 12 – 11 9 15 16 8 10 10 8 10 9 0 22 23 16 10 12 12
16 8 12 22 20 **14** 12 11 13 11 11 12 12

It is also close to L21 null mutations sub-branches with common ancestors of 1300±170 and 1100±250 ybp (Klyosov, 2011), with deviations of only 5 mutations from each one

13 24 **14** 11 11 15 12 12 13 13 13 29 – **17** 9 10 11 11 25 **15** 19 28 15 15 17 17 –
11 11 19 23 16 15 18 **18** 36 37 12 12 – 11 9 15 16 8 10 10 8 10 9 0 22 23 16 10 12 12
16 8 12 22 20 **13** 12 11 13 11 11 12 12 (**L21 null DYS425, 1300±170 ybp**)

13 24 15 11 11 15 12 12 **12** 13 13 29 – **17** 9 10 11 11 25 **15** 19 28 15 15 17 17 –
11 11 19 23 16 15 18 19 36 **38** 12 12 – 11 9 15 16 8 10 10 8 10 9 0 22 23 16 10 12 12
16 8 12 22 20 **13** 12 11 13 11 11 12 12 (**L21 null DYS425, 1100±250 ybp**)

One can see that they are generally the same mutations. Practically the same mutations can be noticed also with yet another young sub-branch of null mutated L21 lineages:

13 24 **14** 11 11 15 12 12 **14** 13 13 29 – **17** 9 10 11 11 25 **15** 19 28 15 15 17 17 –
11 11 19 23 16 15 18 **18** 36 37 12 12 – 11 9 15 16 8 10 10 8 10 9 0 22 23 16 10 12 12
16 8 12 22 20 **13** 12 11 13 11 11 12 12 (**L21 null DYS425, 1050±220 ybp**)

The PALE VIOLET group contains only four haplotypes in the 25 marker format, three of them are extended to the 37 marker ones, however, two of the latter are

identical, and only one haplotype in the 67 marker format. Furthermore, the 25-marker haplotypes appear to belong to two different "local" sub-lineages. All of this makes an estimation of a timespan to their common ancestor very problematic, and can give only a general idea. Eight mutations in four 25-markers haplotypes suggest 8/4/0.046 = 43 → 45 generations, that is around 1125±410 years to a common ancestor. Since three of those four haplotypes are mutated, it gives [ln(4/1)]/0.046 = 30 → 31 generations, that is approximately 775 years to a common ancestor. Since the MAGENTA group has a common ancestor who lived 1325±180 years ago, he (or one of his descendants) might indeed have been the common ancestor of the PALE VIOLET group.

In the PALE BLUE group there two mutations in two 25 marker haplotypes, both mutations in the same haplotype. It technically gives 2/2/0.046 = 22 generations, that is 550±390 years from a common ancestor, and [ln(2/1)]/0.046 = 15 generations, that is 375±375 years from a common ancestor. On average, it gives 460±390 years before present.

Regarding the LIGHT BROWN group, the Clan Donald site has noted:

The group is sometimes referred to as "Frisian". It lies within a larger group, not distinguishable from our haplotypes, that has a positive result for the SNP S21/U106.

Let us verify it. The first 25 marker base haplotype of U106 subclade (4175±430 year "old") is

13 23 14 11 11 14 12 12 12 13 13 29 – 17 9 10 11 11 25 15 19 29 15 15 17 17

There are seven LIGHT BROWN haplotypes available, most of them in 25 marker format only (three of them are available in the 37 marker format, and only one in the 67 marker format). Their base haplotype has only three mutations from that of the U106 base haplotype above (marked in bold):

13 23 14 11 11 14 12 12 12 13 13 29 – 17 9 10 11 11 **24** 15 19 29 15 **16** 17 18

All seven haplotypes have 23 mutations from the above base haplotype, which gives 23/7/0.046 = 71 → 77 generations, that is 1925±445 years from a common ancestor. Three mutations between the LIGHT BROWN and the U106 base haplotypes translate to 1725 years between them, which places THEIR common ancestor to 3900±500 years before present. It is likely the U106 subclade common ancestor himself.

BLUE GREEN and DARK GREEN

The BLUE GREEN dataset consists of three 37 marker haplotypes containing six mutations from their base haplotype. It gives 6/3/0.09 = 22 generations, that is 550±230 years to their common ancestor.

Three of 37 marker haplotypes of the DARK GREEN group contain four mutations, which translates to 4/3/0.09 = 15 generations, that is 350±180 years to their common ancestor. Three mutations in their two of 67 marker haplotypes give 3/2/0.12 = 13 generations, that is 325±190 years to a common ancestor. An average timespan is 17±5 generations, that is 425±125 years.

Two "unclassified" branches, likely of R-L2 and R-U106 subclades

These two branches are clearly visible in Fig. 1 in the most upper right-hand side of the haplotype tree. The first one of 14 haplotypes has the following 37 marker base haplotype

13 24 14 11 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** 15 **19** 17 **36** 38 12 12

in which 4.14 mutation from the L2 base haplotype are marked. This corresponds to 1200 years between their common ancestors. All 14 haplotypes contain 122 mutations from the above base haplotype, which gives 122/14/0.09 = 97 → 108 generations, that is 2700±360 years from their common ancestor. Since the L2 common ancestor lived 4025±410 years before present (Klyosov, 2011), a common ancestor of L2 and the "unclassified" branch lived (4025+1200+2700)/2 = 3960±500 years before present. This was very likely the common ancestor of L2 himself.

The adjacent branch of 17 haplotypes (the next branch, clock-wise in Fig. 1) has 218 mutations from the following 37 marker haplotype

13 23 14 11 11 14 12 12 12 13 13 29 – 17 9 10 11 11 25 15 19 29 15 15 17 17 – 11 11 19 23 16 14 18 18 37 39 12 12

It is identical with the U106 base 25-marker haplotype (see description of the LIGHT BROWN group above), and has exactly the same timespan from the common ancestor. For the U106 subclade it is 4175±430 years, for the "unclassified" branch it is 218/17/0.09 = 143 → 167 generations, that is 4175±500 years to the common ancestor.

CONCLUSIONS

A short version of the R1b-R1b1a2 haplotype tree (the full version is given in http://www.isogg.org/tree/ISOGG_HapgrpR.html) is shown below, with the chronology indicated according to (Klyosov, 2011):

R1b M343 — **~ 16,000 ybp**
R1b1 M415, L278
R1b1a L320
R1b1a1 M73, M478
R1b1a2 L265, M269, M520, S3, S10, S13, S17 — **~ 7000 ybp**[1]
R1b1a2a L23/S141, L49**.1** — **~ 6200 ybp**[1]
R1b1a2a1a L51/M412/S167 — **5300±700 ybp**[1]
R1b1a2a1a1 L11, P310, P311 — **~ 4800 ybp (in Europe)**[2]
R1b1a2a1a1a M405/S21/U106 — **4175±430 ybp**
null mutation U106 — **3325±450 ybp**
R1b1a2a1a1b P312/S116 — **4100±415 ybp**
null mutation P312 — **3575±400 ybp**
R1b1a2a1a1b1 ***M65*** — **~ 1800 ybp**
R1b1a2a1a1b2 M153 — **~ 3640 ybp**
R1b1a2a1a1b3 S28/U152 — **4125±450 ybp**[3]
null mutation U152 — **3525±460 ybp**
R1b1a2a1a1b3c L2/S139 — **4025±410 ybp**
R1b1a2a1a1b3c1 L20 — **3650±400 ybp**
R1b1a2a1a1b3d L4/S178 — **1275±290 ybp**
R1b1a2a1a1b4 L21/M529/S145 — **3750±380 ybp**
null mutation L21 — **3025±460 ybp**
null mutation L21 — **1500±325 ybp**
Unidentified clade — **1650±175 ybp**
R1b1a2a1a1b4b M222 — **1450±160 ybp**
R1b1a2a1a1b4e L144 — **~ 4000 ybp**
R1b1a2a1a1b4f **L159.2** — **1775±200 ybp**
R1b1a2a1a1b4g L193 — **1275±170 ybp**
R1b1a2a1a1b4h L226 — **1500±170 ybp**
R1b1a2a1a1b4i **P314.2** — **2225±300 ybp**
R1b1a2a1a1b5 **L176.2/S179.2** — **3675±560 ybp**
R1b1a2a1a1b5a SRY2627 — **3150±320 ybp**
null mutation SRY — **~ 2800 ybp**
R1b1a2a1a1b5b L165/S68 — **~ 3000 ybp**
R1b1a2a1a1b6 **L238/S182** — **~ 1000 ybp**

[1] In Asia
[2] 4575±580 ybp (a different dataset)
[3] 3800±380 (a different dataset)

History of R1b and its subclades migration is outlined in (Klyosov, 2011). While the R1b1a1-M73 subclade mainly left in Asia, the R1b1a2-M269 subclade had moved to the Russian Plain (the Eastern European Plain) around 12-8 thousand years ago, and descendants of its L23 subclade still remains in Russia (with a

common ancestor of 6775±830 years bp [Klyosov, 2009b], though the analysis has both necessarily discriminated the L23 subclade and some upstream sunclades) and in the Caucasus, with their common ancestors who lived there at least 6000 years before present. The M269 and L23 bearers split, some of them went through the Caucasus to Asia Minor/Anatolia (a common ancestor there of 6000±820 years bp), to the Middle East (a common ancestor in Lebanon of 5200±670 years bp) and then via the North African route arrived to Iberia around 4800 years bp. From there, it moved up North to the continental Europe, and it was considered in (Klyosov, 2011).

As a result, a pattern of the Clan Donald R1b1b2 lineages is rather complicated. Here is a partial list of the Donald subgroups for which some lineages were traced to specific subclades:

Red	L21	1860±250 ybp
Pink	L21	675±170
Brown	L21	675±190
Maroon	L21	285±80
Green	L21→M222	1850±210
Dark Blue	L21 (?)	575±165
Blue Green		550±230
Yellow Gray		1050±210
Violet		410±55
Orange		175±80
Magenta	L21	1325±180
Pale Violet	L21	1125±410
Gray		1000±325
Pale Green	U152 (?)	550±160
Dark Blue – Green		500±325
Black		900±200
Pale Blue		460±390
Light Brown	U106	1925±445
Dark Green		425±125
Yellow Green	L21	290±80
White	L21	350±100
"Unclassified"	L2	2700±360
"Unclassified"	U106	4175±500

MATERIALS and METHODS

This section is described in detail in the preceding article (Klyosov, 2011).

References

Klyosov, A.A. (2009a). DNA Genealogy, mutation rates, and some historical evidences written in Y-chromosome. I. Basic principles and the method. J Genetic Genealogy 5, 186-216.

Klyosov, A.A. (2009b) DNA Genealogy, mutation rates, and some historical evidences written in Y-chromosome. II. Walking the map. J. Genetic Genealogy. 5, 217 – 256.

Klyosov, A.A. (2010a) Irish Haplotypes and haplogroups. Proc. Russian Academy of DNA Genealogy (ISSN 1942-7484), 3, № 6, 1029-1053.

Klyosov, A.A. (2010b) Haplogroup R1b. Part 1. Proc. Russian Academy of DNA Genealogy (ISSN 1942-7484), 3, No.2, 249-299.

Klyosov, A.A. (2010c) Haplogroup R1b. Part 2. Proc. Russian Academy of DNA Genealogy (ISSN 1942-7484), 3, № 3, 406-475.

Klyosov, A.A. (2011) Haplotypes of R1b1a2-P312 and related subclades: origin and "ages" of most recent common ancestors. Proc. Russian Academy of DNA Genealogy (ISSN 1942-7484), 4, № 6, 1127-1192

MacDonalds and Scottish Haplotypes of Haplogroup I

(an updated version)

Anatole A. Klyosov

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

This article concludes a series of articles on the Donald Clan haplotypes of haplogroups R1a1, R1b1b2, and now on I1 and I2. The Donalds Clan is a good representative of Scottish lineages, and in many cases its haplotypes go beyond Scotland, merging with a number of West- and East-European lineages. The Donald Clan haplotypes of haplogroups I ("unclassified"), I1, I2a, and I2b1 (the site organizers use the obsolete nomenclature as Ia, Ib, and Ic, respectively) haplogroups were taken from the Clan Donald USA Project (http://dna-project.clan-donald-usa.org/tables.htm). The "I" list there contains 82 haplotypes, 59 of them are in the 37 marker format, and only 31 in the 67 marker format.

Haplotype trees of those 67 and 37 marker haplotypes are shown in Figs 1 and 2. The numbering has been done in according with the listing of haplogroup I haplotypes in the same order as they were presented by the beginning of September, 2010, as follows:

I ("unclassified")	1-6
I1-M253	7-53
I2a-P37.2	54-63
I2b1-M223	64-82

It have be noticed here that I1-M253 (and its downstream subclades such I1b-M227, I1b1-M72, I1d-L22, I1d1-P109) has the highest frequency in the European North-West (Scandinavia and the Isles). In the Isles it is often associated with Viking and/or Anglo-Saxon "invaders". I2a-P37.2 is considered as mainly the "Balkan" and Mediterranean haplogroup. Unlike I2a, its "parallel" cousin I2b-M436 and its subclade I2b1 occur mainly in the Isles and North-West continental Europe.

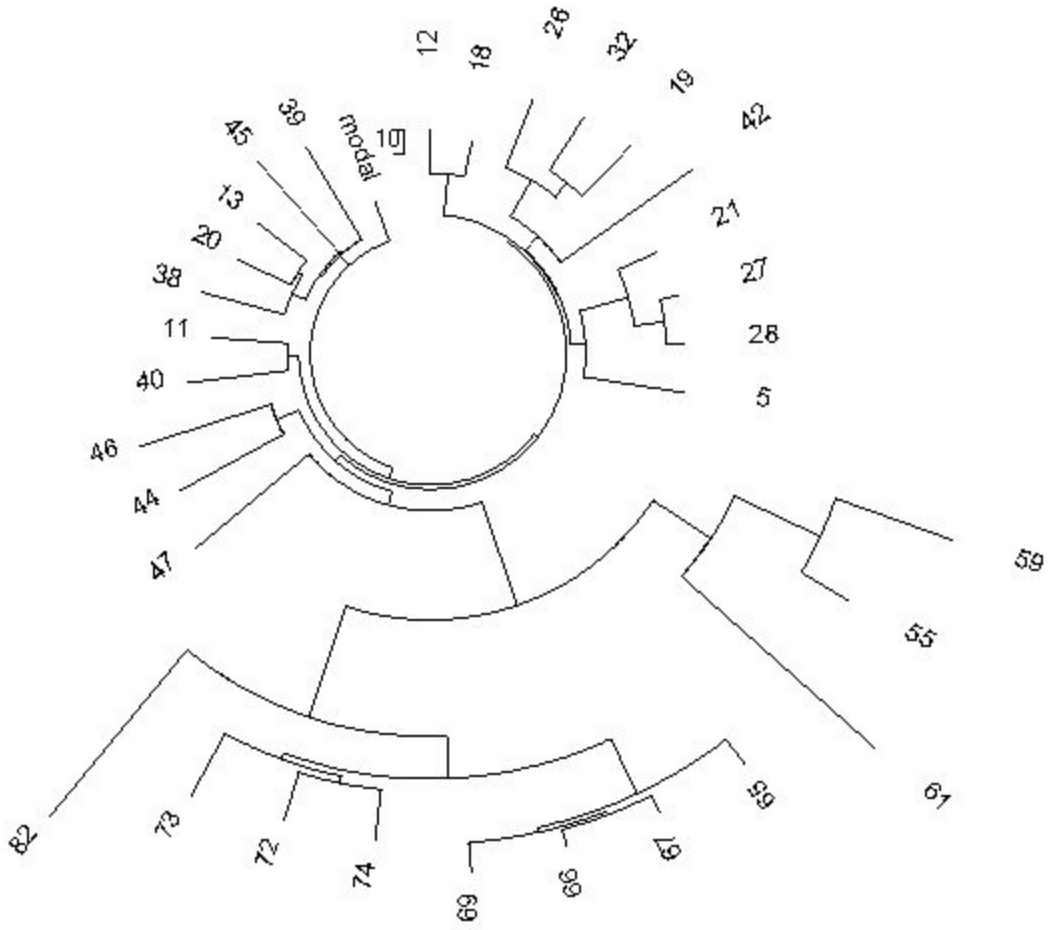


Fig. 1. A 67-marker haplotype tree for 31 haplotypes of I haplogroup of the Clan Donald USA project http://dna-project.clan-donald-usa.org/tables.htm **The tight branch on top of the tree is of the subclade I1, a small branch of three haplotypes on the lower right is of subclade I2a, and the flat branch at the bottom is of subclade I2b1.**

A 67 marker haplotype tree typically gives a better resolution of branches compared with a 37 marker tree, however, a 37 marker tree typically contains more haplotypes. That is why in our methodology the 67 marker tree is used for identification of branches, and the lower-marker tree (the respective dataset) is used for calculations, since it provides better statistics. Haplotype trees in Fig. 1 and 2 perfectly fit to that strategy, particularly because they show exactly the same branches. The branches are resolved so clearly, that there is not any "flipping" of haplotypes between them. It is of no surprise here, since each of the branches corresponds to a different subclade, each with a very different haplotype structure (in terms of their alleles).

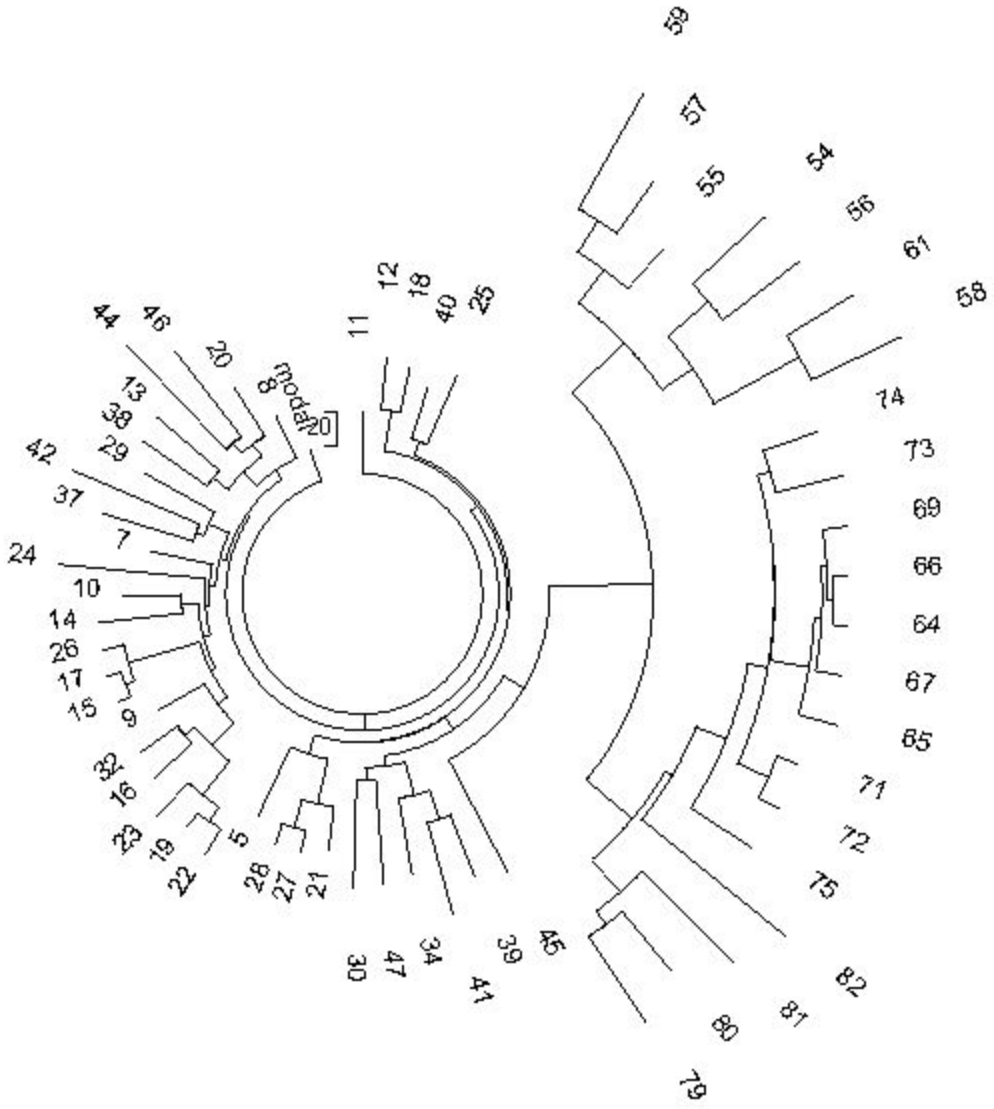


Fig. 2. A 37-marker haplotype tree for 59 haplotypes of I haplogroup of the Clan Donald USA project http://dna-project.clan-donald-usa.org/tables.htm **The tight branch on left, top and bottom belongs to the subclade I1, a small branch of seven haplotypes on the upper right is of subclade I2a, and the flat branch on the lower right-hand side is of subclade I2b1.**

I1-M253 haplogroup, subclade I1d1-P109

Fig. 2 shows that the left-hand side I1 branch consists of a wide branch of 22 haplotypes, flanking with a number of assorted mini-branches or scattered haplotypes, making them total of 37. To make sure that we are not missing anything significant, we have considered the both scenarios. The 22 haplotype I1

branch contain 238 mutations from the first 37 markers in the following base haplotype:

13 23 14 10 14 14 11 14 11 12 11 28 – 15 8 9 8 11 23 16 20 28 12 14 15 16 –
10 10 19 21 14 14 17 **20** 35 37 12 10 – 11 8 15 15 8 11 10 8 9 9 12 22 25 15 10 12 12 16 8 13 25 20 13 13 11 12 11 11 12 11

The mark in bold shows the only one mutation from the base I1d1-P109 haplotype in its 37 marker format, which is a subclade of I1-M253. Apparently, those I1-M253 individuals on the Clan Donald list were "undertyped" to a deeper subclade, which is I1d1 in this particular case. The number of mutations give 238/22/0.09 = 120 generations without a correction for back mutations, or 136 generations with a correction (Klyosov, 2009), that is 3400±405 years to a common ancestor of the branch.

Only 9 haplotypes from these 22 are 67 marker haplotypes, and they have 156 mutations from the above base haplotype. It gives 156/9/0.12 = 144 → 168 generations, that is 4200±540 years to a common ancestor, within margin of error with that for 37 marker haplotypes.

When all 37 of 37-marker haplotypes are considered, they have a slightly different base haplotype (apparently, the flanking haplotypes belong to yet different subclades and/or different local lineages)

13 **22** 14 10 **13** 14 11 14 11 12 11 28 – 15 8 9 8 11 23 16 20 28 12 14 15 16 –
10 10 19 21 14 14 **16** 20 35 37 12 10 – 11 8 15 15 8 11 10 8 9 9 12 **23** 25 15 10 12 12 16 8 13 25 20 13 13 11 12 11 11 12 11

with 420 mutations from it in the first 37 marker haplotypes. It gives 420/37/0.09 = 126 → 145 generations, that is 3625±400 years from their common ancestor.

While the following "signature" in the first 12 markers

13 **23** 14 10 **14** 14 11 14 11 12 11 28

is a typical one for the Scandinavian I1d haplotypes (from Denmark, Sweden, Norway, Finland) (Klyosov, 2010, and it is seen in the Clan Donald branch of the 22 haplotypes in Fig. 2, the "signature'

13 **22** 14 10 **13** 14 11 14 11 12 11 28

is a typical for the Isles I1d haplotypes (England, Ireland, Scotland) and Central and Eastern haplotypes (see below) [ibid.]. In other words, at the extension of the branch to include the flanking haplotypes "the Isles" lineages became prevailing

over the Scandinavian lineages, and the base haplotype tipped to "the Isles" side. It is of interest that a timespan to common ancestors for both "Scandinavian" and "the Isles" are practically equal to each other, namely, 3375±350 and 3425±350 years (Klyosov, 2010), and they are practically equal to the time span to the Donald I1d1 common ancestor, 3400±405 years. Obviously, something has happened before the middle of the 2nd millennium BC (or before that), which results in collapse the I1 lineages in Europe. Only in the middle of the 2nd millennium BC a common ancestor of the I1d1 has appeared, and his descendants re-populated Europe, from the Atlantic to Eastern Europe. The same base haplotypes are identified in Central Europe

13 **22** 14 10 **13** 14 11 14 11 12 11 28

(Austria, Belgium, Netherlands, France, Czech, Hungary, Italy, Greece, Romania, Spain, Switzerland), with 3425±350 years to their common ancestor [Klyosov, 2010] (it is "the Isles" signature, which, however, can be the Central European one in the Isles),

in Eastern Europe

13 **22** 14 10 **13** 14 11 14 11 12 11 28

(Poland, Ukraine, Belarus, Estonia, Lithuania, Russia), with a common ancestor of 3225±360 years before present (Klyosov, 2010),

in Germany

13 **22** 14 10 **13** 14 11 14 11 12 11 28

with a common ancestor of 3225±330 years before present (Klyosov, 2010),

in the Middle East (Jordan, Lebanon, Turkey, Jewish haplotypes)

13 **23** 14 10 **13** 14 11 14 11 12 11 28

with a common ancestor of 3475±480 years before present (Klyosov, 2010). It is interesting that the Middle-Eastern I1d1 haplotype is a superposition of the "Scandinavian" one and the Central European – the Isle base haplotypes.

A more extended dataset (41 of 67 marker haplotypes) has revealed two practically equal in size sub-branches of I1 haplotypes in Eastern Europe (Fig. 3)

Fig. 3. A 67-marker haplotype tree for 41 haplotypes of I1 haplogroup of mainly Eastern Europe (Russia, Poland, Lithuania, Czech, Bulgaria, Germany, Estonia, Romania), haplotypes collected and kindly provided by Igor Rozhanskii.

The 22 haplotype right-hand side of the tree in Fig. 3 has the following base Eastern European haplotype

13 **22** 14 10 **13** 14 11 14 11 12 11 28 – 15 8 9 8 11 23 16 20 28 12 14 15 16 – 10 10 19 21 14 14 16 20 35 37 12 10 – 11 8 15 15 8 11 10 8 9 9 12 23 25 15 10 12 12 16 8 13 25 20 13 13 11 12 11 11 12 11

It is identical in all the 67 markers to the left-hand side branch (Fig. 2) of the Clan

All 22 haplotypes of the Eastern European branch contain 109 mutations in the first 25 markers, which gives 109/22/0.046 = 108 → 121 generations from a common ancestor, that is 3025±420 years. It is the same value within the margin of error with the Donald Clan timespan to their common ancestor (3625±400 years), though might be a little younger (or a little older, within the same margin of error). Second, the left-hand side branch of 19 haplotypes, contains 99 mutations in the first 25 markers of the following base haplotype

13 22 14 10 13 14 11 14 11 12 11 28 – 15 8 9 8 11 23 16 20 28 12 14 15 16 –
10 10 19 21 14 14 16 **19** 35 **36** 12 10 – 11 8 15 15 8 11 10 8 9 9 12 **22** 25 **16** 10 12 12 16 8 13 25 20 13 13 11 12 11 11 12 11

which gives 99/19/0.046 = 113 → 128 generations, that is 3200±460 years from a common ancestor, again within the margin or error with the above figures. There are 4.37 mutations between these two base Eastern European haplotypes, if considering average alleles with their decimal values, which results in only 775 years – cumulatively – between them, and places THEIR common ancestor to 3510±500 years before present, that is within the same margin of error. In other words, it is the same common ancestor of the European I1d haplogroup, including the Clan Donald common ancestor of the considered dataset in this subclade.

I2a-P37.2 subclade

There are only three 67 marker Donald Clan haplotypes in this subclade (Fig. 1), and seven haplotypes in the 37 marker format (Fig. 2). Obviously, it is much less populous subclade among the Donald Clade, as it is in fact in Europe. The I2a branch among the Donalds has the following base haplotype:

13 24 15 11 12 16 11 13 11 13 11 30 – 18 8 10 11 11 25 15 20 30 12 14 14 15 –
10 10 21 21 17 13 18 18 35 36 12 10 – 11 8 15 16 7 11 10 8 13 11 12 22 22 17 10 12 12 15 8 12 24 20 13 12 10 13 10 11 12 11

Those seven haplotypes in the upper right-hand branch in Fig. 2 have 84 mutations in their 37 marker haplotypes, which gives 84/7/0.09 = 133 → 154 generations, that is 3850±570 years to their common ancestor.

The above base haplotype differs by as many as 56 mutations with that of I1d1 subclade (see above) in the 67 marker format, which corresponds to 20650 years between the two base haplotypes. This time is required for 56 mutations to occur in a 67-marker haplotype. Their common ancestor lived in Europe around (20650+3625+ 3850)/2 = 14,000 years ago. This figure is not quite accurate, since

it results from a comparison of two base haplotypes of different subclades, however, it shows a scale of time-related differences between I1 and I2 subclades.

It seems that the Clad Donalds lineage split from a much more ancient lineage with a common ancestor of 6675±910 years before present (Klyosov, 2010). His base haplotype in the first 25 markers was only four mutations away from the above base haplotype (mutations are marked in bold):

13 24 15 11 12 16 11 13 11 13 11 30 – **16** 8 10 11 **12** 25 15 20 **29** 12 14 14 15

I2b1-M223 subclade

The 14-haplotype flat branch on the right-hand side in Fig. 2 contains 141 mutations in the first 37 markers from the following base haplotype:

15 24 15 10 15 16 11 13 11 13 12 29 – 16 8 9 11 11 26 15 20 28 11 11 14 15 –
11 10 19 21 15 14 18 18 31 37 12 10 – 12 8 15 16 8 12 10 8 10 9 0 19 21 16 11 12 12 17
9 14 25 20 9 13 11 13 11 11 12 11

which gives 141/14/0.09 = 112 → 126 generations, that is 3150±410 years from their common ancestor. It is again practically the same figure, within the margin of error, as in all cases of common ancestors of I1 and I2 subclades considered above. The base haplotypes of I2a and I2b1, shown above, differ by as many as 57 mutations in all 67 markers, which translates to 21,250 years between them. This places THEIR common ancestor to approximately (21250+3150+3850)/2 = 14,100 years before present.

Conclusion

The Clan Donald branches of haplogroup I (subclades I1, I2a and I2b1) are well resolved, and their analysis is rather straightforward and unambiguous. It shows that all the three branches descended from a common ancestor who lived around 14 thousand years ago, its (apparently numerous) descendants practically disappeared from Europe at some time which is totally unknown, and reappeared again as three different subclades all in the middle of the 3rd millennium BC. Some of them migrated to the Isles from Scandinavia, some from Central and/or Eastern Europe. It was a relatively mass-migration, therefore common ancestors of the migrants/invaders to the Isles cannot be differentiated from common ancestors on the European continent. Any particular lineages which could have been assigned to the Vikings or another particular tribe do not show up in the Clan Donald haplogroup I (and its subclades) lineages.

References

Klyosov, A.A. (2009). DNA Genealogy, mutation rates, and some historical evidences written in Y-chromosome. I. Basic principles and the method. J Genetic Genealogy 5, 186-216.

Klyosov, A.A. (2010) Haplogroup I. Proc. Russian Academy of DNA Genealogy (ISSN 1942-7484), 3, No.1, 96-158.

The slowest 22 marker haplotype panel (out of the 67 marker panel) and their mutation rate constants employed for calculations timespans to the most ancient common ancestors

Anatole A. Klyosov
http://aklyosov.home.comcast.net

Abstract

The slowest 22 markers from the 67 marker panel have the following mutation rate constants (MRC):

Marker	MRC per marker per "conditional" generation of 25 years
472	0.00001
425	0.00005
436	0.00006
426	0.00009
490	0.00013
454	0.00016
455	0.00016
578	0.00017
641	0.00017
590	0.00017
594	0.00020
388	0.00022
492	0.00023
395S1b	0.00025
450	0.00029
617	0.00050
531	0.00050
640	0.00051
392	0.00052
568	0.00053
395S1a	0.00053
438	0.00055
∑	0.00600/haplotype/generation (25 years)

(This is a shorter version of the article published in Russian in Proceedings of

The literature does not contain values of mutation rate constants for these markers except of three of them determined by Chandler (2006) for DYS426, DYS388 and DYS392 (0.00009, 0.00022, and 0.00052 mutation/marker/generation, respectively). Values reported by Ballantyne et al (2010), based on a number of transmissions and mutations between almost 2000 father-son pairs were very unreliable, because 14 of these 22 markers did not show any mutation (that is, according to Ballantyne et al, the mutation rate constants were lower than 0.0006 mutations per generation); 7 markers showed only one mutation each, and the authors gave the respective values as 0.00058, 0.00057, 0.00059, 0.00059, 0.00061, 0.00063 and 0.00065, which, of course, is the same thing for all of them, besides, with just one mutation the margin of error even for one sigma (68% reliability) will be 100%, and for two sigma (95% reliability) the margin of error will be about 200%, which does not make sense in this context. Only in one marker from these 22 there were two mutations in almost 2000 father-son pairs, and the authors defined the MRC as 0.00113 mutations per marker per generation. The margin of error for two mutations in the system would be ±71% for a 68% reliability, and ±141% for a 95% reliability, however, it is not serious even to mention the 95% reliability in this situation.

We have determined mutation rate constants for all these 22 markers in a system in which most of them showed dozens of mutations, except the three slowest markers for which a number of mutations were between 4 and 22. It was shown that the literature typically shows noticeably higher values of mutation rates than they should be, because authors commonly count mutations "across" branches and subclades of haplotypes, thereby overcounting a number of mutations. For example, several haplotypes with DYS388=10 among a majority of DYS388=12 haplotypes in haplogroup R1a1 would wrongly bring more mutations for DYS388 marker and wrongly increase the "mutation rate constant" for DYS388. In reality, DYS388 form a separate branch in a haplotype tree, in which all (or almost all) haplotypes have DYS388=10, and a number of mutations in this marker would be much less. It would be incorrect to count mutations "across the haplotype tree" or across the haplotype dataset.

The Table above shows that the total (summarily) mutation rate constant for all the 22 markers equals to 0.00600 mutation/haplotype/generation, or 0.00027 mutation/marker/generation (of 25 years per generation). These values were employed for calculations of chronology of the haplogroup tree for Y chromosome, as described in the next article of this issue.

Introduction

There are two principal ways for calculations of TSCA (time span to a [nearest] common ancestor). One way is using haplotype datasets, containing as many haplotypes as possible. In this case the dataset is subdivided into branches, each one of them having its own common ancestor, and a pattern of mutations is analyzed using the linear (mutation-counting) and logarithmic (base haplotype counting) methods. Also, the quadratic and the permutational methods can be used for calculations, as explained in the preceding papers in this issue, as well as in many publications by the author of this study. As a result, two principal parameters are obtained: the base haplotype for each lineage (branch) and a TSCA for each lineage (branch). For example, in the first article in this issue it was shown, that 141 haplotype dataset of the P312-L176.2-SRY2627 in the 67 marker format contain 1896 mutations from their base haplotype

13 24 14 11 11 14 12 12 12 13 13 29 – 17 9 10 11 11 25 15 19 29 15 15 17 18 –
11 11 19 23 16 15 18 17 37 38 12 12 – 11 9 15 16 8 10 10 8 10 10 12 23 23 16 10 12 10
16 8 12 22 20 13 12 11 13 11 11 12 12 **(P312-L176.2-SRY2627, 3150±320 ybp**)

which gives 1896/141/0.12 = 112 → 126 generations, that is 126x25 = 3150 years to their common ancestor. The margin of error is given above. 0.12 here is the mutation rate constant for 67 marker haplotypes.

In the same manner it was calculated that a common ancestor of R1b1a2-P312-L176.2 lived 3675 ybp (years before present), and had the base haplotype

13 24 14 11 11 14 12 12 12 13 13 30 – 17 9 9 11 11 25 15 19 29 15 15 17 17 –
11 11 19 23 16 14 19 16 37 38 13 12 – 11 9 15 16 8 10 10 8 10 10 12 23 23 16 10 12 12
15 8 12 22 21 14 12 11 13 11 11 12 12 (**R1b1a2-P312-L176.2, 3675±560 ybp**)

and a common ancestor of P312-L176.2-L165 lived ~ 3000 ybp

13 25 14 11 11 14 12 12 13 13 13 29 – 18 9 10 11 11 25 15 19 30 14 15 16 17 –
11 11 19 23 16 15 18 17 38 38 12 12 – 11 9 15 16 8 10 10 8 10 10 12 23 24 16 10 12 12
15 8 12 22 20 14 12 11 13 11 11 12 12 (**L165, ~ 3000 ybp**)

These TSCA chronology values fit the phylogeny tree ladder, P312 → L176.2 → SRY2627/L165.

The second principal way for calculating TSCAs considers not haplotype datasets, but base haplotypes. For example, three base haplotypes shown above collectively have 20 mutations between them. In other words, they three represent a dataset having 20 mutation from a deduced base haplotype of THEIR the most recent common ancestor, presumably of the R1b1a2-P312 subclade.

These 20 mutations separate THEIR common ancestor not from the present time (as typically holds when TSCAs are calculated from haplotypes of our contemporaries) but from an average time of the three TSCAs, shown above, that is from (3675+3150+3000)/3 = 3275 ybp.

20 mutations in three base haplotypes translate into 20/3/0.12 = 1500 years plus their average "age", which is 3275 years, which gives 4775 years from a common ancestor of all the three subclades. It is presumably the R1b1a2-P312 himself.

It is important that present-day R1b1a2-P312 haplotype datasets results in TSCAs in the range of 3600 to 4100 ybp, however, a reconstruction of a TSCA for P312 from base haplotypes of different branches and subclades reveals a TSCA for the P132 subclade as around 4800 ybp. It seems that ancient P312 haplotypes went through a population bottleneck, and their common ancestor can be detected more recently compared with the more proper and more ancient date.

Another example regarding a TSCA for P312. If we compare the base haplotypes of P312 and U106, which are "parallel" subclades of the R1b1a2 haplogroup tree

13 24 14 11 11 14 12 12 12 13 13 29 – 17 9 10 11 11 25 15 19 29 15 15 17 17 –
11 11 19 23 15 15 18 17 36 38 12 12 – 11 9 15 16 8 10 10 8 10 10 12 23 23 16 10 12 12
15 8 12 22 20 13 12 11 13 11 11 12 12 **(P312, 4100±415 ybp)**

13 **23** 14 11 11 14 12 12 12 13 13 29 – 17 9 10 11 11 25 15 19 29 15 15 17 17 –
11 11 19 23 **16** 15 **17** 17 **37** **39** 12 12 – 11 9 15 16 8 10 10 8 10 10 12 23 23 16 10 12 12
15 8 12 22 20 13 12 11 13 11 11 **13** 12 **(U106, 4175±430 ybp)**

they differ by six mutations (marked in the U106 base haplotype). It is important that the two base haplotypes which have common ancestors each who lived at the same time, differ by 6/0.12 = 50 → 53 "conditional" generations (25 year per generation), that is 1,325 years between their common ancestors. The sign → means a correction for back mutations. Since a common ancestor of U106 and P312 lived 4175±430 ybp and 4100±415 ybp respectively, THEIR common ancestor lived (1325+4175+4100)/2 = 4800 years ago.

Such an approach is rather convenient when a common ancestor lived several thousand years ago. However, when we move to tens of thousand years deep, or even to 100+ thousand years, we see too many mutations between base haplotypes to be accurately analyzed. Relatively rapidly mutated markers do not seem to be appropriate for very ancient times for common ancestors.

Therefore, in this study we employ the same approaches described above, however, using 22 very slowly mutated markers.

22 slowest marker haplotype panel

The 22 slowest markers (with respect to their mutation rates) are shown in Table 1 in the order of the FTDNA notation, along with their values collected from the literature. The first three markers are from the 1-12 FTDNA panel, the following two markers are from the 13-25 panel, the next one from the 26-37 panel, and the rest are from the 38-67 marker panel. In Table 1 they are divided into the respective sections as indicated.

Table 1.
Mutation rate constants per marker per generation. Values shown in the last two columns were calculated or shown as a ratio of the number of mutations to the number of transmissions between father-son pairs. n/a - not available.

Marker	Mutation rate constant, per marker per generation		
	Chandler, 2006	Ballantyne et al, 2010	Burgarella et al, 2010
426	0.00009	(0/1735)	n/a
388	0.00022	(0/1636)	0.00042
392	0.00052	0.00058	0.00043
455	0.00016	(0/1618)	n/a
454	0.00016	(0/1458)	n/a
438	0.00055	0.00057	0.00043
531	n/a	0.00059	(0/483)
578	n/a	0.00059	(0/403)
395S1a	n/a	n/a	n/a
395S1b	n/a	n/a	n/a
590	n/a	(0/1780)	(0/403)
641	n/a	(0/1768)	(0/403)
472	n/a	(0/1549)	(0/403)
425	n/a	0.00113	n/a
594	n/a	0.00061	(0/403)
436	n/a	(0/1798)	n/a
490	n/a	(0/1759)	(0/403)
450	n/a	0.00063	n/a
617	n/a	(0/1684)	(0/403)
568	n/a	0.00065	(0/403)
640	n/a	(0/1716)	0.00496

492	n/a	(0/1770)	(0/403)

The Table 1 exposes many unclear and dubious things. For majority of the markers their mutation rates either have not been measured, or were not observed, even for as many as almost two thousand father-son pairs. In those cases in which mutations have been observed, there were only one, or two, or three mutations per 1500-1800 father-son pairs. This can result in only rough estimates, such as 0.0006-0.00007, 0.0011-0.0013, 0.0017- 0.0020, respectively. Margins of error in the last two columns in Table 1 are equal about 58-200%, because for one mutation the margin of error is 100-200% for 68%-95% reliability (one and two sigma, respectively), for two mutations it is 70-140%, for three mutations it is 58-115%, respectively. Therefore two or even three decimal digits given in the mutation rate constants in Table 1 do not make any certain sense; for that, their margins of error should have been in the rage of 10-1% or 1-0.1%, respectively. In other words, to indicate 0.00496 as a mutation rate constant for DYS640 is a huge overestimation of precision, when in reality it might have equally been 0.005, 0.004, 0.006, since the authors (Burgarella et al, 2010) have recorded only two mutations in 403 father-son pairs. However, with 2/403 = 0.00496, and with margin of error of 70% (one sigma, 68% reliability) or 140% (two sigma, 95% reliability), it can in reality be any number between 0 and 0.07. Indeed, our data showed that this marker produced 78 mutations per 153,600 transmissions in a series of 1024 R1b1a2-L21 haplotypes, which resulted in 0.00051 mutation/marker/generation (see below). This is close to what Little (2008) has estimated, 0.00047 mut/mark/generation.

In order to examine and verify the data in Table 1 and to compose of a more justified set of mutation rate constants, if it is justified indeed, we have determined a number of mutations in extended series of 67 marker haplotypes. We have considered three haplotype series: 2299 haplotypes of R1b1a2-P312, 1024 haplotypes of its subclade R1b1a2-L21, and 1198 haplotypes of R1a1. The resulted mutation rate constants are listed in Table 2. A careful analysis of the data obtained included whether or not an elevated number of mutations could have come from "across subclades" or "across branches/lineages", whether they seem to be random, how the respective branches looked on the haplotype tree, etc. Besides, comparisons with data already available in the literature have been made, and possible reasons for discrepancies were considered. Some particular cases of unjustified "excessive number of mutations", which have been removed from calculations, are listed in notes to Table 2.

Fig. 1 shows a tree of 352 of 67 marker haplotypes of haplogroup R1a1. They were taken randomly from a dataset of 1198 haplotypes, which were analyzed in Table 2. Why those 352 haplotypes only and not the entire list of 1198 haplotypes? The reason is simple: Fig. 2 shows that the tree is symmetrical one

and does not contain any distinct branches, identification of which is the main goal of a tree composing. A tree composition from more than a thousand of 67 marker haplotypes takes two to three days of continuous computer calculations, and it is not justified for a smooth and a symmetrical tree, which the most likely has one common ancestor. Obviously, the tree contains many local sub-branches, however, it is both natural and unavoidable in any large haplotype tree, and does not affect calculations of a TSCA (time span to a [most recent] common ancestor [of the population]).

Numerous studies of various R1a1 haplotype datasets (Klyosov, 2009b, 2010, 2011) indicate that a common ancestor of European R1a1 haplotypes lived about 4600 ybp (years before present), that is 184 "conditional" generations ago (25 years per generation). The logarithmic method applied to this dataset immediately shows that despite the "smoothness" of the tree, it contains two branches which distort the first order kinetics of the pattern of mutations in the dataset. These two branches belong to "the Tenths" (DYS388=10) [unlike the base DYS388=12] and the "Young Scandinavian branch" (with YCAII = 19, 21) [unlike the base YCAII = 19, 23]. 101 "the Tenths" would immediately add 202 "extra" mutations to the total number of mutations in the dataset, That is how excessive amounts of mutations enter calculations in the literature, if the branches are not resolved before calculations, which in turn elevates the apparent mutation rates, making them inaccurate or flat wrong. Regarding the Scandinavs, YCAII does not belong to the 22 marker panel, however, affects the TSCA for 67 marker haplotypes.

After removal of those two branches from the dataset, 911 haplotypes remain in the dataset. 25 of those are the base haplotypes in the 12 marker format, and only 1 base haplotype remains in the 25 marker format. It gives

[ln (911/25)]/0.022 = 163 →195 generations (the sign → indicates a correction for back mutations), that is approximately 4875 years to a common ancestor, with the base haplotype

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

and

[ln (911/1)]/0.046 = 148 →174 generations, that is approximately 4350 years to a common ancestor, with the 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

Clearly, one base haplotype is not statistically significant, and results in a margin of error of around 100% (see above). The example is given here only to show that the data are not contradictory.

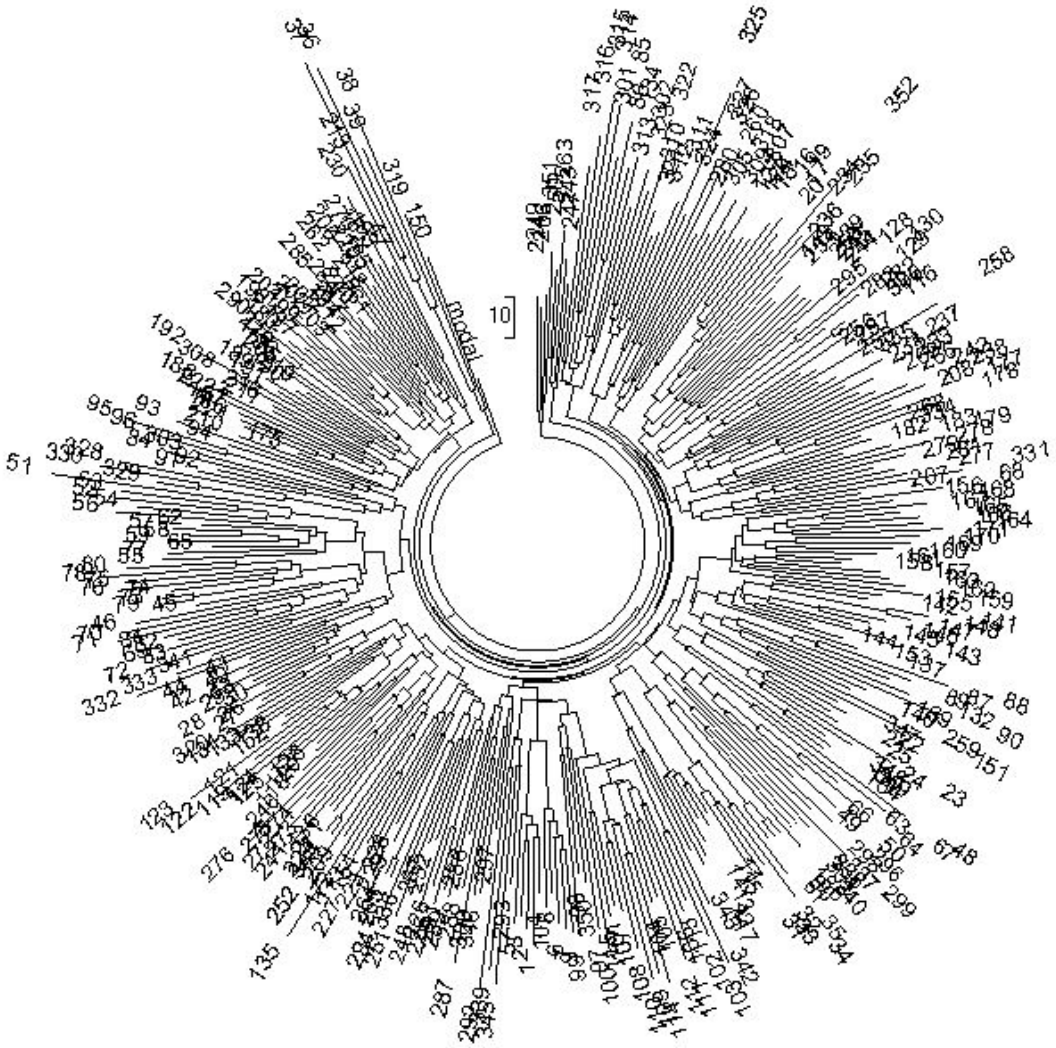


Fig. 1 A tree of 352 of 67 marker haplotypes of haplogroup R1a1 (from YSearch database)

A series of 2299 of 67 marker haplotypes of haplogroup R1b1a2-P312 contained 7 base haplotypes in the 25 marker format, which gave [ln(2299/7)]/0.046 = 126 →145 generations, that is approximately 3625 years to a common ancestor. This is a quite acceptable value for the subclade P312, for which it typically varies from 3600 to 4100 ybp.

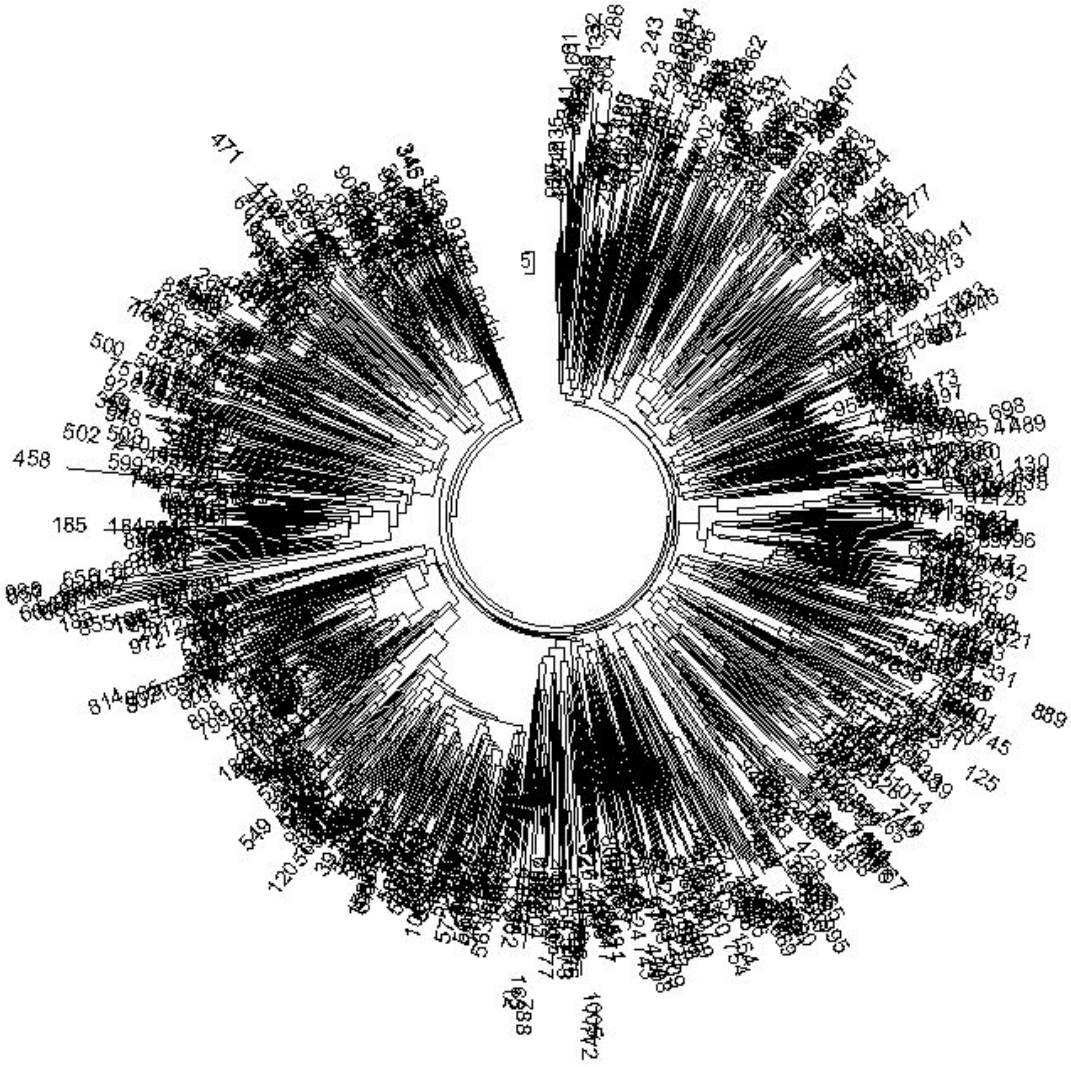


Fig. 2 A tree of 1024 of 67 marker haplotypes of haplogroup R1b1a2-P312-L21 (composed from data provided by Michael Walsh, Administrator of R1b1a2-L21 Project)

Fig. 2 shows a tree of 1024 of 67 marker haplotypes of the downstream R1b1a2-L21 subclade with respect to P312. A relatively young branch in the lower left side could distort calculations of the tree, and this should be taken into account. The whole tree has the following base haplotype:

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

and the young branch on the lower left-hand side has the base haplotype

13 24 14 **10** 11 14 12 12 12 13 13 **30** – **18** 9 10 11 11 25 15 19 **30** 15 15 17 17 –
11 **12** 19 **24** 16 15 18 17 **37** 38 12 12 – **12** 9 15 16 8 10 10 8 10 10 12 **22** 23 16 10 12 12
15 8 **11** 22 20 13 12 11 13 11 11 12 12

108 haplotypes of the young branch have 798 mutations, which gives 798/108/0.12 = 62 →66 generations, that is 1650±175 years to a common ancestor, who lived in the middle of the 1st millennium AD.

It is of interest that such a young branch was initiated by a mutation in a very "slow marker" DYS531, 11→12, which belongs to our new 22 marker panel. Our data has shown that its mutation rate constant equals to 0.00050±0.00020, which means that one mutation in that marker happens once in about 2,000 generations, that is once in 50 thousand years. The fact that this mutation occurred in a founder of the young branch only 1650 years back, reminds that there is no connection whatsoever, between an average frequency of a mutation in any given haplotype, on the one hand, and a TSCA value to a common ancestor the (base) haplotype of whom bears that rare mutation. In other words, any mutation can occur in any haplotype at any time, just some mutations are more rare than others.

All 1024 haplotypes of the R1b1a2-L21 subclade have 16,056 mutations, which gives 16056/1024/0.12 = 131→150 generations, that is 3750±380 years to a common ancestor. The same dataset contains 4 base 25 marker haplotypes, which gives [ln(1024/4)]/0.046 = 121 → 138 generations, that is approximately 3450 years to a common ancestor. These values are the same within margin of error. It was shown above that the series of 2299 haplotypes of the R1b1a2-P312 subclade contained 7 base 25 marker haplotypes, which gives [ln(2299/7)]/0.046 = 126 →145 generations to a common ancestor, that is approximately 3625 years to a common ancestor. They are obviously the same dates, within margins of error.

Nevertheless, in order to double-check the dating, a more careful verification has been performed. All 108 haplotypes of the young branch were subtracted from the list of 1024 P312 haplotypes, and mutations have been recalculated for the residual 916 haplotypes. They contained 14,278 mutations, which gave 14278/916/0.12 = 130 → 149 generations, that is 3725 years to the common ancestor, compared to 3750 years for 1024 haplotypes. This is practically the same figure. It shows that the decrease in the number of haplotypes in this particular case was even (slightly) more important than the decrease in the number of mutations, and after the removal of the young branch the TSCA value slightly increased. However, the number of mutations in the "slow" marker DYS531 in the residual dataset decreased from 161 to 51. It would have been a dramatic effect if only the slow, 22 marker panel was considered.

Table 2.
Mutation rates in an approximate order from the most "slow" markers to the most "fast" ones. The first two columns of the mutation rate constants show a number of mutation per generation, the last three columns show actual number of mutations per the maker in 1024 haplotypes of R1b1a2-L21 subclade, in 2299 haplotypes of R1b1a2-P312 haplogroup (with downstream subclades), and per 1198 haplotypes of R1a1 haplogroup (with downstream subclades)

Marker	Mutation rate constant per marker per generation		A number of mutations per 1024 markers per 3750 years (R1b1a2-L21), per 2299 markers per 4000 years (R1b1a2-P312), and per 1198 markers per 4600 years (R1a1)[a], and the mutation rate per marker per generation		
	Chandler, 2006; Little, 2008[b]: FTDNA, 2008[c]	Ballantyne et al, 2010; Burgarella et al, 2010	R1b1a2-L21 (per 153,600 marker-generations)	R1b1a2-P312 (per 367, 840 marker-generations)	R1a1 (per 220, 432 marker-generations)
472	0.00001	<0.0006	0 <0.000007	1/3 0.000011	0/1 0.000005
425	0.00016 0.00018	0.00113	8/0[d] 0.000052	6/14[e] 0.000056	6/6 0.000054
436	0.00014 0.00018	<0.0006	5/4 (3/4) 0.00059	9/13 0.000060	8/4 0.000054
426	0.00009	<0.0006	20/7 0.00018	11/28[a] 0.00011	6/11 0.00008
490	0.00049 0.00019	<0.0006	9/2 0.00007	3/32[f] 0.00010	16/13 0.00013
454	0.00016	<0.0007	12/2 0.00009	4/42 0.00013	4/29 0.00015
455	0.00016	<0.0006	12/10 0.00014	28/30 0.00016	23/20 0.00019
578	0.00029 0.00008	0.00059	18(17)/2 0.00013	23/35 0.00016	3/35 0.00017
641	0.00017 0.00018	<0.0006	21(18)/8 0.00017	9/53 0.00017	2/22 0.00011
590	0.00017 0.00054	<0.0006	38/1 0.00025	9/44 0.00014	1/13 0.000064
594	0.00030 0.00029	0.00061	27(26)/4 0.00020	6/51 0.00015	11/8[g] 0.00012
388	0.00022	<0.0006 0.00042	23/6 0.00019	13/69 0.00022	4/27[h] 0.00015

492	0.00018 0.00042	<0.0006	17(16)/20 0.00023	25/34[i] 0.00016	5/10[j] 0.000071
395S1b	0.00029 0.00031	n/a	15/8 0.00015	27/36 0.00017	2/24 0.00012
450	0.00061 0.00020	0.00063	33/12 0.00029	18/44 0.00017	0/12 0.000054
617	0.00124 0.00042	<0.0006	170(169)/6 0.00110	33/27 0.00016	8/42 0.00023
531	0.00104 0.00037	0.00059	153/8 (44/7) 0.00037	16/91 0.00029	6/163 0.00077
640	0.00047 0.00034	<0.0006 0.00496	77/1 0.00051	6/133 0.00038	2/43 0.00020
392	0.00052	0.00058 0.00043	87/20 0.00070	43/31[k] 0.00025	4/35 0.00018
568	0.00068 0.00053	0.00065	20(18)/17 0.00023	47/38 0.00023	18/22 0.00018
395S1a	0.00053 0.00031	n/a	52/11 0.00041	24/98 0.00033	81/1 0.00037
438	0.00055	0.00057 0.00043	38/26 0.00042	45/83 0.00035	23/40 0.00029

[a] Format such as 11/28, and then 0.00011 means that this marker (DYS426 in this particular case in haplogroup R1b1a2-P312) all 2299 haplotypes have 11 mutations "up" (from the base allele 12 to 13) and 28 mutations "down" (to allele 11), that is 39 mutations total. These 39 mutation per 2299 markers per 160 "conditional" generations (25 years per generation) give 39/367840 = 0.00011 mutations per marker per generation.

[b]http://freepages.genealogy.rootsweb.ancestry.com/~geneticgenealogy/MR.htm

[c] http://www.kin.marshdna.com/DNAinfo.htm

[d] Without 48 null mutations in DYS425

[e] Without 69 null mutations in DYS425 in 2299 markers-haplotypes of haplogroup R1b1a2-P312.

[f] Without 77 mutations in DYS490=10 in subclade R1b1a2-L176.2/SRY2627, which forms a separate branch.

[g] Without 325 mutations in DYS594=11 in "Scandinavian" branches of haplogroup R1a1.

[h] Without 101 «the Tenths" (DYS388=10), which form a separate branch in haplogroup R1a1. They would have added 202 mutations.

[i] Without 53 mutations in DYS492=14, which form a separate branch in haplogroup R1b1a2.

[j] Without 57 mutations in DYS492=14 of North-Western branch of haplogroup R1a1.

[k] Without 441 mutations in DYS392=14, which is the base for subclades M222, making 25% in Irish haplotypes of R1b1a2 haplogroup.

Let us consider one more example of how a number of mutations in markers is often overcounted. Fig. 3 shows a haplotype tree of haplogroup I from British Isles. Obviously, if the entire list of 194 haplotypes is considered as one population in terms of DNA genealogy, its "the most recent common ancestor" would be a phantom one, along with its phantom "base haplotype", and phantom "TMRCA". Furthermore, a number of mutations in the list will be so high, that the "mutation rate" derived for each marker will be VERY overstated. This, again, shows why so many "mutation rates" in the literature are too high when compared with those calculated accurately, using separated lineages.

However, a haplotype tree can be sometimes misleading. The tree in Fig. 3 might create (a false) assurance that the tree is well-resolved and can be employed for calculating mutation rate constants for the slow, 22 marker panels, particularly using the right-hand side, much older half of the tree. However this will be a mistake, and the mutation rate constants will be significantly overestimated, as many in the literature.

The thing is that the right-hand side includes haplotypes from different haplogroups, as shown in Fig. 3. Furthermore, those haplogroups are further subdivided into subclades and branches.

Let us take the I2b branch (in fact, a combination of branches) in the lower right-hand part of the tree. It contains 66 haplotypes, and their TSCA equals 220 "conditional" generations (25 years each), that is 5500 ybp, if to employ the 67 marker haplotypes.

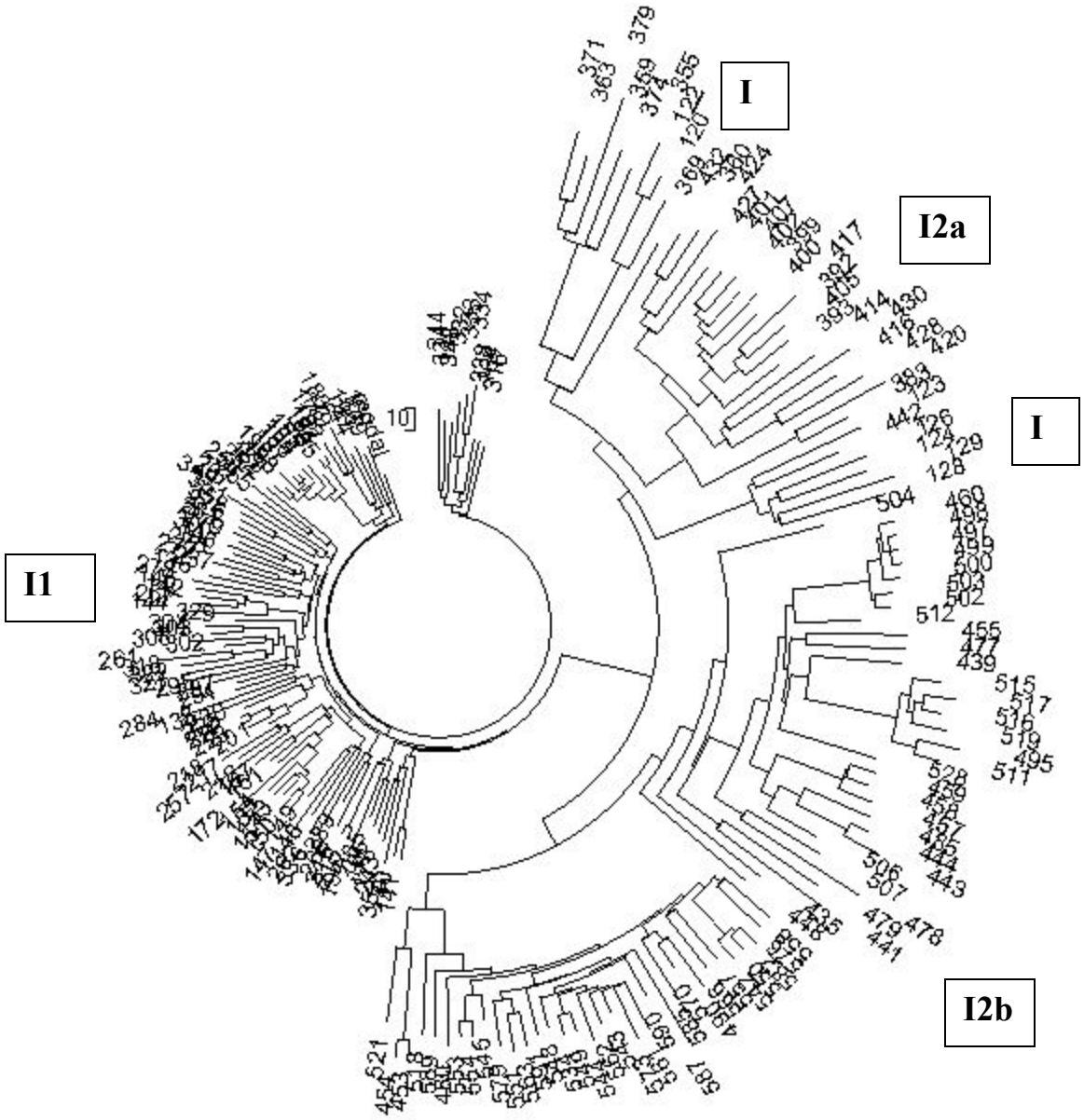


Fig. 3. A 67 marker haplotype tree for haplogroup I from British Isles, containing 194 haplotypes, listed in http://www.familytreedna.com/public/IrelandHeritage/default.aspx

Among 22 of the "slowest" markers of the I2b branch the following nine showed no mutations at all: DYS 472, 436, 426, 590, 490, 641, 578, 392, 438. Another seven markers showed just a few mutations (the notation is as "up/down" mutations from the base allele):

DYS 450 1/2

DYS 594	2/1
DYS395S1b	0/2
DYS454	3/0
DYS455	1/1
DYS492	2/8
DYS617	6/0

total 29 mutations in all of them. However, the remaining six markers shot up with their mutations. One of the slowest markers, DYS425, showed 33 null mutations, which should not be counted in this particular case. They are not mutations per se, they are just do not exist. They cannot be determined. DYS388 gave 12 mutations in only one direction. Clearly, it is a block of mutations, a feature of a separate branch. DYS531 gave 10/2, DYS395S1a – 17/1, DYS640 – 33/0, DYS568 – 30/4. Summarily it is 138 mutations, without counting null mutations. For 66 of 22 marker haplotypes over 220 generations, the average mutation rate appear to be 138/66/220/22 = 0.00950 mutations per haplotype, and 0.000430 mutations per marker per generation. It is significantly higher compared to average mutation rates considered in this study. The reason is the same as explained earlier, namely, not separation of haplotypes into branches. Below are examples of how overcounted were mutations in this case when branches were not resolved:

DYS388 – 0.0083, when Table 2 shows 0.00022, 0.00022 и 0.00015, difference is 40-55 times.

DYS531 - 0.0083, Table 2 shows 0.00037, 0.00029, 0.00077, difference is 11-20-30 times.

DYS395S1a – 0.00124, Table 2 shows 0.00031 and 0.00033, 0.00037, 0.00053, difference is 2-4 times.

DYS640 – 0.00227, Table 2 shows 0.00051, 0.00020, 0.00038, difference is 5-10 times.

DYS568 – 0.00234, Table 2 shows 0.00023, 0.00023, 0.00053, 0.00065, 0.00068, 0.00018, difference is 4-13 times.

This is the "mechanism" of how mutation rate constants are overstated in many publications in the literature.

A more detailed consideration showed that practically all those "excessive mutations" went into a separate branch on he lower right-hand side of the tree in Fig. 3. This is subclades I2b1. As a result, it turned out that besides the said nine markers which did no have mutations, another three did no have them as well:

DYS 425, DYD388, and DYS568. Another seven markers from the 22 marker panel had 29 mutations described above. The residual three markers showed just a few mutations: DYS640 2/0, DYS531 0/2, DYS395S1a 0/1, making 34 mutations total. This, however, decreases a number of haplotypes and mutations in the branch, and the system becomes less statistically robust, having more than a half haplotypes not having mutations at all.

This above example was given not to make calculations, but to show how "excessive mutations" get into "mutation rate constants" in the literature, if not to resolve haplotype datasets into distinct branches.

The optimized summary Table 3 is shown below. The average mutation rate for all 22 marker haplotype is 0.00600 mutations per "conditional" generation of 25 years, the average mutation rate for a marker in this panel is 0.00027 mutations pr generation. This value is almost an order of value lower compared to the mutation rate constants for 12-, 17-, 25-, 37- и 67-marker haplotype panels (0.00183, 0.00200, 0.00183, 0.00243 и 0.00179 mutations per marker per generation, respectively).

Table 3.
Mutation rate constants (MRC) for the slowest 22 markers from the 67 marker panel of haplotypes in an order from the most "slow" markers to the most "fast" ones.

Marker	MRC per marker per "conditional" generation of 25 years
472	0.00001
425	0.00005
436	0.00006
426	0.00009
490	0.00013
454	0.00016
455	0.00016
578	0.00017
641	0.00017
590	0.00017
594	0.00020
388	0.00022
492	0.00023
395S1b	0.00025
450	0.00029
617	0.00050
531	0.00050

640	0.00051
392	0.00052
568	0.00053
395S1a	0.00053
438	0.00055
$\sum$	0.00600/haplotype/generation (25 years)
Average	0.00027/marker/generation

This 22 marker panel is more accurate an reliable when time spans to common ancestors reach 10 thousand years and up to 50 and 100 thousand year and higher. They are more reliable also in that corrections for back mutations are much less for "slower" markers and haplotypes in terms of rates of their mutations. For example, for a "canonic" mutation rate constant of 0.002 mutation per marker per generation, an average number of mutations of 1.0 per marker correspond to 1/0.002 = 500 generations without a correction for back mutations, but as many as 925 generations with the correction, that is 23,125 years before present (the time span to a common ancestor of that population). Correction factor in this particular case is 1.85, that is +85%. For the average mutation rate constant of 0.00027 per marker per generation the distant of 23,125 ybp will correspond to only 0.22 mutation per marker, that is a correction actor will be only 1.12, +12%. This follows from the equation (Klyosov, 2009a)

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

where:
λ is a corrected average number of mutations per marker,
λ_{obs} is the observed number of mutations per maker.

Straightforward calculations show that at λ_{obs} = 0.22, the corrected value is only 12% higher. At 0.300 mutations per marker the correction will be 17%, however, it corresponds already 33 thousand years to a common ancestor of the population. At those historical depths corrections for back mutations can be neglected for some approximate calculations.

Practical applications of the 22 marker panel of haplotypes and its mutation rate constants are described in the subsequent study in this issue, in which TSCA values were calculated to as ancient times as 130 thousand (and higher) years to common ancestors of present day humans.

References

Ballantyne, K.N., Goedbloed, M., Fang, R., Schaap, O., Lao, O., Wollstein, A., Choi, Y., van Duijn, K., Vermeulen, M., Brauer, S., Decorte, R., Poetsch, M., von Wurmb-Schwark, N., de Knijff, P., Labuda, D., Vezina, H., Knoblauch, H., Lessig, R., Roewer, L., Ploski, R., Dobosz, T., Henke, L., Henke, J., Furtado, M.R., Kayser, M.(2010) Mutability of Y-chromosomal microsatellites: rates, characteristic, molecular bases, and forensic implications. Am. J. Human Genet. 7, 341-353.

Burgarella, C, Navascues, M. (2010) Mutation rate estimates for 110 Y-chromosome STRs combining population and father-son pair data. Eur. J. Hum. Genet., advance online publication, 8 September 2010

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

Klyosov, A.A. (2009a). DNA Genealogy, mutation rates, and some historical evidences written in Y-chromosome. I. Basic principles and the method. J Genetic Genealogy 5, 186-216.

Klyosov, A.A. (2009b) DNA Genealogy, mutation rates, and some historical evidences written in Y-chromosome. II. Walking the map. J. Genetic Genealogy. 5, 217 – 256.

Клёсов, А.А. (2010) Haplogroup R1a1 and its subclades in Asia. Proc. Russian Academy of DNA Genealogy (ISSN 1942-7484), 3, № 11, 1866-1896.

Klyosov, A.A. (2011a) Biological chemistry as a foundation of DNA genealogy: the emergence of "Molecular History". Biochemistry (Moscow) (ISSN 0006-2979) 76, No. 5, 517-533.

Klyosov, A.A. (2011b) DNA genealogy of the major haplogroups of Y chromosome. Proc. Russian Academy of DNA Genealogy (ISSN 1942-7484), 4, № 5, 988-1014.

DNA genealogy of the major haplogroups of Y chromosome. (Part 1)

Anatole A. Klyosov
http://aklyosov.home.comcast.net

(This is a shorter version of the article published in Russian in Proceedings of the Russian Academy of DNA Genealogy, May 2011, vol. 4, No. 5, pp. 988-1014)

Resume

Using the slowest 22 marker panel, described in the preceding article, the chronology of appearance of haplogroups A, B, C, I, I2, R1 has been calculated from their base haplotypes. The principal premise of this approach is that haplotype alleles evolve in a continuous process of random mutations of their markers in Y chromosome since the time immemorial. Certain SNPs, chosen by researchers for subdivision of Y chromosomes into 20 (currently) categories called haplogroups, have "punctured" Y chromosomes in another random process, absolutely irrelevant to the (random) mutational evolution of haplotypes. As a result, we now have on the Earth an enormous branched pyramid of haplotypes, each one marked with a certain (Latin) letter identifying a haplogroup, which (haplogroups) themselves consist of many lineages arranged into various subgroups, called subclades, then into branches, lineages, etc. Each haplogroup, subclade, branch, lineage, etc. can be characterized with a base haplotype, which was carried by a founder of the haplogroup, subclades, branch, lineage, etc. a certain time ago. Most of them were terminated long ago as a result of natural cataclysms, wars, famine, daughters out, etc., and those which survived to the present time continue this endless continuation of their haplotypes, all connected with haplotypes of our ancestors thousands, tens and hundreds of thousands, and millions years ago. According to current knowledge, this connection had passed through a population bottleneck between 200 and 100 thousand years ago, and haplotypes of all people on Earth originated from one common ancestor who was coined with a name "chromosomal Adam".

The oldest haplogroup among males on our planet is believed to be haplogroup A which – again – believed to arose in Africa, as well as the second oldest haplogroup B. All other haplogroups (C-T) are considered to be descending from haplogroup B, and believed to have a non-African origin (meaning they have appeared outside of Africa, but – technically – from the African haplogroup B).

This study which employed the "ultra-slow" 22 marker panel, has identified the most stable (in terms of mutations) 22 marker base haplotypes of all principal haplogroups, and "connected" them to each other on an absolute time scale. This led to an identification of a common ancestor of the "alpha-haplogroup" who lived 136,000 years before present (ybp), and from whom haplogroup A (but not haplogroup B) descended approximately 85,000 ybp.

Overall, haplogroups, according to this study, arose:

Alpha-haplogroup, 136,000 years bp ("Chromosomal Adam")
Haplogroup A, 85 thousand years bp (descended from "alpha")
Beta-haplogroup, 64 thousand years bp
Haplogroup B, 46 thousand years bp (descended from "beta")
Haplogroup C, 36 thousand years bp (descended from "beta")
Haplogroup I – 46 thousand years bp
Haplogroup I2 – 34 thousand years bp
Haplogroup R1 – 23 thousand years bp

Other haplogroups are analyzed in the Part 2 of this study, which will be published in the next issue of the Proceedings.

The data showed that haplogroup B did not descend from haplogroup A, but from the "beta"-haplogroup. All non-African haplogroups descended also from the "beta-haplogroup".

The base haplotypes of said haplogroups were as follows (that of R1a-M17 is for the Russian Plain [RP], of R1b1a2-M269 is for Europe [Eur]):

12 11 11 – 9 11 – 10 – 10 8 14 15 7 10 8 12 13 11 16 8 13 9 11 12	(A)
11 12 11 – 11 11 – 10 – 11 8 16 16 8 10 8 12 10 11 15 8 12 11 12 11	(B)
11 13 11 – 11 11 – 10 – 10 8 16 16 8 10 8 12 12 12 13 8 12 11 11 11	(C)
11 14 11 – 8 11 – 10 – 11 8 15 15 8 10 8 12 10 12 12 8 13 11 11 12	(I1d1)
11 13 11 – 11 11 – 10 – 11 8 15 16 8 10 8 12 11 12 12 8 13 10 12 12	(I2)
12 12 11 – 11 11 – 11 – 11 8 17 17 8 10 8 12 10 12 12 8 12 11 11 12	(R1a, RP)
12 12 13 – 11 11 – 12 – 11 9 15 16 8 10 8 12 10 12 12 8 12 11 11 12	(R1b1a2, Eur)

Common ancestors of the haplogroups lived as the following times (margins of error are estimated around 10-15%):

A and B,	130,000 ybp.
A and C,	122,000 ybp.
A and R1a1,	130,000 ybp.
A and R1b1b2,	136,000 ybp.
A and I2,	127,000 ybp.

All of them seemingly descended from "alpha-haplogroup" with a common ancestor who lived 136,000±9,000 ybp from these andother data; realistically the margin of error should be larger.

B and C,	62,000 ybp.
B and I2,	67,000 ybp.
B and R1a1,	56,000 ybp.
B and R1b1a2,	64,000 ybp.

All of them seemingly descended from "beta-haplogroup" with a common ancestor who lived 64,000±6,000 ybp from the above calculations; realistically the margin of error should be larger.

Haplogroup C did not have a common (direct) ancestor with haplogroups I2, R1a1, andR1b1a2.

These data show that common ancestor of alpha-haplogroup, from which haplogroup A descended, lived 136,000 years before present, that is a huge timespan before haplogroup A arose (or it had passed a population borrleneck). Haplogroup A descended from the "alpha" common ancestor, and haplogroup B descended from the "beta" common ancestor, about 40,000 years apart. There is a huge gap in 18 mutations in the 22 marker panel between base haplotypes of haplogroups A and B, which places the common ancestors of haplogroup A and B by more than 120,000 years apart (the "lateral" mutational difference).

Haplogroups B, C, I2, and R1 all descended from a common ancestor who lived 64,000±6,000 years before present. It was apparently a haplogroup which initiated the non-African population of our planet.

If this (tentative) conclusion is correct, then the non-African alpha-haplogroup (136,000 ybp) was an ancestral to the (pesently) African haplogroup A (85,000 ybp), which currently live in Africa in their majority. This alpha-haplogroups was also ancestral to the beta-haplogroup, which was not identified as yet in the phylogeny of haplogroups, and which arose 64,000±6,000 ybp, apparently outside of Africa. It was ancestral to B and (in parallel) all C-T haplogroups and their downstream subclades, including haplogroups C (36,000 ybp), I (46,000 ybp), I2 (34,000 ybp), R1 (23,000 ybp), R1a1 (21,000 ybp), R1b1a2 (16,000 ybp). If so, a major discrepancy between "population genetics" and anthropology/archaeology might be resolved. The discrepancy means that the first (population geneticists) insist on the "out-of-Africa" theory, despite the haplogroup tree is not rooted, while many anthropologists insist that anatomically modern man had appeared in Eurasia and certainly not in Africa, and specialists in genome studies take an uncertain central position, albeit prefer to interpret their finding (often still rather vague) leaning towards the "out-of-Africa" concept.

Introduction

The preceding paper (Klyosov, 2011) explains why a "slow-mutating" panel of markers is necessary for studying haplotypes in a time frame below 20-40 thousand years ago, and down to 100+ thousand years before present. The 22 marker panel was arranged (Table 1), with the average mutation rate constant of 0.0060 mutation/hapl/gen, and 0.00027 mutation/marker/generation, which is almost ten times slower compared to those for 12, 17, 25, 37 and 67 marker haplotypes. In fact, individual mutation rate constants, listed in Table 1, are practically useless for TSCA (time span to a [most recent] common ancestor) calculations, since for that we need thousands of haplotypes (that is, thousands of individual markers in them) to achieve satisfactory statistics. The individual mutation rate constants are useful in order to obtain an average mutation rate constant for the entire 22 marker haplotype per generation, and – if needed – a average mutation rate constant per marker per generation. It does not have much sense to follow individual mutations in individual markers, since statistics is not there.

Table 1.
Mutation rate constants (MRC) for the slowest 22 markers from the 67 marker panel of haplotypes in an order from the most "slow" markers to the most "fast" ones.

Marker	MRC per marker per "conditional" generation of 25 years
472	0.00001
425	0.00005
436	0.00006
426	0.00009
490	0.00013
454	0.00016
455	0.00016
578	0.00017
641	0.00017
590	0.00017
594	0.00020
388	0.00022
492	0.00023
395S1b	0.00025
450	0.00029
617	0.00050

531	0.00050
640	0.00051
392	0.00052
568	0.00053
395S1a	0.00053
438	0.00055
∑	0.00600/haplotype/generation (25 years)
Average	0.00027/marker/generation

The last two lines in Table 1 have been used in this study for analysis of chronology of haplotypes and haplogroups of Y chromosome.

Haplogroup A

The literature offers a number of estimates of an "age" of haplogroup A, which is believed to arose in Africa, though the haplogroup tree is not rooted. Some researchers believe that the haplogroup is 200-160 thousand year old, based on anthropological data on a timing of an alleged appearance of *Homo sapiens*, and assume that it was the "chromosomal Adam". Of course, such an assumption is built on sand, since we know nothing on possible offspring (or lack of it) of those bones excavated in Africa and elsewhere, survival (or not) of their lineages and their ancestral connection to present-day people. The data itself, such as 200-160 thousand years, mentioned above, is also constantly disputed and challenged.

Other researchers place an origin of haplogroup A closer to 80 thousand years, an postulate that it was time of a possible "out-of-Africa" migration. However, other postulate that it was haplogroup B, not A, which migrated from Africa, while yet other insist that there was no "out-of-Africa" migration at all, and non-African population arose from non-African people, whether it was in the Middle East, Asia, or Europe. Many representatives of the Russian school of anthropology are convinced that anatomically modern people first appeared on the Russian Plain (Eastern European Plain) where the most of relevant artifacts have been found with datings close to 50,000 ybp.

Let us see what haplotypes can tell in that regard.

Figs 1 and 2 show 37 and 67 marker haplotype trees of haplogroup A and some of it subclades. Obviously, the lineages are fairly young, particularly in subclades A1a and A3b. Either they have appeared rather recently, or they recently passed a population bottleneck, or else.

Subclade A1a

The A1a branch in the tree has the following base haplotype in the 67 marker format:

13 23 14 11 16 17 12 10 12 13 11 31 – 16 8 9 7 13 23 14 20 34 11 12 13 16 – 11 11 17 19 15 14 18 18 31 32 11 8 – 10 8 15 17 6 11 10 9 9 8 12 21 21 20 13 11 16 15 8 12 21 23 12 13 11 14 10 11 12 11

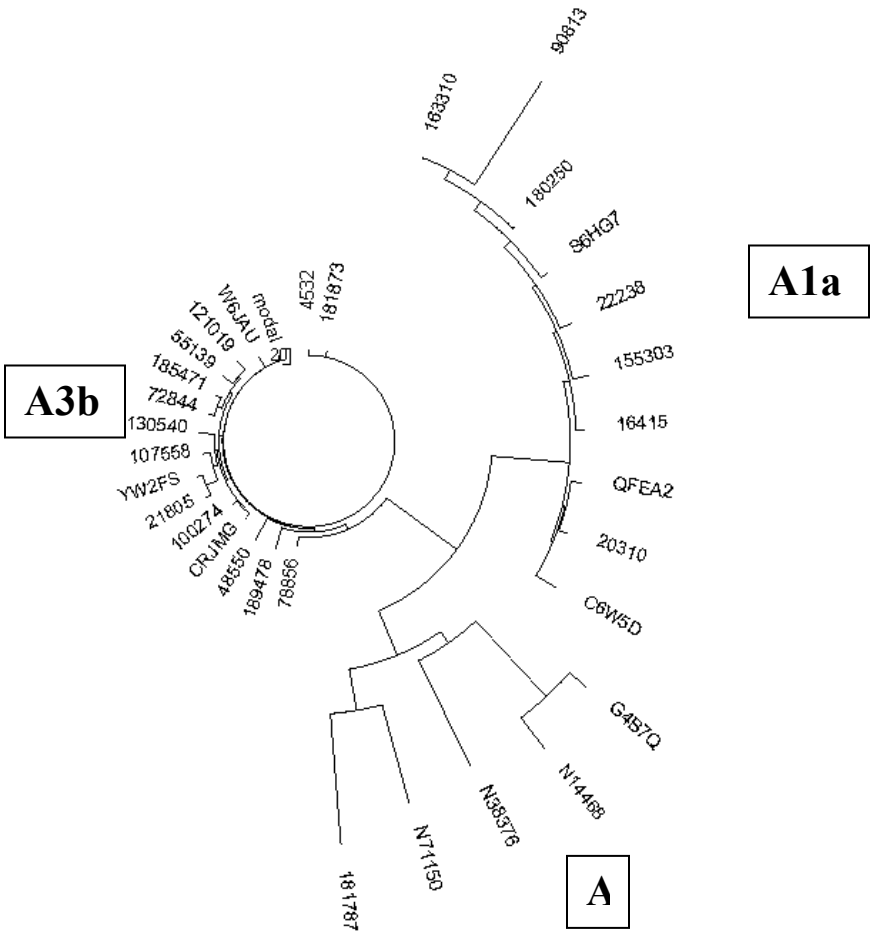


Fig. 1. A tree of 31 of 37 marker haplotypes of haplogroup A with some subclades. Haplotypes were taken from YSearch и FTDNA's «African Project» (http://www.familytreedna.com/publicwebsite.aspx?vgroup=African.DNAProject&section=yresults).

Nine 37 marker haplotypes of the branch (the tenths haplotype, No. 90813, is considered below) have only 18 mutations, which gives 18/9/0.09 = 22 generations, that is 550±140 ybp. It is possible that the whole branch consists of descendants of an African slave who was taken from Africa in the 15th century, and now his descendants could have afforded their DNA analysis unlike their more lucky tribe members who continue to live in Africa.

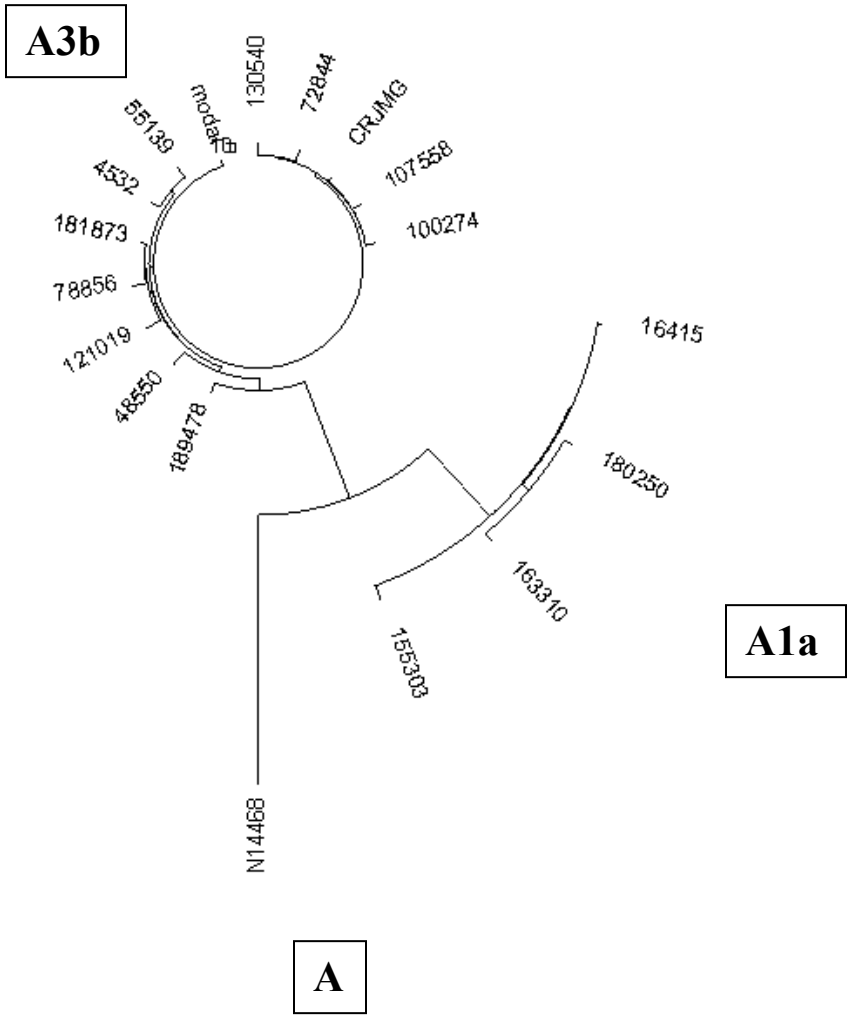


Fig. 2. A tree of 17 of 67 marker haplotypes of haplogroup A with some subclades. Haplotypes were taken from YSearch and FTDNA's «African Project» (http://www.familytreedna.com/publicwebsite.aspx?vgroup=African.DNAProject&section=yresults).

Haplotype 90813 which is an outlier and therefore was not included into calculations above, clearly belongs to quite a different lineage. It differs from the above base haplotype by 29 mutations (!) on 37 markers (mutations are marked):

14 22 14 11 16 **20** 12 10 **11 12** 11 30 – **15** 8 9 7 13 23 14 20 34 **13 14 14 17** – 11 **10** 17 19 15 **16 16** 18 **34 37 12** 8

Those 29 mutations place the above two haplotypes by 11700 years apart, and put their common ancestor at about 6125 years before present. In other words, in

order to accumulate 29 mutations in one 37 marker haplotype compared to another, the base haplotype of THEIR ancestor needs 6125-550 = 5575 years to the base haplotype of that (presumably) slave in the 15th century, and 6125 years to the present-day haplotype 90813, making those 11700 total. That is how a TSCA value to a common ancestor of two haplotypes is calculated.

Let us move to 22 marker haplotypes. The base A1a haplotype described above in the 67 marker format, in the 22 marker format looks as follows:

12 10 11 – 7 13 - 8 – 10 8 15 17 6 10 9 12 13 11 16 8 13 11 11 12 (**A1a, 550±140 ybp**)

The single haplotype 90813 exists only in the 37 marker format, therefore, in the 22 marker format it has only the first six alleles:

12 10 11 – 7 13 – 8

As one can see, the above two haplotypes cannot be resolved in the "slow-mutating" format. This was the idea of the example. In the first six markers the resulting mutation rate was approximately 0.00170 per haplotype per generation (see Table 1), therefore, one mutation in those six markers would occur – on average - in 1/0.0017 = 588→640 generations, that is once in 16,000 years. It is of no surprise that at a difference in 11,700 years between the two base haplotypes there is not a single mutation between them.

Let us compare data in 7 and 67 marker haplotypes. From the nine of 37 marker haplotypes it was determined, that a common ancestor of the branch lived 550±140 ybp (see above). From the four of 67 marker haplotypes o the same subclades A1a, which contain 10 mutations, we obtain 10/4/0.12 = 21 generations, that is 525±170 ybp. The fit is almost ideal. This means that the branch is the proper one in terms of DNA genealogy, and indeed has one common ancestor.

Subclade A3b

16 haplotypes of the subclades have the following 67 marker base haplotype (mutations compared with that of subclades A1b are marked):

13 **21 15 9 11 12** 12 **11** 12 13 11 **29** – 16 **9** 9 **9 11** 23 **15 21** 34 **10 11 15** 16 – **10** 11 **21 21 14** 14 18 18 **29 30** 11 **10** – 10 **9 12 12 7 12** 10 **8 10** 8 **0 23 27 17** 13 11 16 15 **10** 10 **27** 23 **16 14 9 15 9** 11 **11** 11

In the first 37 markers those 16 haplotypes have 32 mutations, which gives 32/16/0.09 = 22 generations, that is the same 550±110 years from a common ancestor, as in subclades A1a. It might have been the same wave of the imported

slaves. 12 of 67 marker haplotypes from the same series have 41 mutations, which gives 41/12/0.12 = 28 → 29 generations, that is 725±130 ybp. This is within margin of error with the TSCA, calculated for the 37 marker haplotypes.

12 11 11 – 9 11 – 10 – 10 9 12 12 7 10 8 0 13 11 16 10 14 9 11 11 (**A3b, 550±110 ybp**)

Branches of subclades of haplogroup A, and their common ancestor

There are huge amount of mutations between the base haplotypes of A1a and A3b subclades, 88 in their 67 marker haplotypes. This corresponds to 46 thousand years between their common ancestors, and places THEIR common ancestor to 24,000 ybp. Here we have, however, a problem: very recent common ancestors of the both subclades, and a very ancient THEIR common ancestor. In such a situation a use of 67 marker haplotypes becomes rather problematic. Therefore, we will consider their 22 marker haplotypes:

12 10 11 – 7 13 – 8 – 10 8 15 17 6 10 9 12 13 11 16 8 13 11 11 12 (A1a)

12 **11** 11 – **9 11** – **10** – 10 **9 12 12 7** 10 **8 0** 13 11 16 **10 14 9** 11 **11** **(**A3b)

There are as many as 25 mutations between them, and most of them have occurred in the slowest, last part of the haplotypes, sometimes reaching as many as five mutations in the same marker. This signals on their VERY ancient common ancestor. Luckily, we have a third base haplotype, of subclades A, which forms a branch in the lower right-hand part of the haplotype tree in Fig. 1. A calculation with three base haplotypes would increase accuracy of the estimate.

Five individual haplotypes of haplogroup A have the following 37 marker base haplotype:

14 20 15 11 17 19 12 12 12 13 11 30 – 17 8 9 10 11 22 14 19 31 13 14 15 16 – 10 10 19 19 14 13 17 18 35 36 12 10

They all have 136 mutations from the above base haplotype, which gives 136/5/0.09 = 302 → 428 generations, that is approximately 10,700 years to a common ancestor. In the 22 marker format the base haplotype is as follows:

12 12 11 – 10 11 – 10 – 11 8 15 15 8 9 8 0 10 9 14 8 12 8 11 12 (**branch A, 10,700 ybp**)

A consideration of all the three base haplotypes in the 22 marker format results in the following, presumably base (ancestral) haplotype of haplogroup A

12 11 11 – 9 11 – 10 – 10 8 14 15 7 10 8 12 13 11 16 8 13 9 11 12 (**A, ~85,000 ybp**)

All three base haplotypes, of a branch of subclade A, A1a, and A3b, are collectively remote of the last base (ancestral) haplotype by 41 mutations, which gives 41/3/0.006 = 2278 → 3258 generations from a common ancestor (a correction factor here is 1.43), that is 81,450 ybp, plus 4 thousand years of the averaged "age" of all the three subclades/branches. As a result, we obtain that the common ancestor of all the three identified branches/subclades lived approximately 85,000 years before present.

Since the base haplotype for the common ancestor of haplogroup A is a deduced one, and 41 mutations from it could also be subject of distortions, we need to double check it. Therefore, we have applied the permutation method (Klyosov, 2009a) for which there is no need to find a base haplotype and count mutations from it. In this method a squared difference in alleles should be counted for each allele, that is for all 66 alleles in three experimentally identified base haplotypes, shown above. The sum of these squared mutations was equal 344 in this particular case, when each difference was counted twice, from "up" to "down", and again, from "down" to "up". This value should be divided by 2 (since we counted each difference twice), then by square of a number of haplotypes (9), then by a number of markers (22), and obtain 344/2/9/22 = 0.869 mutations per marker, already corrected for back mutations, since the correction is already "built in" into the permutation method (Klyosov, 2009a).

This gives 0.869/0.00027 = 3219 generations, that is 80,500 +4,000, approximately 85 thousand years to a common ancestor of haplogroup A, practically the same "age" which was calculated above using the "linear" method of calculations, with a correction for back mutations.

Haplogroup B

According to current views «B is thought to have arisen approximately 50,000 years ago» (International Society of Genetic Genealogy, Y-DMA Haplogroup Tree 2100, http://www.isogg.org/tree). The words "is though to" are rather peculiar; it would be good to see based on what this thought has appeared. Anyway, if it is a wild guess, it is a good guess, as it is shown below (46,000 ybp, as follows from our data).

Figs 3 and 4 show haplotype trees of haplogroup B in the 37 and 67 marker formats. The left-hand side of the both trees shows "young" lineages, the right-hand side shows "old" lineages, as it can be seen from shapes of the branches.

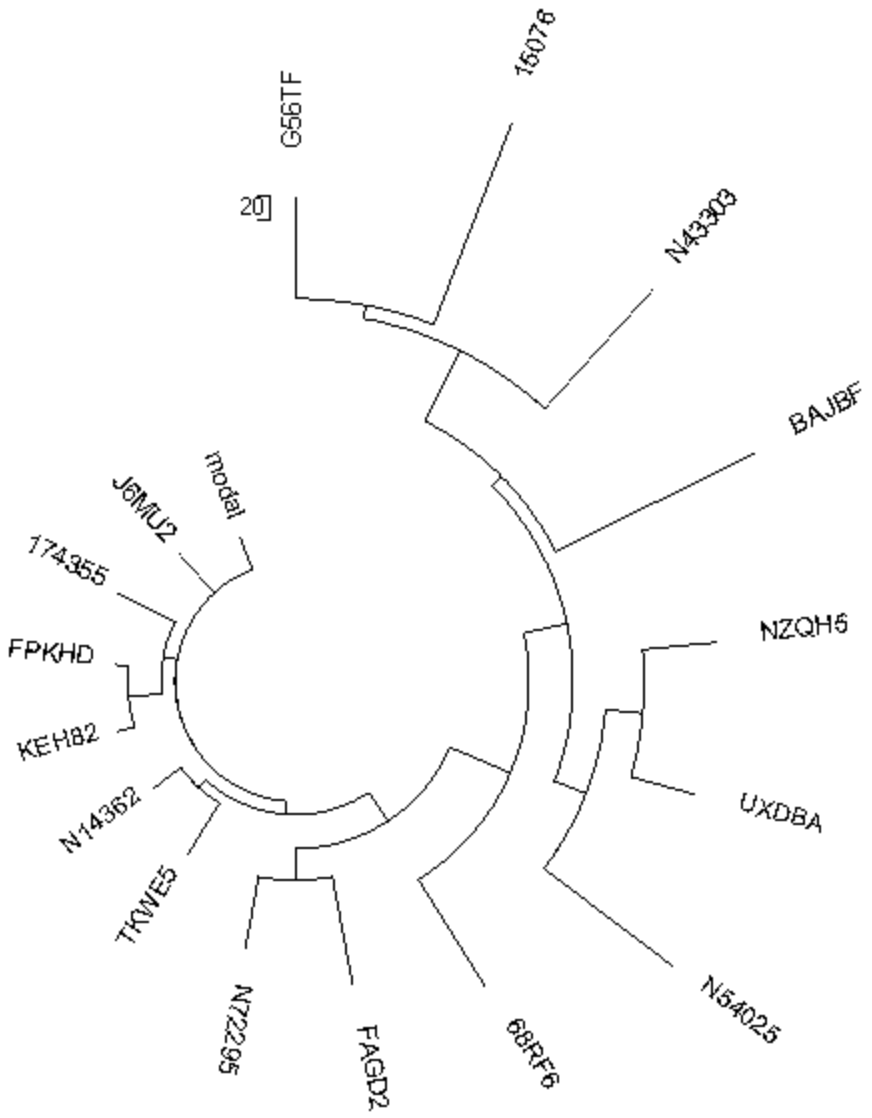


Fig. 3. A tree of 16 of 37 marker haplotypes of haplogroup B. Haplotypes are taken from YSearch and the FTDNA "African Project" (http://www.familytreedna.com/publicwebsite.aspx?vgroup=African.DNAProject&section=yresults)

The first four haplotypes in Fig. 3 from the top in the counter-clock direction have the following base 67 marker haplotype:

13 23 15 11 11 12 11 10 13 13.5 11 32.5 – 16 8 8 10 11 26 14 21 38 16 16 17.5 18 – 12 11 21 22 14 12 18 17 33 37 8 10 – 11 8 15 16 8 11 10 8 10 10 12 20 20 17 10 11 14 14 8 12 24 21 15 12 11 14 10 12 11 11

Their first 37 markers have 304 quadratic permutation mutations, which gives 304/2/16/37/0.00243 = 106 generations, that is 2650±570 ybp. 16 here is the

The permutation method here is the preferred one, since four haplotypes are not statistically robust number for counting mutations from a deduced base haplotype.

The following two haplotypes (N14362 and TKWE5) differ by only 7 mutations in the first 37 markers from the deduced base haplotype:

13 24.5 15 10 11 11 11 10 12.5 14 11 32.5 – 16 8 8 10 11 26 15 21 34.5 16 16 17 18 – 11 11 21.5 22 14 12 17.5 17 33 36.5 8 10 – 11 8 16 16 8 12 10 8 10 10 12 20 20 20 10 11 14 13 8 12 24 21 15 12 11 14 10 12 10 11

This gives approximately 1025±400 years to their common ancestor.

The following two haplotypes (N72295 and FAGD2) differ by 22 mutations in the first 37 markers from the deduced base haplotype:

13.5 23.5 16 10 11 11 11 10 12 13.5 11 31.5 – 18 8 8 11 11 26 14.5 23 35 15 16 17 17.5 – 11 10.5 20.5 22 13 12.5 18 16.5 33 35.5 8 9 – 11 8 16 16 8 11 10 8 10 10 12 20 21 15 10 11 15 15 8 11 27 22 17 12 11 14 10 12 11 11

This gives approximately 3500±820 ybp.

Calculations are complicated for the ancient right-hand branches, since statistics is not there. However, even the three preceding small branches differ by as many as 44 mutations in their 67 marker haplotypes, which gives approximately 14,200 years to their common ancestor.

Now let us consider eight of 22 marker haplotypes, shown as 67 marker ones in Fig. 4. Two on the left-hand side are practically the same, as well as three haplotypes in the lower part of the tree. They represent one branch. Five 22 marker haplotypes (N72295, N54025, UXDBA, G56NF and KEH82) when considered by the permutation method have 540 quadratic mutations, which gives 540/2/25/22/0.00027 = 1818 generations, that is approximately 46,000 years to a common ancestor of haplogroup B. His 22 marker haplotype was apparently as follows:

11 12 11 – **11** 11 – 10 – **11** 8 **16 16 8** 10 8 12 **10** 11 **15** 8 **12 11 12 11**

It differs from he base 22 marker haplotype of haplogroup A by 18 mutations as marked, which gives 18/0.006 = 3,000 generations, with a correction factor (for back mutations) of 1.633, that is 123,000 years between common ancestors of haplogroup A and B. Because they lived 85 and 46 thousand years before present, THEIR common ancestor lived approximately 136,000 years before present.

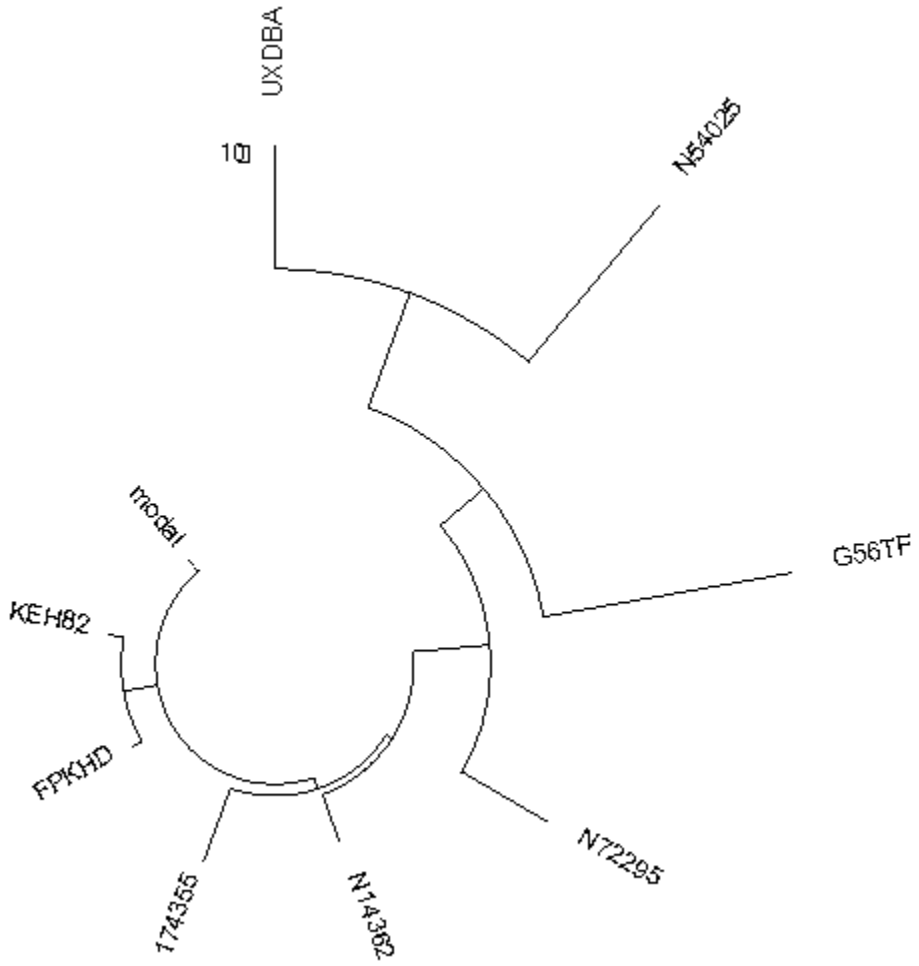


Fig. 4. A tree of eight of 67 marker haplotypes of haplogroup B with subclades. Haplotypes are taken from YSearch and the FTDNA "African Project" (http://www.familytreedna.com/publicwebsite.aspx?vgroup=African.DNAProject&section=yresults).

It indicates that haplogroup B did not descend from haplogroup A. They descended from two different common ancestors who lived 136,000 ybp and 64 ybp, respectively, and not necessarily that they lived in Africa. Since the second one is a haplogroup upstream from the first one, the first haplogroup can be named "alpha haplogroup" due to lack of a more proper name.

Haplogroup C

The International Society of Genetic Genealogy, Y-DNA Haplogroup Tree 2100, http://www.isogg.org/tree) defines a time of appearance of haplogroup C as it "appears to have arisen shortly after modern humans left Africa and is estimated to be approximately 50,000 years old". Again, it is not clear whether this is a wild (or a sort of educated) guess or it is based on certain data.

Haplogroup C is named differently by different research groups. Some call it "CT", some "CF". It is just a matter of naming, since each haplogroup includes its downstream subclades by definition, since all of them carry the all the upstream SNPs. In this study we call it haplogroup C, with a full understanding that it can be named otherwise as described above.

Fig. 5 shows a haplotype tree for haplogroup C with some subclades. A common ancestor of two Hawaiian haplotypes lived 1925±550 ybp (see the legend to Fig. 5). This is a common "age" for Polynesian haplotypes (Klyosov, 2009b).

The base haplotypes of the branches are as follows (numbers on the right shows a number of mutations in the respective branch):

11 13 11 – 11 12 – 11 – 10 9 16 16 8 10 8 12 11 12 13 8 12 12 11 10 (0)
11 13 11 – 11 12 – 10 – 10 10 16 16 8 10 8 12 11 12 13 8 13 12 11 10 (5)
11 14 11 – 11 12 – 9 – 10 9 16 16 8 10 8 12 11 12 14 8 14 11 11 10 (2)
11 13 11 – 11 12 – 11 – 10 9 16 16 8 10 8 12 11 12 13 8 12 11 11 10 (0)
11 13 11 – 11 12 – 10 – 10 9 16 16 8 10 8 12 11.5 12 13 8 12 10 11 10 (1)
11 13 11 – 11 11 – 10 – 10 8 16 16 8 10 8 12 12 12 12 8 12 11 11 11 (35)
11 15 12 – 11 11 – 11 – 10 8 16 16 8 8 8 13 12 12 12 8 12 10 11 11 (0)

"Ages" of the respective branches based on numbers of mutations in the 22 marker format are: 0 (haplotypes 1-4), 7320 ybp (haplotypes 7, 13, 14), 4175 ybp (haplotypes 5 and 6), 0 (haplotypes 8 and 9), 2080 ybp (haplotypes 10 and 11), 17,900 ybp (the rest of nine haplotypes on the ancient branch in Fig. 5, except subclades C2a), 0 (haplotypes 23 and 24). The base (ancestral) haplotype of haplogroup C based on the data above is as follows:

11 13 11 – 11 11 – 10 – 10 8 16 16 8 10 8 12 12 12 13 8 12 11 11 11

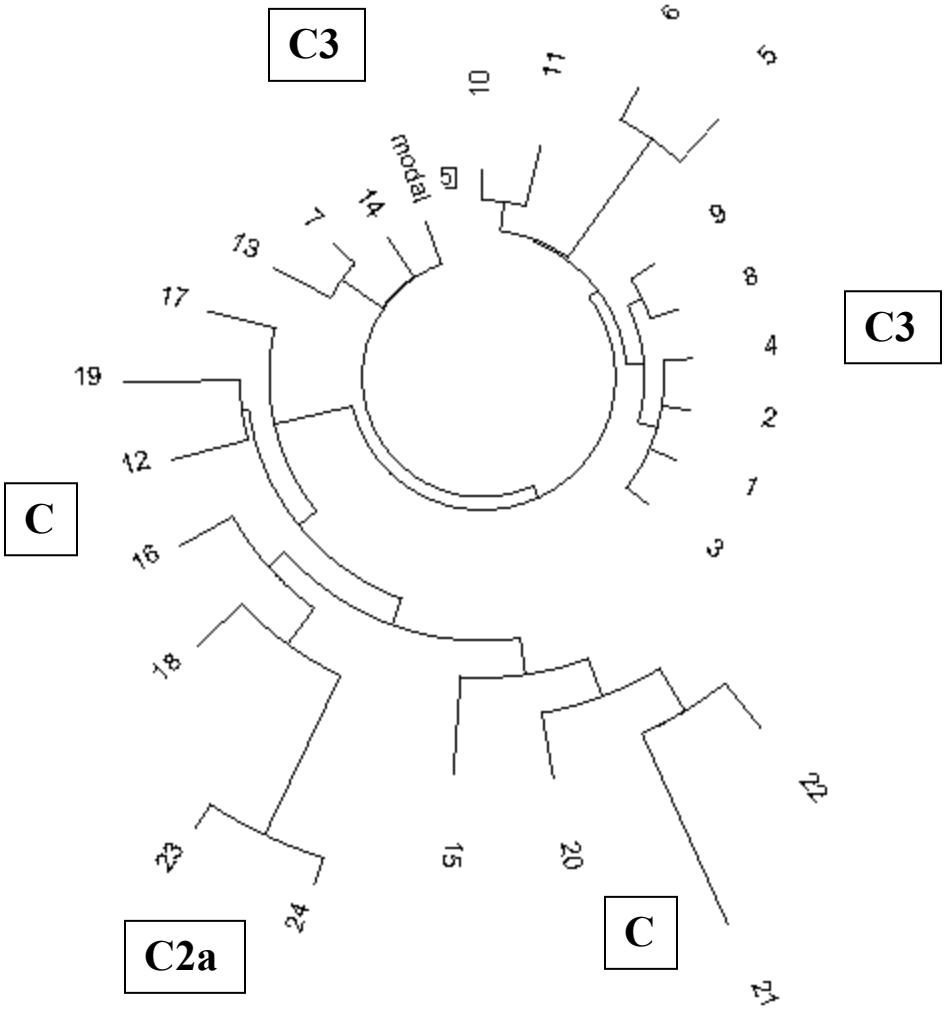


Fig. 5. A tree of 22 of 22 marker haplotypes of haplogroup C (15-22 and, probably, 12), C2a (23 and 24), and C3 (1-11, 13, 14). Haplotypes were taken from YSearch and FTDNA Projects. Haplotypes 1-4 (Poland and Czechoslovakia [N61150, N67268, 3R5AW, 125097]), 5-6 (Ukraine and Hungary [N23982, PFJW3]), 7 (Germany [PETCD]), 8-11 (Kazakhstan [BK4A3, CQYS8, ED5XE, FDTJ6]); haplotypes of the right-hand side, older branch of haplogroup C were mainly not identified, except 20-21 (India [N80268, N76603]), 22 (OAE [ACF8Z], 12 (Korea [6UPP9]) and 15 (Poland [57671]); haplotypes 23 and 24 [6DSW7 and 9YADN] from Hawaii (haplogroup C2a was determined in one of them, however, both haplotypes are identical in 22 marker format, and differ by 14 mutations in the 67 marker format, placing their common ancestor at 1925±550 years before present).

All seven base haplotypes of the branches above have 45.5 mutations from the ancestral haplotype, which gives 45.5/7/0.006 = 1083 → 1273 generations, that is 21,800 years, plus 4500 years of the

averaged "age" of all the branches, resulting in 36 thousand years to a common ancestor of haplogroup C based on available haplotypes (Figs 4 and 5).

A collection of 27 haplotypes of haplogroup C in Ysearch is represented by mainly short haplotypes (one-third is 12 marker haplotypes, one-third is 25 marker haplotypes, the rest is mainly 17 marker haplotypes), however, all of them fit the base 22 marker haplotypes show above in the respective slow markers from the 22 marker panel.

22 marker base haplotype of haplogroup C differs from the base haplotypes of haplogroups A and B, shown below, by 18 and 8 mutations, respectively. This generally fits well with the expected dynamics of mutations in base haplotypes of these haplogroups. 18 mutations translate into 3000 generations, with correction for back mutations (a correction coefficient is 1.633) to 123,000 years between common ancestors of haplogroups A and C. Since they lived 85,000 and 36,000 ybp, respectively, their common ancestor should has been lived about 122,000 ybp. This is close to 136,000 years for a common ancestor of haplogroups A and B, within margin of error.

8 mutations between the base haplotypes of haplogroups B and C corresponds to 1333 generations multiplied by 1.219 (the correction factor for back mutations), which gives 41,000 years between common ancestors of haplogroups B and C, who lived 46,000 and 36,000 ybp, respectively. It seems that a common ancestor of haplogroups B and C lived approximately 62,000 years before present. It is significantly earlier then the haplogroup B itself, but a reason for such discrepancy is unknown as yet. Either there was a common ancestor of haplogroup C and some extinct haplogroup who lived around 60,000 years bp, or it is some inaccuracy in calculations of the TSCA for haplogroup C, or some yet unknown reason. It is more likely that haplogroup C descended from haplogroup B, despite the discrepancy.

Verification of the method using haplogroups R1a1 and R1b1a2

R1a1 and R1b1a2 are relatively young haplogroups, among the last ones on the phylogenetic ladder of haplogroups. It is therefore expected that mutational distances between the base haplotypes of haplogroup R1a1 and R1b1a2, on the one hand, and those of haplogroups A, B, and C, on the other, will be large. Hence, a question, how large and whether they (the distances) would fit into the emerging chronological pattern?

Let us revisit the conclusions of the preceding material. All the three base haplotypes are shown below, the respective "age" of their common ancestors are 85, 46 and 36 thousand years.

12 11 11 – 9 11 – 10 – 10 8 14 15 7 10 8 12 13 11 16 8 13 9 11 12 (A)
11 12 11 – 11 11 – 10 – 11 8 16 16 8 10 8 12 10 11 15 8 12 11 12 11 (B)
11 13 11 – 11 11 – 10 – 10 8 16 16 8 10 8 12 12 12 13 8 12 11 11 11 (C)

Common ancestors of haplogroups A and B, and A and C lived, respectively, 130 and 122 thousand years before present, which is practically the same thing within margin of error, and those of haplogroups B and C lived approximately 62 thousand years ago, however, likely 46 thousand years before present.

R1a1

The base haplotype of haplogroup R1a1 on the Russian Plain in the 22 marker format is as follows (it is derived from the data of Rozhanskii and Klyosov, 2009):

12 **12** 11 – **11** 11 – **11** – **11** 8 **17 17 8** 10 8 12 **10 12 12** 8 **12 11** 11 12 (**R1a1, ~5000 ybp**)

It differs from the base haplotype of haplogroup A by 22 mutations (marked above). This is indeed the largest mutational difference compared with that between A and B (18 mutations), A and C (18 mutations) and B and C (8 mutations). Therefore, the overall pattern is supported with the new data.

22 mutations correspond to 170 thousand years between the common ancestors of haplogroups A and R1a1 (the European branch), who lived 85 and 5 thousand years ago, respectively. Therefore, a common ancestor of haplogroups A and R1b1 lived (170,000 + 85,000 + 5,000)/2 = 130,000 years before present. We again came to a common ancestor of the "alpha haplogroup". It should be emphasized, that this dating was obtain based on independent data, that is the base R1a1 haplotype of 5,000 ybp. 22 mutation difference between the base haplotypes A and R1a1 is an objective information, a direct observation, which has not been used earlier in this study. The "alpha" common ancestor was a common ancestor of both haplogroups A and "beta" haplogroups, and R1a1 (the European branch) descended from the "beta".

We do not know as yet where this "alpha" ancestor lived 136 thousand years ago. It was not necessarily Africa. The haplogroup tree is not rooted.

The base haplotypes of haplogroups B and R1a1 (the European branch) differ by 11 mutations. It gives 60,700 years to THEIR common ancestor. This in turn places a common ancestor of both haplogroup B and R1a1 (the European branch) to 56 thousand years before present. This apparently is the haplogroup B itself, since all these estimates have margin of error not less than ±5,000-10,000 years. In this case 46,000±5,000 ybp (haplogroup B) and 61,000±10,000 ybp (a common

ancestor of haplogroups B and R1a1), as well as 62,000±10,000 ybp (a common ancestor of haplogroups B and C) are within margins of error.

The base haplotypes of haplogroups R1a1 and C differ by 10 mutations, which corresponds to 53,700 years between their common ancestors. Haplogroup C which is only 36,000 of "age" cannot be this common ancestor. Indeed, the phylogenetic tree of haplogroups does not show haplogroup C as an ancestor of R1a1, and bearers of R1a1 do not have the SNP M130, which identifies haplogroup C and its downstream subclades. Calculations show that a common ancestor haplogroups C and R1a1 lived 47,000 years ago, and it is the most likely haplogroup B, which is 46,000 years of "age" (see above).

R1b1a2

A typical base haplotype of the European branch of haplogroup R1b1a2 with an "age" of around 7,000 years (if to count from the subclade M269) is as follows:

12 12 13 – 11 11 – 12 – 11 9 15 16 8 10 8 12 10 12 12 8 12 11 11 12 (**R1b, ~7000 ybp**)

It is expected that between the base haplotypes of R1a1 and R1b1a2 is at least 25,000 years. This comes from an estimate that the first arose 21,000 ybp (Klyosov, 2009b) and evolved for 16,000 years to become the base haplotype with the "age" of 5,000 years (see above), and the second one arose 16,000 years ybp (Klyosov, 2009b) and evolved for 9,000 years to become the base R1b1a2 base haplotype shown above. In other words, they were evolving for at least 16,000 + 9,000 = 25 thousand years separately from each other.

In fact, the base 22 marker R1a1 and R1b1b2 haplotypes shown above differ from each other by 7 mutations, which gives 7/0.006 = 1167 → 1380 generations (the correction factor for back mutations equals to 1.187 in this particular case), that is 34,600 years between their common ancestors. It means that THEIR common ancestor lived approximately (34600+5000+7000)/2 = 23,300 ybp. This was a common ancestor of haplogroup R1. This consideration shows that the methodology of this study is appropriate.

There are 23 mutations between the base haplotypes of haplogroup A and R1b1a2 (marked below):

12 **12 13** – **11** 11 – **12** – **11 9 15 16 8** 10 8 12 **10 12 12** 8 **12 11** 11 12 (R1b1a2)

It is close to 22 mutations between A and R1a1, which could have been expected. This places a common ancestor of haplogroups A and Rib1a2 by approximately 136,000 ybp, that is to the same timeframe as that for haplogroups A and R1a1,

130,000 years before present. The difference is less than 5%, and well within margin of error of calculations.

The base haplotypes of haplogroups B and R1b1a2 (the European branch) are separated by 13 mutations (there were 11 mutations between B and R1a1):

11 12 11 – 11 11 – 10 – 11 8 16 16 8 10 8 12 10 11 15 8 12 11 12 11 (B)
12 12 **13** – 11 11 – **12** – 11 **9 15** 16 8 10 8 12 10 **12 12** 8 12 11 **11 12** (R1b1a2)

This corresponds to 76 thousand years between their common ancestors, and placed THEIR common ancestor at 64,000 ybp. This is somewhat higher than 56,000 ybp for a common ancestor of haplogroups B and R1a1, however, could be within margin of errors of calculations, as it was considered above.

As it was already discussed above, haplogroups R1b1a2 were not supposed to descend from haplogroup C, however, let us check their base haplotypes. There are 13 mutations between them, which corresponds to 76,000 years between them. Clearly, that haplogroup C with its "age" of 36,000 years cannot be an ancestral R1b1a2. A common ancestor for both of them should have lived approximately 60 thousand years before present. It was likely a common ancestor of haplogroup B.

Haplogroups I, I1 and I2

It was suggested (Klyosov, 2010b), based on craniological data, archaeology, anthropology and DNA genealogy, that haplogroup I (or its parent haplogroup) arose on the Russian Plain more than 40,000 ybp, and then migrated westward, to Europe. Let us verify this dating using "slow" 22 marker haplotypes.

Fig. 6 shows a 67 marker haplotype tree of haplogroup I1d1, the most wide-spread subclade of haplogroup I1 in Europe (Klyosov, 2010a). Its 67 marker base haplotype is as follows:

13 23 14 10 14 14 11 14 11 12 11 28 – 15 8 9 8 11 23 16 20 28 12 14 15 16 – 10 10 19 21 14 14 17 21 35 37 12 10 – 11 8 15 15 8 11 10 8 9 10 12 23 25 15 10 12 12 16 8 13 25 20 13 13 11 12 11 11 12 11

Taking into account mutations, it was calculated that its common ancestor lived 2275±330 ybp (Klyosov, 2010a).

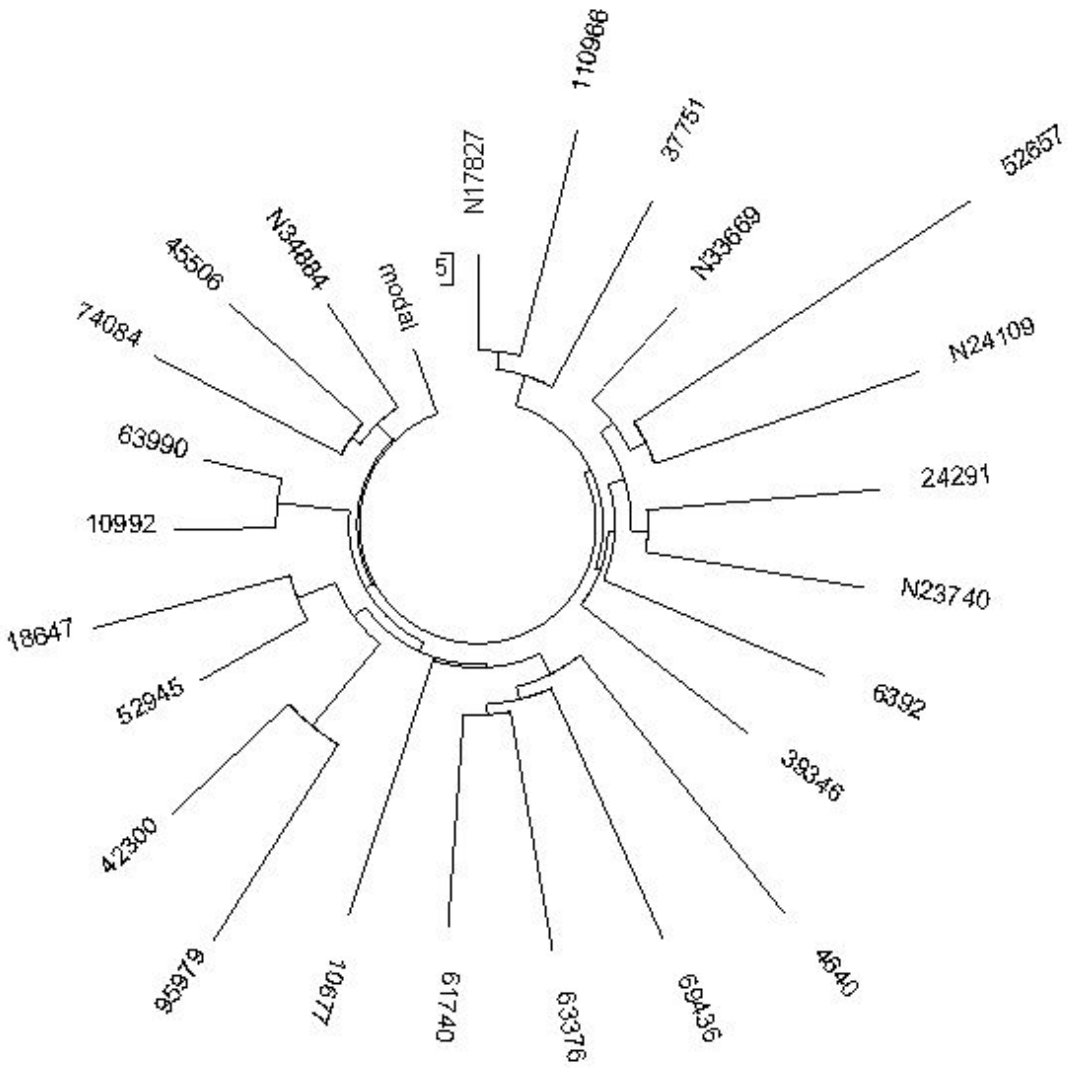


Fig. 6. A tree of 24 of 67 marker haplotype of haplogroup I1d1-P109. Haplotypes were kindly provided by Rebekah Canada, Administrator of Y-DNA Hg Project I-P109.

This base haplotype in the 22 marker format can be presented as follows:

11 14 11 -- 8 11 – 10 – 11 8 15 15 8 10 8 12 10 12 12 8 13 11 11 12
(**I1d1-P109, 2275±330 ybp**)

This base haplotype is typical for populations of I1 haplogroup in Europe, from British Isles to Scandinavia, Germany, Central and Eastern Europe and to the Middle East (Jordan, Lebanon, Turkey, Jews of haplogroup I1). All of them have a common ancestor who lived between 3500-3200 ybp, and it is much less than for one mutation to occur in 22 marker haplotypes.

Since the same study (Klyosov, 2010a) provides the base 67 marker haplotypes for subclades and branches of haplogroup I2, they can be presented in the 22 marker format (the base haplotypes of two I2a2 branches are also shown):

11 13 11 -- 11 11 – 10 – 11 8 16 16 8 10 8 12 11 12 12 8 13 10 11 12	(I2a1)
11 13 11 -- 11 11 – 10 – 11 8 **15** 16 **7** 10 8 12 **10** 12 12 8 **12** 10 11 12	(I2a2 – old)
11 13 11 -- 11 11 – 10 – 11 8 **15 15 7** 10 8 12 **10** 12 12 **7 14** 10 11 12	(I2a2 – young)
11 13 12 -- 11 11 – 10 – 11 8 15 16 8 10 8 12 11 12 12 9 13 12 12 12	(I2b1)
11 13 **11** -- **10 12** – 10 – 11 8 15 16 8 10 8 12 11 12 12 **8** 13 12 12 12	(I2b2)

Mutations are marked between the three I2a haplotypes, and between two I2b haplotypes.

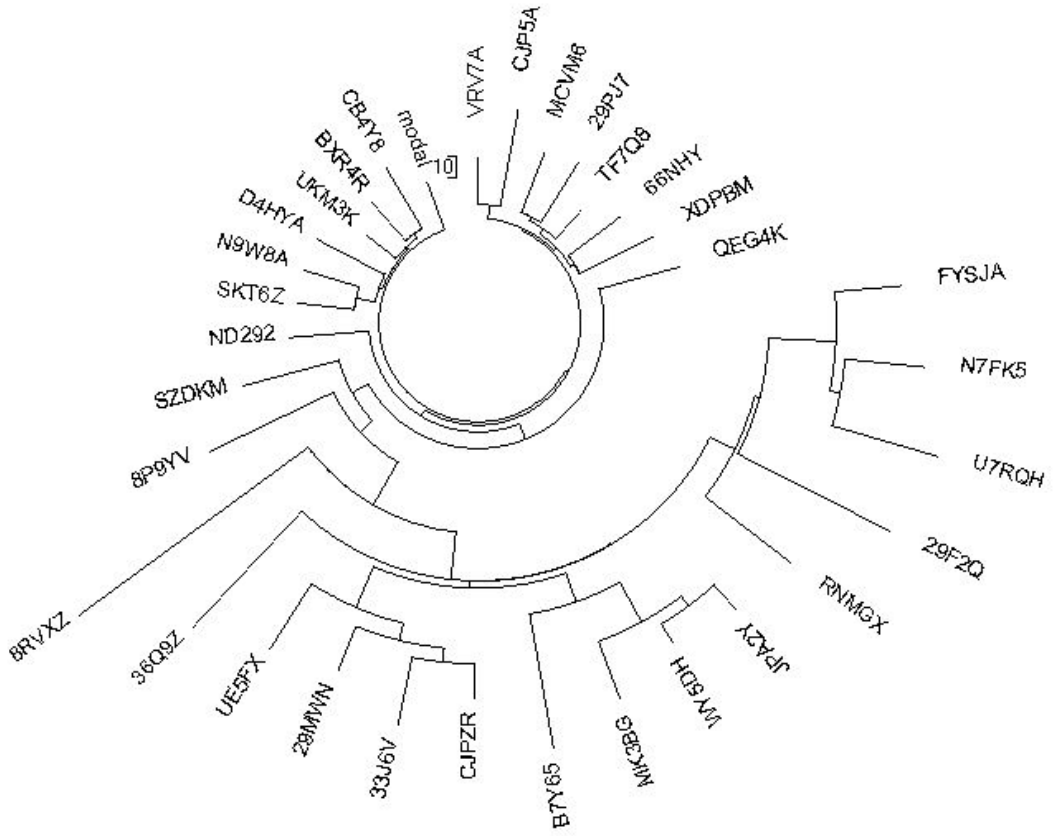


Fig. 7. A tree of 32 of 67 marker haplotypes of haplogroup I2a2-M26. The haplotypes were kindly provided by Tibor Feher, Administrator of the Project. The "age" of the younger (on the top) and the older (the bottom) branches is 2275±380 and 6250±800 years, respectively (Klyosov, 2010a).

These mutations immediately show that lineages of these subclades are descending from very ancient common ancestors, since one mutation corresponds to 167 "conditional" generations (25 years each), that is to 4125 years.

As an example, a 67 marker haplotype tree of haplogroup I2a2-M26 is shown in Fig. 7. The tree contains a young and an old branch, with their "age" of 2275±380 and 6250±800 years, respectively (Klyosov, 2010a). Haplotype trees for all three haplogroups listed above are more uniform ones, and their "age" is 5600±620 years (I2a1), 5700±590 (I2b1) and 5000±630 (I2b2) (Klyosov, 2010a).

Two I2a2 branches, the young and the old ones, differ by 4 mutations, which corresponds to 667 → 733 generations, that is 18,300 years between their common ancestors, and place THEIR common ancestor of haplogroup I2a2 at 13,400 ybp. Then, the base haplotypes I2a1 and I2a2 differ by 4 mutations, which places a common ancestor of haplogroup I2a at 18,675 ybp.

Haplogroups I2b1 and I2b2 differ by 4 mutations as well, which places a common ancestor of haplogroup I2b at 14,500 years bp.

It is already shows that I2 is a very ancient haplogroup, since its two parallel sub-branches are old, and on top of it they further differ by 7 mutations, which corresponds to 34,600 years between their common ancestors. This places THEIR common ancestor, that is a common ancestor of haplogroup I2, at 34,000 ybp. Its base haplotype could have looked as follows:

11 **13** 11 -- **11** 11 – 10 – 11 8 15 **16** 8 10 8 12 **11** 12 12 8 13 **10** **12** 12

Obviously, a common ancestor of its parent haplogroup, I, should have lived even earlier, that is seemingly earlier than 40,000 ybp. It seems that an estimate of "more than 40 thousand years before present" based on a number of considerations (see above) is already justified.

A date when haplogroup I1 arose is uncertain, since almost all haplotypes of this haplogroup that are available come back to around 4,000 ybp, when its bearers passed a population bottleneck. There are a few allegedly I1 haplotypes available with a common ancestor of approximately 17,000 before present (Klyosov, 2010a). If these haplotypes are taken into consideration, then the base haplotypes of haplogroup I1 and I2 differ by 8 mutations, or 41,000 years of difference, and THEIR common ancestor, that is a common ancestor of haplogroup I lived 46,000 years before present. In fact, this fits a set of archaeological data on the Russian Plain, and that those paleoeuropeans seemingly migrated to Europe around 45-40 thousand years before present (Klyosov, 2010b, and references therein).

Let us compare the base haplotypes of haplogroups A, B, and I2. There are 18 mutations between A and I2, which place THEIR common ancestor at 127,000 ybp, almost exactly with the "alpha" common ancestor of 136,000±9,000 ybp.

There are 10 mutations between the base haplotypes of B and I2. It corresponds to 53,700 years between their common ancestors, and places THEIR common ancestor at approximately 67,000 ybp, at the same timeframe (64,000±6,000 ybp) for common ancestors of haplogroup B and C, B and R1a1, B and R1b1b2, and B and I2. It seems that there was a common ancestor of all non-African haplogroups, who lived 64,000±6,000 years before present, that is much earlier than the common ancestor of the African haplogroup B, and who is missed in the phylogeny of non-African haplogroups.

If this (tentative) conclusion is correct, then the non-African alpha-haplogroup (136,000 ybp) was ancestral to the African haplogroup A (85,000 ybp) which currently live in Africa in their majority. This alpha-haplogroups was also ancestral to the beta-haplogroup, which was not identified as yet in the phylogeny of haplogroups, and which arose 64,000±6,000 ybp, apparently outside of Africa. It was ancestral to haplogroupB and all C-T haplogroups and their downstreams, including haplogroups C (36,000 ybp), I (46,000 ybp), I2 (34,000 ybp), R1 (23,000 ybp), R1a1 (21,000 ybp), R1b1a2 (16,000 ybp). Then a major discrepancy between "population genetics" and anthropology/archaeology might be resolved, according to which the first insist on the "out-of-Africa" theory, despite the haplogroup tree is not rooted, many anthropologists insist that anatomically modern man had appeared in Eurasia and certainly not in Africa, and specialists in genome studies take an uncertain central position, albeit prefer to interpret their finding (often still rather vague) towards the "out-of-Africa" concept.

The Part 2 of this study presents data for other principal haplogroups, such as D, E, F, G, H, J, K, L, M, N, O, T.

ADDENDUM

In the Table below:

- The first column shows a number of mutations between two base haplotypes in the 22 marker format,
- The second column show a number of "conditional" generations (25 years each) based on the data of the first column and employing the mutation rate constant equal to 0.0060 mutation/haplotype/generation (Klyosov, 2011),
- The third column lists a correction factor for back mutations, calculated with the following equation (Klyosov, 2009a)

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

Where:

λ is the corrected average number of mutations per marker,

λ_{obs} is a observed number of mutations per marker(explanations and references are given in [Klyosov, 2009a, Klyosov, 2011),

- The fourth column shows the distance in years between the common ancestors of the two base haplotypes with a mutational difference listed in the first column, with a correction for back mutations. A time span to a common ancestor of the both base haplotypes equals to this distance in years (fourth column) plus a timespan to the first common ancestor plus a timespan to the second common ancestor with the sum divided by 2.

A number of mutations between two base 22 marker haplotypes	A number of generations between the common ancestors of the two base haplotypes w/out a correction for back mutations	The correction factor for back mutations	The distance in years between the common ancestors of the two base haplotypes with a correction for back mutations
1	167	1.023	4,250
2	333	1.048	8,700
3	500	1.073	13,400
4	667	1.100	18,300
5	874	1.128	24,600
6	1000	1.157	28,900
7	1167	1.187	34,600
8	1333	1.219	40,600
9	1500	1.253	47,000
10	1667	1.288	53,700
11	1833	1.325	60,700
12	2000	1.363	68,150
13	2167	1.430	76,000
14	2333	1.445	84,300
15	2500	1.489	93,000
16	2667	1.535	102,300
17	2833	1.583	112,100
18	3000	1.633	122,500
19	3167	1.686	133,500
20	3333	1.741	145,100
21	3500	1.799	157,400
22	3667	1.860	170,500
23	3833	1.922	184,200
24	4000	1.989	198,900

25	4167	2.058	214,400

The author wishes to thank Vladislav Ryzhkov and Igor Rozhanskii for help in collection of haplotypes and for discussion of the study and its results. The author also appreciates help by Rebekah Canada and Tibor Feher in providing haplotype series.

References

Adamov, D.S., and Klyosov, A.A. (2008) Calculations of "age" of populations based on Y chromosome data by the method of averaged quadratic deviations. Proc. Russian Academy of Genealogy (ISSN 1942-7484), 1, № 5, 855-907.

Klyosov, A.A. (2010a) Haplogroup I. Proc. Russian Academy of Genealogy (ISSN 1942-7484), 3, № 1, 96-158.

Klyosov, A.A. (2010b) DNA genealogy, craniometry, and origin of Europeoids. Proc. Russian Academy of Genealogy (ISSN 1942-7484), 3, № 8, 1256-1309.

Клёсов, А.А. (2011) The slowest 22 marker haplotype panel (out of the 67 marker panel) and their mutation rate constants employed for calculation timespans to the most ancient common ancestors. Proc. Russian Academy of Genealogy (ISSN 1942-7484), 4, № 5, 968-987.

Rozhanskii, I.L. and Klyosov, A.A. (2009) Haplogroup R1a1: haplotypes, lineages, history, geography. Proc. Russian Academy of Genealogy (ISSN 1942-7484), 2, № 6, 974-1099.

Ballantyne, K.N., Goedbloed, M., Fang, R., Schaap, O., Lao, O., Wollstein, A., Choi, Y., van Duijn, K., Vermeulen, M., Brauer, S., Decorte, R., Poetsch, M., von Wurmb-Schwark, N., de Knijff, P., Labuda, D., Vezina, H., Knoblauch, H., Lessig, R., Roewer, L., Ploski, R., Dobosz, T., Henke, L., Henke, J., Furtado, M.R., Kayser, M.(2010) Mutability of Y-chromosomal microsatellites: rates, characteristic, molecular bases, and forensic implications. Am. J. Human Genet. 7, 341-353.

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

Klyosov, A.A. (2009a) DNA Genealogy, mutation rates, and some historical evidences written in Y-chromosome. I. Basic principles and the method. J. Genetic Genealogy, 5, 186-216.
http://www.jogg.info/52/files/Klyosov1.pdf

Klyosov, A.A. (2009b) DNA Genealogy, mutation rates, and some historical evidences written in Y-chromosome. II. Walking the map. J. Genetic Genealogy. 5, 217 – 256.

"Service by Blood and Death". Klyosov (R1a1, West-Eurasian branch) Family Story as a Mirror of Russian History

Anatole A. Klyosov

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

Between us, I like to consider myself as a descendant of Boyar Children. This is a deeply Russian thing. It carries a flavor of ancient stories, legends, folk songs.

However, if you were to randomly ask folks out there on the street who Boyar Children were, nine out of ten would have only a vague idea except they would say that it sounds awfully ancient. Maybe even 99 out of 100 would say no more than that. And I am talking not of Boston where I live now. I am talking about present day Russian cities or villages. The reason is a simple one. The 20th century in Russia was so violent and turbulent that memories of our roots and our past were wiped out rather effectively.

Dvoryane and Boyar Children

Boyar Children were a part of the top society in ancient Russia. The top society in its entirety was consisted of Boyars, Counts Men, Courtiers and Boyar Children. After a while, the first two groups had merged with the last two, and the last two groups had in turn merged and became Dvoryane (in Russian), or Courtiers. This merging occurred during 17-18 centuries.

The word Courtiers literally means "people of the Court"; first the Count's court, then the Royal, Tsar Court. Initially, the Dvoryane were assembled into the Count Squadron. The Squadron was the best part of the local military forces. Every city had a squadron, or a platoon, or a division of Boyar Children. Ordinary folks had taken arms only in extraordinary situations, and peasants never did. However, the peasants had to provide horses and food for Dvoryane and Boyar Children.

Dvoryane and Boyar Children always were military folks. Using present day language, they were military officers. More than that, they were dynasties of military officers. They had formed a rather closed society of dynasties of combat servicemen. As Russian historians of the past have written, Court Squadrons

were formed from people who not only had clearly shown bravery and valor on a battlefield, but were also of a distinct "breed". It does not sound too politically correct nowadays, but those who enrolled them into the Squadrons probably knew better. Or at least they had their own opinion.

Boyars were the top strata in ancient Russia, or Rus as it was called back then. Initially, they were children and close relatives of the Count, the closest blood, combat buddies, the backbone of his squadron. Over time, they became the closest advisors to the Count. Their children were called, literally, Boyar Children. Since that collective name had carried an important message of nobility, their children and grandchildren continued to be called Boyar Children. They gradually became a rather large group within the military elite, and formed a middle- and junior-level contingent military. This, of course, was not related to the lower and lowest levels of troops since those were not dynasty servicemen. Using present-day language again, Boyar Children commonly filled positions from plain horsemen, in noble dynasty cavalries, to majors, depending on experience and duration of service. Colonel and General ranks were held by Counts, Voevodas and, preferably, other Royal blood people.

The word Dvoryane was recorded the first time in a historical document called the Letopis (a series of ancient formal diaries all "signed and numbered") of 1175. The use of the term "Boyar Children" as a top society can be found in the Letopis of 1281 which described the capture of a traitor Count named Mikhailo Glinski when he was leaving to join Polish troops who were enemies that time. The Letopis says "Mikhailo Glinski is riding alone a mile ahead of his Dvoryane, and Count Mikhailo Golitsa captured him, and the Boyar Children captured all of Glinski's Dvoryane".

Initially, Boyars and Boyar Children were of a higher status compared to the Dvoryane since they were strictly military people while many Dvoryane had just maintained the Court though holding responsible positions. In the Court Squadron, they were at the same level and practically indistinguishable from Boyar Children, although as late as in the first half of the 16th century, Boyar Children were of a higher class compared with the Dvoryane. But then with an elevation of prestige of the Tsar and his court, a status of Courtiers/Dvoryane began to exceed the military status of the Boyar Children. In the 16th and 17th centuries, both terms have practically merged, or rather, stayed together and are practically always referred to as "Dvoryane and Boyar Children", in contrast to other people, draft soldiers, peasants, villagers. Both Dvoryane and Boyar Children have served "from their land", as it was called those times, meaning that they were receiving land for their military service and had to justify the possession by continuing their service and service of their sons and grandsons.

Until the 18th century, the term "Dvoryane and Boyar Children" was the equivalent to the term "warriors". An expression "to send ahead Dvoryane and Boyar Children" meant to direct the best and brightest cavalry. But this was not all it meant. In the Kursk area, the very South of Russia in those days, with Tatars from the South and South-West, Poles from the West, Lithuanians from the North-West, the Kursk area that will be our focus later on, with Kursk as its ancient capital, was completely destroyed by Tatars in the 13th century. It was rebuilt in the 16th and 17th centuries by Dvoryane and Boyar Children who were also military engineers and builders. There were no peasants in the Kursk area at that time to be drafted to build fortresses. In fact, all cities at that time in the Kursk area were military garrisons. It was too dangerous for peasants to move there voluntarily. And peasants were free to move at that time anywhere in Russia.

As described by historians of those times, Dvoryane and Boyar Children were the most educated people. They were carriers of culture, they were the most decent strata of Russian society, and role models in all aspects of life.

In the 18th century, the term Boyar Children was disappearing and being completely replaced by the term Dvoryane. For those who retired from military service to their private land, particularly for those who have not served in military forces during the First government revision (1710-1720s), both terms were gradually replaced by the terms "single-courters". Furthermore, those retirees settled on their land which they had earned for their military service and were becoming free farmers. This was typically the end of their nobility. Some of their descendants were able to return to nobility if they successfully claimed former Dvoryane status. This status and title could be given back to them only after special consideration and was additionally confirmed by returning to military service.

In 1566, Ivan the Terrible (Ioann IV, son of Vassily) had determined three tiers for dynasty servicemen. The top, first tier were Moscow Dvoryane and provincial Dvoryane with Moscow Dvoryane serving and owning land in the Moscow area. Provincial or City Dvoryane, had served and owned land in other areas, in and around other cities and towns in Russia. The second tier was Boyar Children. They had the same rights as did Dvoryane, however, they typically kept lower military positions and ranks. They could move to the first tier for military achievements. The third tier was for shooters ("strelzy"), cannon-men, and other servicemen. They could be from Dvoryane, but not from peasants and other "plain people". They formed regiments which were headed by Dvoryane and Boyar Children. The first two tiers had served "by heritage", from their land, and children often continued to carry military ranks of their fathers. The third tier had served "by draft".

In 1642, Tsar Mikhail Romanov (the first Tsar Romanov) had issued a verdict ("ukaz") which banned by-heritage-servicemen to become peasants and plain soldiers. A verdict issued in 1675 prohibited tier three draft soldiers to become Boyar Children for whatever achievements. These verdicts had expedited forming of a closed Dvoryane society in Russia.

When Feodor, son of Ivan the Terrible, was Tsar of Russia (1584-1598), the Russian Army included 80 thousand horsemen in the Dvoryane Cavalry. This cavalry was formed exclusively from Dvoryane and Boyar Children. "High Dvoryane", that is Boyars, had received a salary of 70-100 rubles a year, "Middle Dvoryane" 40-60 rubles, and Boyar Children 20-30 rubles a year.

There were few cases in Russia when the status of Boyar Children was granted to lower class people and even to peasants, however, these cases, according to historians, were typically unsuccessful. Boris Godunov, son of Feodor, had exercised several such cases, but they happened to be "infirm" and were retracted by subsequent Tsars.

Boyar Children were an absolutely inherent part of the Russian military. On the Southern frontiers of Russia, in the Kursk area, they served as reconnaissance officers, entered into combats, leading their "tens" and "hundreds" of troops, set defense of their cities and towns, storm enemy fortresses, and carried all uneasy military duties. For their service, they were paid in the form of land besides monetary compensation. This was called "outlanding".

Kursk was the largest city in the South-West and South-East of Russia. To the South-West of the Kursk area there was a huge territory called the Wild Plain.

Outlanding of Dvoryane and Boyar Children

There were two principal mechanisms of outlanding. One mechanism was a scheduled draft, it was announced in a formal document and indicated who and how much land and money was granted. Another mechanism was an "application for land" to the local or central authorities, including the Tsar himself, when a serviceman or his sons asked for a grant of land for their father.

And here I will mention my forefather, Ivan Klyosov (born 1580), who was the earliest Klyosov found in the archives, and who was granted 300 acres of land in the Kursk area in 1639. The Granted Land Book describes it as follows (the Russian National Archives of Ancient Acts [RNAAA], 1209-188-15684, p. 159)

Year 7147 (1639) April 12 day according to the manuscript by Tsar and Grand Count Mikhail son of Feodor of all Rus, a Boyar Son Ivan Klyosov, of Reitar Service, following

an application by Kursk citizens, Boyar Children Kirey Klyosov, Frol Evsyukov, and Denis Pyzhov and Mina Vozhov and Ostakh Shipilov and by directives by Stolnik and Voevoda Ivan son Vassiliy Butourlin, in Kursk stan of Kursk region, fields and forests were signed off from the mouth of Khmelevsk River upstream and downstream along the right-hand bank of Prut River and after the field investigation the land was measured and recorded as hundred acres into the field and twice as much according to Tsar and Grant Count Mikhail son of Feodor of all Rus the manuscript and the investigation.

After 56 years, *the Records of Boyar Children in the City of Kursk and Kursk region of 1695* (Archives of the Kursk Region, 1555-1-168, p. 678) mentioned for the first time the Klyosov village. In 15 more years, according to the Census of 1710, the Klyosov village was recorded in a Section of "Reitars and Single-Courters", that it had 12 courts and 35 inhabitants, of which seven were adult men and nine were adult women. The rest were children, and everyone there had Klyosov as their last name.

It is worth mentioning here that until the 15th century, last names in Russia were not in use even among Boyars. By the end of the 15th century, last names were more and more frequent among Dvoryane. In the 16th century, practically all Dvoryane and Boyar Children had last names, unlike lower class citizens, whose last names had begun to appear only in the 19th century.

The fact that those people in 1639 who were listed in the Tsars manuscript had last names places them among rather distinguished people.

The Ukaz (verdict) of Tsar Alexey Mikhailovich of 1648 lists norms of land compensations for Boyar Children in the Kursk-Belgorod region:

-- Upper size grants (1st and 2nd tier): 400 and 300 acres
-- Middle size grants (3rd, 4th and 5th tiers): 250, 200 and 150 acres
-- Lower size grants (7th and 8th tiers): 100 and 70 acres

Land was granted only to dynasty service men. An "Order" of 1649 by Alexey Mikhailovich banned people not of Dvoryane and Boyar Children heritage to own land. As a rare exception, plain soldiers were granted 25 acres at best. When a landowner's sons, and only dynasty service men were landowners, reached 15-18 years of age, they were signed up to the Tsar's service and thereby acquired rights to have land themselves. After that, they were obliged to respond immediately to each military call and depart to military expeditions with weapons and supplies, bringing other horsemen and foot soldiers depending on the size of their private land. Typically, the requirement was one armed horseman per each 100 acres of private land (an Ukaz by Ivan the Terrible of September 20, 1556). Landowners could not sell their land, the land had to stay

with the dynasty service men while they served. The rule was that "the service should not diminish and the land should not leave the service".

Sons of wealthy Dvoryane and Boyar Children were registered "from the father's land" and often were not granted extra land. Sons of not-so-wealthy and have-not Dvoryane were registered "in addition to the father's land", that is, extra land was granted.

An "Ukaz of outlanding" of 1555 specified: "Father's land should not be taken away from sons if they are fit to serve". The same rule was confirmed by the "Order" of 1649. The main principle was that "those who serve must have land". This was the underlying principle for the system of outlanding. The follow-up principle was that "those who have land must serve". Only church lands presented an exception.

Present-day readers may have the impression that Dvoryane and Boyar Children were wealthy people and had land with many peasants on them. This was not always true. In 1670's, for example, out of 168 Dvoryane and Boyar Children of the Kursk area registered for service, 99 have not had or were not granted land, the rest were outlanded but as scheduled, "Some got it in a half-force, some even less, and some got just a little". Military service was a tough service. Monetary compensations were minimal, and mainly they were received at war, as a rule, as a prize for capturing or killing enemies, for being wounded, for escaping as a POW. During peaceful times, the compensation was one ruble per man per four-year period.

Last but not least, not all land granted as a compensation was actually obtained. Actual land was often substantially less compared to that granted. Actual land should have been found, and its measurement and legal recording should have been arranged. That was not too easy. Available land often was of a poor quality (marches, ravines, etc.), or too far away, or too close to the enemy. That is why, among provincial Dvoryane, there were many "small land-owners", with much less than 100 acres of land, which was not even enough to present even one horseman to the duty. They often received only 80 or 40 acres or 30 or even 10 acres of land. This was close in size to land which was owned by peasants.

There were 731 Boyar Children in Kursk in 1642. They represented a core of Dvoryane local society. Outlanding compensation for them was between 70 and 500 acres per serviceman, however, most of them did not have peasants on their land. At that time, there were almost no peasants in the Kursk area. Therefore, the landowners had lived by their courts which were single courts and later called "single-courters". Furthermore, in the 18th century they had formed a strata in the Russian society that was formally named "Single-Courters". I will come back to this later with a reference to my ancestors.

In general, in dynasty service families, all children had served either directly or indirectly, regardless of their gender. Sons after 15-18 years of age would get on their horses to defend fatherland, and daughters would get married to supply new defenders for the fatherland. There were, of course, some cases when sons of a wounded retired father were not called into service, but left with him "on tillage", and they would loose all privileges of a dynasty servicemen.

Returning to Ivan Klyosov, a Reitar, for what service and achievements a 59-year old Boyar Son, was land granted by the Tsar's Ukaz? The archives do not provide an answer to this specific question, however, let's take a look at the history of the Russian State from the beginning of the 17th century to 1639 when Ivan received his land.

"Fuzzy times" and the beginning of the House of Romanovs

In 1604, when Boyar Son Ivan Klyosov was 24 years of age, Russia entered into "Fuzzy Times" that lasted nine years until 1613. Boris Godunov was ending his seven-year Tsarist rule. He died unexpectedly in April of 1605. The end of his government was, by all means, a catastrophic one. The Southern lands were devastated by Tatars, there was plague and four years of drought in a row. To add insult to injury, some rumors had appeared and spread that a little son of Tsar Ivan the Terrible had not died in 1591 at the age of nine but had miraculously survived. This had initiated turbulences in the masses who cited God's punishment, because the rule of Boris was unlawful, was achieved through deceit and would not do any good.

In October of 1604, "Dimitry the Impostor" (as he was known later) had entered Moscow State and went to the Kursk region. Initially, he was met with a "fire fight" (an expression of that time, meaning cannon shots) but as a result of a political disagreement between the defenders of the city, Kursk surrendered. After the death of Boris Godunov, the Russian military took Dimitry's side. In June of 1605, the successor of Tsar Boris, little Feodor, was killed. The following month, the widow of Ivan the Terrible and mother of little Dimitry recognized the newly appeared Dimitry as her son. That same month July, Dimitry was approved for the throne. However, his Tsarship did not last long, and he was assassinated less than a year afterwords, in May of 1606, after mother of little Dimitry had admitted that her recognition of Dimitry the Impostor was false.

Vassily Shuisky became the next Tsar for four years period. However, provincial Dvoryane did not attend the ceremony. Fuzzy Times continued.

Meanwhile, the Tatars were taking an advantage of a political and military mess in the center of the Moscow State and increased the intensity of their invasions

into the Kursk-Belgorod region. Besides Tatars, Kossaks and other "thieve groups" invaded as well. That kept Dvoryane and Boyar Children on a permanent lookout and in fights. In August of 1610, a Tsar manuscript arrived to Kursk and other cities of Russia that said "*disturbances and disagreements between the Russians are making the country weaker, and Poles and Lithuanians are moving on to Russian soil and shedding blood and devastating Orthodox Christian churches and monasteries. The Polish King is staying at Smolensk, and the Polish Getman at Mozhaisk, and the Impostor in Kolomensk. Seeing all of this, Tsar Vassily Shuisky decided to resign and become a plain citizen, and wished that all our people unite and protect our Russian Orthodox beliefs and our wives and children*".

Four months after it the Dimitry Impostor the Second was killed in October 1612, Moscow was liberated from Polish and Lithuanian troops.

During these nine years of Fuzzy Times, and the changing Tsars, Kursk Dvoryane and Boyar Children continued to carry out their duties during all the instability and disorganization. Ukaz of the Boyar Duma had directed that all Dvoryane without exception must enlist their service by May 29, 1611, and those who declined would lose their land. The Ukaz, however, stated: "Land should not be taken away from wives and children of fallen and wounded Dvoryane".

In 1613, Mikhail Feodorovich Romanov became the Tsar before he completed 17 years of age. Kursk had sent a whole delegation of Dvoryane and Boyar Children to the Sobor which approved the new Tsar. All of them signed the Approval Manuscript on behalf of Kursk Dvoryane.

Tsar Mikhail was the Russia ruler from 1613 through 1645. This time period was saturated with military fighting for Kursk by Dvoryane and Boyar Children. These were times of placing life in Russia back in order, times of flourishing and elevating Russia. At the same time, it was a period of heavy fighting on the Southern and South-Western borders of Russia, deadly dangerous times for Dvoryane and Boyar Children. What role has Boyar Son and Reitar, Ivan Klyosov, participating in during those times?

In 1612, a year before Mikhail had become Tsar, and when Ivan Klyosov was 32 year of age, Polish Getman Zholkevsky with his 70,000-strong army besieged Kursk. The Polish troops could not take the city and left after several bloody battles in which the Poles had suffered great losses.

In February and March of 1613 there were battles in the Kursk region with Polish and Lithuanian troops that ended with their defeat. Reports to the Tsar described the valor of Dvoryane and Boyar Children. These reports were the first ones received by the newly established Tsar and brought a renewed spirit to the Russian troops.

Meanwhile, there were practically non-stop military expeditions on the Western and Southern frontiers of Russia against Poles, Lithuanians, Tatars and Cherkas (Kossaks). In 1616, the Dvoryane and Boyar Children cavalry in Kursk alone was 753-members strong.

In 1616, Lithuanian troops had invaded the Kursk region, and a 340-man cavalry of Dvoryane and Boyar Children were sent to fight them. In 1617, Lithuanians had invaded the Kursk region again, and the Kursk forwarded 380 mounted Dvoryane and Boyar Children along with foot soldiers with the "fire fight". In 1618, Lithuanians had invaded the Kursk region again. In 1620, the Tsar sent an Ukaz to Kursk in which he ordered Dvoryane and Boyar Children to have "two horses each" in case the Tatars came. Meanwhile, Polish and Lithuanian kings had sent envoys to the Tsar to sign "a truce for 14 years, from the present 127 (1619) to 7141 (1633)".

During this relatively peaceful time period, a number of Dvoryane and Boyar Children had reached significant numbers – 885 in 1625, 864 in 1629, 997 in 1631, 1130 in 1642. However, peaceful times were relatively short, since the Tatars did not sign a truce and did not have any intention for peace. In 1616, they came near Kursk and were completely defeated within 15 miles of the city. Men, women and children who had been taken prisoner were freed. In 1622, a large troop of Tatars was crushed after they entered the Kursk area. In 1623, another large troop returned, and a 300-strong cavalry of Dvoryane and Boyar Children was sent to meet them along with mounted Kossaks and 100 foot soldiers with the "fire fight". There was a heavy battle that devastated the Tatars. In 1628, a regimen of Dvoryane and Boyar Children met the Tatars who had captured many Russians, and freed them. The Tatars went towards Kursk, and within 10 miles from the city, they were defeated and their commanders captured.

In 1632, Lithuanians and Tatars had attacked Belgorod, South of Kursk, and took many prisoners and approached Kursk. United troops from several cities in the area led by Dvoryane and Boyar Children defeated the Tatars in a battle. In 1633, Polish troops burnt Belgorod and the suburbs around it. Being encouraged by that, another Polish army went by Kursk and besieged it. After a long period of time, during which they devastated the suburbs, they lost many troops and left the city. The Poles invaded the area in 1634, and besieged Kursk with a 12,000-strong army. They had stayed at Kursk from the 4th to the 16th of April, and threw fireballs into the city but could not take the city and left.

In response, troops from the Kursk region stormed and took the Polish cities of Borzna and Poltava. This went down in history as the war between Russia and Poland of 1633-1634. A truce was signed on May 17, 1634.

Service in action by Dvoryane and Boyar Children was practically uninterrupted during formal peace and real wars. As a result, Dvoryane and Boyar Children were granted land by verdict of the Tsar Ukaz. Among those who obtained land by this Ukaz in 1639 was Boyar Son, Ivan Klyosov.

Boyar Children and Reitars

The Book of Kursk Registry, 1652, Section "City Boyar Children" (RNAAA 210-10-191) indicates that Kirey, son of Ivan Klyosov, checked in for duty "being on a horse and with a *pishal* (a rifle – AK) and was granted a compensation of 300 acres and ten rubles". Ten rubles was pretty good money at that time. An acre of land in the first half of the 17th century cost about two rubles or a bit less, and this price was practically the same from the North of Moscow to the South, as well as the Kursk region.

Kirey was a seasoned man when called to duty at 47. Indeed, by that time, Dvoryane and Boyar Children served all their lives – literally, until death or disability. Not many of them lived long enough to resign.

Who were City Boyar Children, anyway? Unlike Moscow Boyar Children, they lived and served in other cities and towns. This name appeared when the Moscow State was formed, and courts of the former Counts did not merge with the central, Moscow Court, but remained in other cities and towns. Hence, City Boyar Children were local landowners and "*served from their land*". They were obliged to serve, and represented the principal military force. They represented Dvoryane dynasty servicemen.

Reitars were a separate contingent of Dvoryane and Boyar Children. They were a heavy cavalry. Unlike Dragoons, a light cavalry who would come down from their horse in order to shoot from a rifle and then mount again, the Reitars shot from on top of their horse. Their armor was so heavy that they often would not have been able to mount again in the field after coming down from their horse. Reitar's rules of engagement were basically copied from German and Dutch ones. A Reitar is a close copy of the German Reiter or a rider. Russian Reitars had served in Divisions of Foreign Lines along with Western instructors. The very word Reitars had gone along with the times of Peter the Great, as "Reitar Order" (1651-1680) and then the following "Foreign Order" (until 1701) and "Order of Military Matters" which became the Military Collegiums in 1717.

Various troops with foreign names, including Reitars, were introduced since the beginning of the 17th century under Tsar Boris Godunov, and particularly under Tsar Mikhail Romanov, when the Russian military system had become poor

organization. Reitar regiments were composed of City Dvoryane, Boyar Children and other free citizens primarily with their own horses. To be enrolled as a Reitar, one should come "on horse and armed". Reitars were trained by foreigners and by Russian Stol'niks and Dvoryane, who were already trained.

"Free citizens" were enrolled along with Dvoryane and Boyar Children although it somewhat "*tainted the purity*" of dynasty servicemen. However, there were not too many nobles left after heavy and long wars during the Tsarship of Alexey Mikhailovich, son of Mikhail Feodorovich Romanov (Ruler in 1613-1645). The following numbers speak for themselves. If in 1631 there were 997 Dvoryane and Boyar Children in Kursk, and in 1642 there were 1130, then in 1672, four years before the end of the rule of Tsar Alexey, they were only 614 Boyar Children in Kursk.

Military and combat service of Dvoryane and Boyar Children

The registry "Tales of Kurskers on Service, 1675" (RNAAA, 210-732-40) tells us that Sergey, the son of Kirey Klyosov (born ~1630), Ivan's grandson, reported to Duma Dvoryanin Semyon Ivanovich Zaboryansky that he served his duty to the Tsar in a city regimen, and had a compensation of 250 acres and a salary of 11 rubles. His actual land was 35 acres, and he did not have peasants in his possession. In the past year (1673-74) and the current one (1674-75), he was in a military expedition in the regiment of Count Gregory Romodanovsky.

The fact that Boyar Son Sergey Klyosov did not have peasants in his possession is not surprising since there were very few peasants during the 17th century in the Kursk region. Practically the entire population consisted of military men. Cities and towns were military garrisons. Peasants, who were free to move before 1591, did not have any desire to settle in the dangerous Kursk region. Only in the 18th century did peasants start to move into those areas.

Another registry, "A Sorting Out Book of Boyar Children in the City of Kursk and the Kursk Region, 1695" (The State Archive of Kursk Region [SAKR], 1555-1-168-419) says that Kondrat, son of Sergey Klyosov (born ~ 1655), "is serving regiment Reitar service from Boyar Children instead of my deceased brother Maxim Klyosov" and that he had just returned this 203 (1695) year for a vacation from the Kazikermen expedition where he had served in a regiment of Stol'nik and Voevoda Ilya Mikhailovich Dmitriev-Mamonov. Kondrat's land was 30 acres. The Book also indicates that according to Kondrat's report "he will continue to be on horseback serving our Great Tsars, the rifle of his state *carobin* and a couple of pistols ready for an expedition *and to shooting in its entirety*". Kondrat was then 40 years of age. He had obtained a grant of 200 acres and 7 rubles of money.

Kondrat's brother, Vassily, son of Sergey Klyosov, had reported in the same Book (SAKR 1555-1-168-678) that he had served in the City service in Kursk since 1679, with a compensation of 150 acres and 6 rubles of money, and his land of 17 acres is situated in the Klyosov village of the Kursk stan of the Kursk region.

Here we see two different kinds of service, that is "the regiment service" and "the city service". The city service was a garrison service while the regiment service was service in a military expedition. The regiment service was understandably more honorable.

From time to time, particularly when the intensity of combat service was significantly elevated, there was no available Dvoryane and Boyar Children to call into service. In those cases old resigned dynasty men were called to battle. In the second half of 17th century, the Voevoda of Kursk was allowed by Tsar Ukaz to draft and to direct frontier posts of retired Dvoryane and Boyar Children, who retired because they were old or were sick. These were not isolated cases. Indeed, service in the Kursk region was called "service by blood and death". Documents from the end of the 16th century describe forces of "Turks and Crimea's and Nogay's Tatars, and the Lithuanian King, who fiercely united with Pols and Ugrs [Finns], Germans and other Swedes". They commonly hit on the Southwestern part of Russia, namely Kursk, Rylsk and Putivl regions. Tatars had constantly penetrated the area often dragging captured civilians by the hundreds. Cities and suburbs were often devastated. This kept Dvoryane and Boyar Children on a constant lookout by means of reconnaissance and skirmishes. Distances that should have been covered by riding dynasty servicemen are mind-boggling. Observation towers were spread over huge territories. There were, for example, seven observation towers to the South of Kursk out in the Wild Field, the nearest one in four days of horseback riding or about 300 miles through marshes, forests, prairies. The last one was in 430-450 miles away. The towers were reached in a half-day, a day and sometimes two or three days of horseback riding between them. And all of them could be reached by Tatar troops. To deliver notes, Boyar Children would ride over huge distances often through an enemy territory. That is why their reconnaissance service was considered truly dangerous. Those areas were saturated with servicemen blood.

This was the service of Dvoryane and Boyar Children of the Kursk region. Among them were Boyar Children Ivan Klyosov and his son Kirey, and Kirey's son Sergey, and Sergey's son Kondrat, and Kondrat's brothers Vassily and Maxim, mentioned earlier, and Kondrat's son Afanasy Klyosov. And Kirey's brothers Savva and Luka Klyosovs, Savva's son Ostakhei (Boyar Son, service calls in 1626, 1636, 1639, and 1645), Ostakhei's son Ivan (Reitar, service call in 1675), Ivan's son Mikhail (Reitar, call in 1697), Luka's son Anton (Boyar Son, calls in 1639 and 1645), Kirey's son Prokofiy (Reitar, calls in 1675 and 1682),

Prokofiy's son Ivan (Reitar, call in 1697), Ivan's brother Antip (Reitar city service, call in 1697). And also Iov Klyosov and his son Samoyla (Reitar, call in 1697), Dementiy Klyosov and his son Afanasiy (Boyar Son, call in 1630), Afanasiy's son Pavel (Reitar, call in 1675), Pavel's son Potap (Reitar, call in 1697), Karp Klyosov and his son Leon (Reitar, call in 1697), Leon's brother Mikhail (Reitar city service, call in 1697), one more Leon Klyosov and his son Stepan (Reitar, same call in, 1697), Timofey Klyosov and his son Savva (Boyar Son, call in 1645), Kiril Klyosov and his son Trofim (Reitar city service, call in 1697) and Zakhar Klyosov, a cannonman (call in 1645).

All thirty-two Klyosovs were Kursk servicemen.

As it has been repeatedly stated, those times in the Kursk area were quite dangerous. As mentioned earlier, Ivan Klyosov was granted land in 1639. Seven years after that, in 1646, Tatars undertook a large invasion of the Kursk region. Thirty four settlements on granted land were completely destroyed and their inhabitants were kidnapped, mainly young Dvoryane, women and children, whole families, whose men were serving at that time or were killed. Being encouraged by their success, Tatars moved to another area to the neighboring Rylsk. They managed to devastate several settlements there, however, a regiment of Rylsk Boyar Children had attacked them, and in a bloody battle, freed the women and children, a total 439 people. As a historian wrote, "This explains why every Dvoryane and Boyar Son had armaments hanging on a wall next to his bed.

One of the largest *pogroms* (destruction of households) by Tatars had happened in 1680 in the Kursk-Belgorod region. They killed and kidnapped 471 men and 368 women. Most of those killed and captured were Dvoryane and Boyar Children with their wives and children. According to historical sources, Tatars beat to death defenseless children from infants to 15-year olds. Two hundred and ninety four children died including twenty eight boys of 4-years old or less, and forty two girls were burnt alive. In 1680 alone, Tatars killed and led into captivity 3,258 men, women and children, and took away 24,193 cows and horses, 4,828 beehives and burnt to the ground four churches, 688 courts, four windmills and eight settlements. Captured people were sold into slavery in Crimea, Constantinople, Asia Minor, and other countries and regions.

When Tatars appeared in the neighborhood, inhabitants yelled and cried "Tatars are coming, our death is coming". This was a yell of desperation. Military horseback riders and Dvoryane and Boyar Children should have battle day and night. This was the lifestyle in the Kursk region for Dvoryane and Boyar Children in the 16th and 17th centuries. This was directly related to the military expeditions in which Sergey, grandson of Ivan Klyosov, and his great-grandson, Kondrat, participated.

Chigirin, Azov, and Kazikermen Expeditions

What were the expeditions in the division of Gregory Romodanovsky in 1674-1675, and what was the Kazikermen expedition in the regiment of Mamonov in 1694-1695?

These expeditions were part of the military operations undertaken by Tsar Alexei Mikhailovich before the end of his rule, to secure the right-hand side of the river Dniepr, which was a part of Russia. The left-hand side of the Dniepr, also the Russia side, was united with Russia twenty years back, in 1654. The opposite side was still under the heavy influence of the Turks. *Getman* (the title – AK) Samoylovich was the highest leader in Ukraine in 1674. He ruled on the both sides of the Dniepr, and was recognized by Russia. Doroshenko, the former Getman in the Western Ukraine, with a residence in Chigirin on the other side of the river, did not recognize Samoylovich and was maneuvering between the Kossacks and Moscow. At the same time, he was playing a game with the Turk Sultan, and, as it was reported at that time while negotiating with Russia, sent a request for military help from Crimea by the Turks. Samoylovich, alarmed, called for help from Moscow.

Count Gregory G. Romodanovsky with his troops had moved to Chigirin and seized the city. Doroshenko, understanding the hopelessness of the situation and having no news from the Turks, surrendered Chigirin and forfeited his getmanship. This was the first, but far from the last, joint military expedition of the Russian troops and Ukrainian Kossacks to Chigirin. In fact, this was the beginning of a series of Russian-Turkish wars, that is, wars of the Osman Empire and alignment to the Crimean Khanship. Sergey Klyosov participated in this, the first Russian-Turkish war.

Sergey's participation in the war was taken over by his son, Kondrat. The Kazikermen military expedition, which he had mentioned in "A Sorting Out Book of Boyar Children in the City of Kursk and the Kursk Region, 1695" (The State Archive of Kursk Region, 1555-1-168-419), was a part of the advancement to Azov by Tsar Peter, who later became Peter the Great. This was called The First Azov Expedition of 1695. A Cannon Regiment was headed by *Bombardir Piter*, as he called himself. The entire summer of 1695, Tsar Peter with his 35,000-strong troop had tried to capture the Turkish fortress Azov by throwing into it thousands of bombs but was unsuccessful and had to pull away by fall. In other words, his rule had started with an unsuccessful military operation.

However, it was not completely unsuccessful. At the end of July 1695, a regiment of Boyar Children of Kursk, among which was Reitar Kondrat, son of Sergey Klyosov, and headed by Kursk *Voevoda* (a military commander - AK), Ilya

Dmitriev-Mamonov, had stormed and took over the Turkish fortress Kazikermen at the mouth of the Dniepr river. The storming the fortress involved Kossack troops headed by Getman Ivan Mazepa and Moscow troops on horseback and on foot headed by Boyar B.P. Sheremetev. The siege and bombarding of the fortress had lasted five days and nights by use of cannons and small firearms. A Kossack historian of that time, Samyila Velichka, wrote: "One night, Kazikermen was hit by dozens of bombs and small grenades with fire coming from all carriages. All around the dark night was illuminated as if from lightning". Finally the fortress wall was destroyed and the troops stormed in. Hand-to-hand combat lasted five hours, and Kazikermen capitulated. Garrisons of the two neighboring Turkish fortresses, Aslam-kermen and Mouberek-kermen, fled to Crimea, abandoning large cannons. This was probably the only good news in the whole Azov campaign. Kondrat Klyosov was probably proud when he reported back home about the expedition.

Single-courters

By the beginning of 18th century, frontiers of Russia went from South to further West, and life of the military elite in the Kursk region became less tense. Dvoryane and Boyar Children were allowed to retire and permanently settle on their own land. If earlier their land was given to them in a temporary ownership only during the military service, then later, step by step, the restrictions were removed, and ownership rights began to move towards their direct descendants, then to wives and children, then to widows, and finally to side relatives. When sons in the dynasty military family were "*ripe to serve*", they either "let in" to the father's land and would not receive any extra land, or were granted their own land "*to be derived*" from their father.

Eventually, when life of the dynasty military people in the Kursk region had became more stable, it turned out that many Dvoryane and Boyar Children did not want to carry their saber on their hip all their life and be called to the troops until they are disabled or dead, and they began to move to their land to become farmers. This was the beginning of the forming of a separate strata of the Russian society of civil land-owners and their children, so-called single-courters. According to records of those times, in the 18th century, many Dvoryane who did not want to continue their military service moved to single-courts. In the 19th century, they received a right to become Dvoryane again after they were able to prove their noble roots and enrolled back into military service. As an historian wrote: "Some very old Dvoryane lines, reduced to own relatively little land, became single-courters. During the times of Peter the Great, some Dvoryane who owned as many as 100 and 200 peasant courts, did not want to serve and legally registered as single-courters".

In many cases, however, that transition was a forced one. Reforms by Peter the Great, who had changed Russian society, led to a significant reduction of the number of military dynasties, primarily City Dvoryane and Boyar Children. They faced few options: either to struggle to stay in the military service, or move to administrative work, that is, to become a kind of bureaucrat in the new society, or to settle as farmers on their own land. Many preferred the latter choice.

After some time, a number of the Tsar's verdicts created a new strata of society. Single-courters had taken a new level between Dvoryane and peasants, and became middle-class farmers. Many documents of that time called them free farmers, later state farmers, unlike possessed farmers, monastery farmers, or church farmers. State farmers in the Kursk region were not so much field workers; typically, they kept themselves busy with gardening, fishing, beehive-keeping, poultry-keeping, sheep-breeding and horse breeding.

Most Boyar Children did not have their own peasants or "possessed farmers". After they became single-courters, they labored themselves sometimes hiring workers. However, when they had served in the military, their land was tax-exempt, because they paid for the land with their service. Now their land became state taxed.

The legal unit for taxation was the single-court. Single-courts were further divided into smaller single-courts in which children and other close relatives lived. Single-court families often were quite large, with grandfather and grandmother, their children with wives and children, sometimes children of the third generation. Everyone in the court was called a single-courter. Only the court was taxed as a unit. Census documents, so-called Revision Tales, which were collected once in a while, have recorded single-courters by courts. There were ten Revision Tales between 1720 and 1858, the number does not include earlier revisions, so-called landrat censuses.

An example of such a large court was the first court in the Klyosov village, according to the First Revision of the Kursk district of 1710 (GAKO 184-4-12-177), which was registered as "Single-courters, Reitars". That court would nowadays be called a cluster. The Revision lists Afanasyi, son of Kondrat Klyosov (grandson of Boyar Son Sergey), 35 years of age; his wife, a 7-year old son and 1-year old daughter, two nephews of 16 and 18 years of age, a wife of the older cousin and their 6 months old daughter. Eight single-courters in one court. The second court was assigned to the second grandson of Sergey Klyosov, 50-year-old Eremei, son of Vassiliy Klyosov with his family; the third court was registered to Eremei's brother, a 40-year old Ivan with us family; the fourth court was their brother 30-year old Gregory, with his family; the fifth court was the

widow of their brother, Matvei with his family, and so on. There were 12 courts total with 35 "single-court *souls*".

The Second Revision in 1744 has shown that there were 70 people who lived in the Klyosov village that time: 69 single-courters and one hired worker. At the beginning of the 20th century, an incomplete family tree of Klyosovs included 176 individuals, and now, including us – my wife and two grandchildren makes 16 generations. A total number of descendants of Ivan Klyosov from the 16th century until now would be hard to count. There should be many hundreds of them. However, World War I, the Russian revolution of 1917, and the following Civil War of 1918-1922, as well as the World War II, have greatly reduced their numbers both directly and indirectly.

The Klyosov courts were not situated too close to each other. The land was not crowded. Three hundred acres is a pretty wide territory, which is the equivalent of 290 football fields. Courts were situated on two river banks on different sides of the land. It was a matter of convenience of Klyosovs. Those who had only 20 acres, almost 20 football fields, had enough land for their family. With time, of course, relatives multiplied and courts divided. Children commonly divide the father's land fairly, in equal parts, unless children died or moved elsewhere. Single-court villages were closed to outsiders, who were seldom allowed to get the land. Rarely son-in-laws were such outsiders. They were rare because daughters commonly left to go to their husband's village, also typically a single-court bearing the husband's name, or – much later - a court in the same village. Sons-in-law rarely came to their bride's court, because a sister of brothers would not get land. She was supposed to go to her husband. Exceptions were when the daughter was the only heir. And indeed, "*our land was inherited, a deserved inheritance, earned by tough military service, granted and registered by our ancestor. A son-in-law did not belong there*".

The Klyosov family records are filled with examples of departures of daughters to single-court villages:

From the 1763 Revision Tale:

-- Paramon, son of Afanasiy (yob 1713), wife Ekaterina, brought from Kursk district, village Anakhin, daughter of a single-courter Anakhin.

-- Matvei, son of Paramon (yob 1740), wife Akilina, brought from village Yakshino, daughter of a single-courter Yakshin.

From the 1782 Revision Tale:

-- Gregory, son of Paramon (yob 1748), wife Ustinia, brought from Fatezh district, village Shuklina, daughter of a single-courter Anisim Shuklin.

-- Vassilisa, Daughter of Paramon Afanasievich (yob 1761), given out to marriage to Akim Perkov, a single-courter from village Perkovo of Kuresk district.

-- Avdei, son of Yakov (yob 1740), wife Mavra, brought from L'gov district, village Polyachkovo, daughter of a single-courter Ivan Polyachkov.

-- Evpat, son of Lukyan (yob 1734), wife Ekaterina, brought from Kursk district, village Umrikhin, daughter of a single-courter Stepan Umrikhin.

And so forth.

Sometimes funny things had happened. Here is a record in a church book, section "Church Marriages". Groom – Timofey Klyosov from village Klyosov, Christian, 19 years of age. Bride – Elena Klyosov from Village Klyosov, Christian, 17 years of age. Witness from the groom side – Sergey Klyosov from village Klyosov and Kondrat Klyosov from village Klyosov; witnesses from the bride's side - Timofey Klyosov from village Klyosov and Sidor Yakovlev from village Yakovlev.

Another record: Groom – Ivan Klyosov from village Klyosov, Christian, 26 years of age. Bride – Tatyana Klyosov from Village Klyosov, Christian, 19 years of age. Witness from the groom side – Vlasiy Klyosov from village Klyosov and Kondrat Klyosov from village Klyosov; witnesses from the bride's side - Stefan Klyosov from village Klyosov and Stefan Pykhtin from village Pykhtin.

There were quite a few cases like these. However, it was already the 19th century and relatives spread along different and distant family lines.

Land granted to the ownership by the outlanding of dynasty military people, was called quarterland, from the Russian word translated as "quarter" (*chetvert'*), a land measuring unit equal to an acre. Hence, a right to own this land was called a quarter-right. A son-in-law, who had married an heir, and hence, entered a quarter-community, would not acquire a quarter-right legally, and it was not recorded in the land registry. There was no "possessed" peasants among them, of course, because a free single-courter female would not marry a slave. That is why grooms and brides were only single-courters from the same circle of the society. Historians, who had lived during those times described that single-courters were known for their vanity with their heritage and stiff upper lips to lower classes. And then "*many of them were Dvoryane of a former high society*", but became farmers for their unwillingness to serve in the Army.

And now we come to taxation "*from a soul*", or *soul-taxation*, and to the fate of the single-courter strata in Russia. This fate had resulted from a number of particular historical processes in Russia, namely disappearance of a noticeable part of the Dvoryane by moving it to agriculture, the displacement of an old, conservative lifestyle in rural Russia by a more mobile, aggressive, *soulless lifestyle.* The driving force of it was the role of a merciless state "assembly line" in grinding the conservative Russian lifestyle, which had eventually led to the fall of the Russian State.

This process can also be described in other terms. All of this evolutionary change had reflected an actual process of industrialization of Russia, the development of its modern agricultural system, which was initiated by Peter the Great. As a result, old Russia's backbone had collapsed in 1917.

The process had begun to accelerate from the time of Peter the Great's re-organization of the Russian Army. Former military check-ins "*on horse and armed*", care for "*purity of the military structure*" and "*purity*" of the military elite above all along with care of the dynasty military men themselves by their granting of land, was replaced with unified troops and maintaining unification. A kind of military "assembly line" structure had started to form, in which everything was paid, not by the military officer as before but from the state coffers. The state paid for horses, armaments, uniforms, food, and everything else. The country was divided by those who served and those who supported the army through taxes. Russia, with its conservative lifestyle, was not ready for it, and this led to a remarkable stress for the Russians. For such a sharp turn towards the new political and military system, as well as the whole way of life, Tsar Peter had pushed with his reforms.

Among the slogans of Tsar Peter was "nobody owns land in Russia for free". This clearly meant taxation of privately owned land. Therefore, single-courters, as a strata of the Russian society, was an invention of Peter the Great. Indeed, single-courters were initially mainly Dvoryane and Boyar Children, who served according to their heritage and from their land, and who then left the service and settled on their private, granted land. However, as a result of the "industrialization" of Russia for the following 100-150 years, that strata was mercilessly homogenized and stripped its former rights. For starters, this strata was mixed with representatives of a lower class of the society who were drafted for military service, and who received much smaller pieces of land, typically no more than 10-25 acres. Then with a series of Ukazs, or Tsar verdicts, that strata was transferred from the quarter-right to a "soul-right" in which taxation was from "souls", not from the land. Finally, single-courters were officially named not private farmers, as they used to be, but as "*state farmers*".

Of course, nobody in the Russian Government had such a far-fetched plan. It was just happening by itself and pushed forward with each new Ukaz as a part of a general process of pushing Russia forward to an industrialized society.

Flipping through Revision Tales of single-courters from the first one in 1710 (it was a preliminary one, and finally set up in 1724), to the tenth, in 1858, one can follow milestones in destroying of the former conservative, patriarchal lifestyle of Boyar Children settled onto their land granted for military service.

In the first Revision, Klyosovs in all 12 courts were described as "Single-courters, Reitars" or "Single-courters, City" (GAKO 184-4-12-177-179). "City" meant that they retired from Kursk military service. In 1711, an Ukaz by Tsar Peter assigned single-courters to a separate strata of the society which had legal right to discuss matters between themselves through representatives. The next Ukaz by Tsar Peter enacted on May 14, 1723, assigned to single-courters all former military people who have land, including City Dvoryane. At the same time, taxation from single-courters was significantly increased.

In 1724 the next Ukaz enacted by Peter stated that all single-courters were "fixed" on their land which they could not sell or leave. Furthermore, a new obligation was introduced, a rather severe one, that was called "a collective responsibility". In other words, if a single-courter left his land and the community, all other villagers had to pay the taxes for those who had left. State coffers needed money.

Needless to say, such measures did not encourage single-courters. They saw a clear curbing of their rights. Overall, Tsar Peter's reforms treated landowners quite heavy-handedly. In 1724, all single-courters who were not returned to the state service, military or otherwise, were officially named "state farmers" which in the Russian language also meant "state peasants". "Farmers" was never a respected term in Russia, then and now. It was a proverb in Russia during those times that "State peasants live as God says, possessed peasants live their master says", however, it did not help much. This was a way of converting retired Dvoryane and Boyar Children to farmers and that transfer was significant.

After Peter the Great, the Russian Tsars were Elizaveta, Ekaterina, and in 1796, after Ekaterina's death, her son Pavel became the next Tsar of Russia. He had initiated a change from the quarter-tax to the soul-tax. In other words, taxation from a court as an administrative unit was removed and replaced with head-taxation, a taxation from every individual inhabitant of courts and villages in general. This was a painful process, mainly for those who had had large land inherited from their ancestors 100-200 years earlier. The reason was that the soul-taxation was introduced along with re-distribution of land between the haves and have-nots, and to allocate to every farmer 10 acres of land per soul, or 40

acres to every court. It was accomplished mainly by taking land from large land-owners. It turned out that there was not so much land available. Then it was decided to still give 10 acres per soul, and where there was not that much land available, to give 5 acres per "revision soul". This was very beneficial to small land owners but not to large land-owners.

That decision had largely divided society and stressed it. Pavel was Tsar for only five years, however, his "soul-taxation" reform had been implemented for about 100 years and led to many fights, including physical ones between the have and have-not farmers and peasants.

According the last 10th Revision in 1858, village Klyosov had 14 courts, 71 male single-courters and 73 female single-courters (GAKO 184-2-1112-626/639). It seemed that nothing had changed in village Klyosov for the last 150 years except the number of courts and their inhabitants, which created a kind of a still picture. Such a stillness was very typical for large land single-courters. This stillness was their happiness and was at the core of a patriarchal lifestyle. There was one more feature in their life clearly distinguishing large land single-courters from other peasants. Single-courters knew their ancestors and predecessors. They knew their relatives, past and present, even quite distant ones and how they were related to each other. An old historian wrote: "These memories with their exact precision are carried by only Boyar Children who never forget their nobility and past service of their great-grandfathers; there are no tales in modern villages composed with recent visitors". This was considered as "vanity" of single-courters by "recent visitors" not having such memories.

Eventually, the Russian Government took over all single-court villages move to a soul-taxation. The patriarchal lifestyle had gone, and the villages had started to collapse. My direct predecessors had moved as an extended family to Siberia in 1898, and built their life back again. And then came the 1917 Revolution "of workers and peasants". My great-grandfather, Ermolay Klyosov, a priest, was shot by a firing squad following a verdict by a "revolutionary tribunal". During those times, many Klyosovs were accused and sentenced for being rather "haves" than "have not's". This was their principal guilt, and this was a fact of life then. Here is a far from complete list of Klyosovs sentenced and exiled by the Soviet Government and its representatives:

-- Klyosov Fedor, 1872 yob, born Tomsk, Siberia, Russia. Was arrested: September 28, 1937. Sentenced December 9, 1939 by "troika" (meaning – his case was considered by three representatives, they were the court, jury, and judge at the same time and place) of NKVD (the predecessor of KGB). The case was described as "an anti-soviet activity". Shot on January 22, 1938. The case was reversed and nullified in May, 1989.

-- Klyosov Ivan, 1905 y.b., born Kursk Region, Russia. A head of a local farm. Was arrested: March 3, 1937. Sentence: imprisonment for 3 years. The case was reversed and nullified in September, 1989.

-- Klyosov Petr, 1896 yob, born Stavropol Region, Russia. Was arrested: March 10, 1932. Shot on January, 1933.

-- Klyosov Georgy, 1881 yob, born Stavropol Region, Russia. Was arrested: October 30, 1930. Sentenced May 22, 1931 for 10 years an exile in the Siberian North.

-- Klyosov Longvin, 1883 yob, born Stavropol Region, Russia. Was arrested: October 20, 1930. Sentenced May 22, 1931 for 10 years an exile in the Siberian North.

-- Klyosov Prokop, 1875 yob, born Omsk, Siberia, Russia. Was sentenced on August 8, 1931. Sentenced August 8, 1931 for exile in the Siberian North, died there in 1933.

-- Klyosov Ilya, 1870 yob, born Ishim District, Russia. Was arrested: February 17, 1938. Sentenced March 4, 1938 by NKVD "troika". Shot on March 12, 1938. The case was reversed and nullified in January, 1959.

My father, Alexey Klyosov, was born in 1923 in Siberia and graduated from Air Force school in 1941. He was drafted right away to the front to fight with Nazi Germany. The war was ended for him in April 1945 with the defeat of Konigsberg in East Prussia, Germany, in which he was taking an active part and was awarded for valor. He had served with the military transportation troops in Insterburg (now Chernyakhovsk) in East Prussia which was later annexed by the Soviet Union and became the Kaliningrad Region, USSR, and now Russia. Then he served and our family lived in Riga, Latvia, and we all moved in 1955 to the largest Soviet missile military base, which then became a space complex Kapustin Yar. Being mortally ill there he was directed to serve at a famous Russian resort called Sochi in a semi-tropical area of Russia on Black Sea as an attempt to save his life. It gave him indeed a few more years of life. His son and my brother, Eugene, had continued family traditions and served for more than 20 years in Siberia with the military transportation troops, and then in Sochi. I live in Boston with my wife Gail, for the last 20+ years. Our daughter, Svetlana with her family and our two grandsons, live in France, in a *village* between Nice and Cannes. They are all naturalized French nationals.

This is kind of normal biography, and a mirror of Russian history with a transition to modern times.

My haplogroup is R1a1, haplotype 13 24 16 11 11 15 12 12 10 13 11 30 – 16 9 10 11 11 24 14 20 34 15 15 16 16 – 11 11 19 23 15 16 17 21 36 41 12 11 – 11 9 17 17 8 11 10 8 10 10 12 22 22 15 10 12 12 13 8 15 23 21 12 13 11 13 11 11 12 13

All of us Klyosovs must have about the same haplotype, with only few mutations, as well as most of inhabitants of Russian cities, towns and villages, having the same haplogroup R1a1. It is a typical Slavic haplotype and is derived from our ancestors who had lived about five thousand years ago. But that is a different story.

Acknowledgements: The author is indebted to a professional genealogist, a chairman of Kursk Family History Society Eugene S. Karpuk for a truly professional work in archives, and to expert genealogists Vladimir N. Orlov and Vladimir G. Volkov (both from Russia) for a great help and valuable comments and suggestions. I greatly appreciate a valuable help of William P. Kisseloff-McCulloch (California) in preparation of this paper.

Climatic Correlations

The report presented at the XIX session of the seminar "Earth system".
Geological Faculty of Moscow State University – February 1, 2011

Valery P. Yurkovets
valery.yurkovets@gmail.com

A scheme of climatic events and their correlations with the ages of paleosols, archaeological cultures and paleogeographic data of the Caspian Sea - Sea of Azov - Black Sea - Mediterranean Sea in the upper Pleistocene and Holocene is proposed. The scheme and correlations are given in Tables 1 - 3. The totality of these comparisons show that the climate on our planet is subject to periodic changes, which are based on two space factors - the precession of the Earth rotation axis and the superposition of the Earth, Moon and Sun orbits (Yurkovets, 2010). The first factor determines the change of glacials and interglacials, and it has a period of approximately 26 thousand years. In the Table 1 it is presented in the form of large sinusoid. The second factor has a two thousand year period. It is presented in Table 1 in a form of a sawtooth curve. It determines the change of a cooling-warming stages during the glacials and interglacials.

The first curve shows that the glacials take place in the hemisphere, which is opposite to the Sun at aphelion (due to inclination of the Earth rotation axis). As a result, winters in this hemisphere are the most severe, that is the cause of glaciations. Over the full period of precession glacials on Earth occurred twice - first in the northern hemisphere, the second in the south hemisphere.

The mechanism of small cycles (superposition of the Earth, Moon and Sun orbits) was discovered by Petterson at the beginning of the 1900s, and then confirmed by archaeoclimatic reconstructions of Shnitnikov, Matyushin, Shilik, and others (Yakushev, 2008). Large and small cycles determine changes in the global sea level and inland water bodies, including the Caspian Sea. A comparison of these cycles to the chronology of archaeological cultures from the Upper Paleolithic is described in Table 2. As can be seen, the archaeological events are determined indeed by the climatic cycles.

The proposed scheme fits well to a paleogeographic reconstruction of the Caspian and Black Seas by Janina (Janina, 2009), Gorlov (Gorlov et al., 2004), Badyukova (Badyukova, 2006) and others, see Table 3.

These considerations allowed us to refine both the time and the cause of Manych Strait's formation, on which the dammed waters of Eurasia discharged to the Black sea during the glacials (Grosman, 1989), as well as the penetration of the

Black Sea waters in the Caspian Sea during the warm Black Sea transgressions. The first stage was accompanied by erosion which reached the marks -5 meters (Badyukova, 2006). The second one, as a small circle, followed immediately after the first one, when Manych was not blocked by silt, as shown in Table 1. This also explains the two observed levels of Khvalyn transgression, which were erosional and accumulative. The first one had occurred in the maximum of the glacial, when the base level lowered to the level of the Pontus regression. The second one was during the maximum warming, which followed by immediately, that is when decreasing Khvalyn transgression's water met in the Black Sea which has increased the Black Sea transgressions. Their levels were equalized, gradually reducing the base level to zero values.

Based on data of Table 1, the simultaneous climatic events were:

1. Wurm III – Upper Valday Glacial in Eastern Europe - Sartanian Glaciation in Siberia - Wisconsin III Glacial in North America, during the period of 10 to 17 thousand years ago;

2. Wurm II / III Interglacial - Bryansk Interglacial in eastern Europe - Kargin Interglacial in Siberia - Wisconsin II / III Interglacial in North America, during the period of 17 to 36 thousand years ago;

3. Wurm II - Leyastsiem cold snap in Eastern Europe - Konoschel cold snap in Siberia - Wisconsin II Glacial, during the period of 36 to 43 thousand years ago;

4. Wurm I / II Interglacial - Kostenki (working title) Interglacial in Eastern Europe and Siberia - Wisconsin I / II Interglacial in North America, during the period of 43 to 62 thousand years ago;

5. Wurm I – Lower Valday Glacial in Eastern Europe - Zyryanka Glacial in Siberia - Wisconsin I Glacial in North America, during the period of 62 - 69 thousand years ago;

6. Riss-Wiirm Interglacial - Mikulin Interglacial in Eastern Europe - Sangamonian Interglacial in North America.

DNA genealogical aspects. The climatic history of the Upper Pleistocene and Holocene was recorded not only in the Quaternary geological chronicle and as the material of archaeological cultures, as shown in Tables 1-3, but also indirectly, through the history of haplogroups of the DNA. The right side from the precession curve in Table 1 shows the main events of the Y-chromosome phylogeny according to Karafet (Karafet et al, 2008) and Klyosov (Klyosov, 2009). This comparison suggests a link between the main events in phylogeny of the DNA and the global climate changes. The reason for this is that any significant change in climate forced people to migrate, and this led to a formation of new

branches on the phylogenetic tree of the Y-chromosome. Those coordinated phylogenetic and climatic events sometimes can be distinctively noted. Thus, the warming subatlantik not only led to the collapse of the great empires of the Iron Age in Eurasia (the "Great Migration of Peoples"), but had resulted in a "population bottleneck" of the people in the east part of the continent. As it turned out (Rozhanskii, 2010), at least 70% of the Mongols and Kirghiz, as well as a large percentage of the Kazakhs happen to be the direct descendants along the male lines possibly originated by only five individuals (common ancestors) who lived there at those times. This is remarkably correlated with the historical events which led to the unification of China, and then to the emergence of the empire of Genghis Khan (Yurkovets, 2011).

Table 1. Climatic Correlations

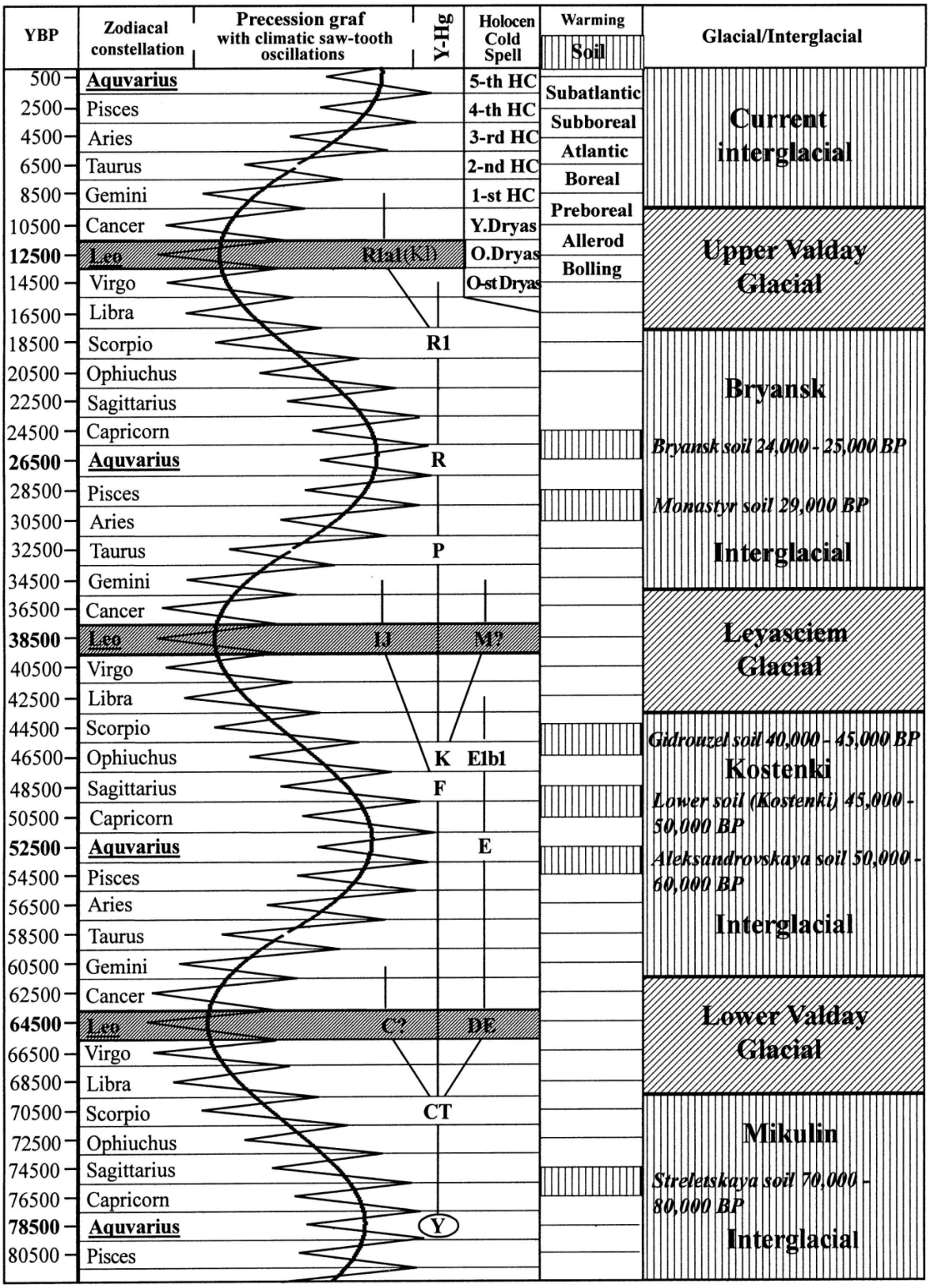


Table 2. Climatic and Archeological Correlations

t of CE max.	Millenium	Period	Climatic Event (CE)	Ecological and Historical outcome
1	2	3	4	5
+500	3	Current Interglacial (Holocen)	Current warming	**"Global warming"**. Rising sea levels by 25 meters, the desertification of the steppes, stepping of the forest-steppes. Melting Arctic ice field.
-500	2		Holocen Cold Spell 5	**The flowering of the humid zone**. Settling a man of huge spaces that have become suitable for animal husbandry and agriculture. As a result - a global population explosion.
-1500	1		Subatlantic Warming	**The collapse of the great empires** as a result of the great droughts. Global phenomenon known as the **Migration Period (Barbarian Invasions).**
-2500	1 BC		Holocen Cold Spell 4	**The flowering of the Iron Age in Europe**. Time of the great empires of Eurasia - the Persian, Roman, Mauryan in Hindustan, China Qin. Maya civilization in Mesoamerica.
-3500	2 BC		Subboreal Warmin	The **Bronze Age collapse.** Decay of the Bronze Age cultures in the humid zone of desertification steppes. **Seima-Turbino** phenomenon. Stonehenge abandoned. The collapse of ancient Egypt.
-4500	3 BC		Holocen Cold Spell 3	**The flowering of the Bronze Age**. The spread of **Corded Ware** culture, the spread of **Bell-Beaker** culture, the heyday of the **Yamna** culture. Stonehenge 1-2. The beginning of the civilization of Ancient Egypt.
-5500	4 BC		Atlantic Warming	**"Chalcolithic collaps"** – decay of the **Sredny Stog** culture and the **Cucuteni-Trypillian** culture. The decay of **Samara** culture, Agidel, Surtandin cultures. End of **Ubaid** culture. In fact - as the **"Migration Period"**.
-6500	5 BC		Holocen Cold Spell 2	**Flowering Sredny Stog culture, Cucuteni-Trypillian culture** and a **Vinèa culture, the Danubian cultures**. Spreading of animal husbandry hroughout Eurasia, the beginning of the northern **Comb Ceramic** culture.
-7500	6 BC		Boreal Warming	**"Creation" of the modern world**: the end of melting glaciers, the rise of the oceans up to date, the breakthrough the Bosphorus, the formation of the Baltic Sea, the beginning of the Sahara desert. Peak of **Lepenski Vir, Starèevo** culture.
-8500	7 BC		Holocen Cold Spell 1	**The Neolithic Revolution** - the transition from hunting and gathering to agriculture and settlement. Goats, sheep, cow, pig were domesticated, horse were domesticated on the southern Urals (Mullino II). **Lepenski Vir.**
-9500	8 BC		Preboreal Warming	**The beginning of melting ice sheets**. End of periglacial – **«sandur»** - type of economy. End of **Magdalenian** as an ecological phenomenon.
-10500	9 BC	Upper Valday Glacial	Young Dryas Cold Spell	**Last phase of glaciation**. The last period of the Volga-Oka interfluves's **Ressetinskaya** culture, dominant in the central regions of the Russian Plain entire period of the Late glacial.
-11500	10 BC		Allerod Warming	**Glaciers are retreating. Magdalenian 6. Ressetinskaya** culture. Flourishing cultures of the Baltic **Magdalenian - Ahrensburg** culture, the **Federmesser** culture. **Swiderian** culture. End of **Natufian** Levant culture.
-12500	11 BC		Old Dryas Cold Spell	**Maximum glaciation. Magdalenian 5**. In Europe 11-13 thousand years BC beings only Magdalenian cultures - **Ressetinskaya** culture, **Swiderian** culture to the Centre Europe and itself **Magdalenian** in the West Europe.
-13500	12 BC		Bolling Warming	**Glaciers are retreating. Madeleine 4, Ressetinskaya** culture. End of the Hamburg culture. Start **Swiderian** culture.
-14500	13 BC		Oldest Dryas Cold Spell	**Glaciation. Madeleine 3, Ressetinskaya** culture (inherits of East Gravette), **Hamburg** ñulture (is similar to Ressetinskaya).
-15500	14 BC		Warming	**Glaciers are retreating. Madeleine 2. Ressetinskaya** culture. Start the **Hamburg** culture (culture is wholly owned by the Oldest Dryas «from heat to heat»: 13500 - 11100 BC).
-16500	15 BC		Cold Spell	**Start Upper Valday glaciation**. The beginning of the third wave of **megafauna** death. **Upper Solyutre. Madeleine 1. Ressetinskaya** culture (?).
-17500	16 BC	Bryansk Interglacial	Warming	**End of Bryansk interglacial. Badegulskaya** culture. Intermediate **Solyutre**. End of **East Gravette: Kostenki-Avdeyevo** culture (inherited by **Ressetinskaya** culture).
-18500	17 BC		Cold Spell	**Climatic Optimum. Lower Solyutre. East Gravette: Kostenki-Avdeyevo** culture.
-19500	18 BC		Warming	**Climatic crisis. Protosolyutre. East Gravette: Kostenki-Avdeyevo** culture.
-20500	19 BC		Cold Spell	**Climatic Optimum. East Gravette: Kostenki-Avdeyevo** culture.
-21500	20 BC		Warming	**Climatic crisis**. End of **Gravette, East Gravette: Kostenki-Avdeyevo** culture.

Table 2. Climatic and Archeological Correlations (cont.)

1	2	3	4	5
-22500	21 BC	Bryansk Interglacial	Cold Spell	**Climatic Optimum. Gravette, East Gravette - Kostenki-Villendorf** community of cultures, 30,000-20,000 BC (including **Kostenki-Avdeyevo** culture, 26,000-16,000 [Zaraysk] BC).
-23500	22 BC		Warming	**Climatic crisis. Gravette, East Gravette - Kostenki-Villendorf** community of cultures, 30,000-20,000 BC (including **Kostenki-Avdeyevo** culture, 26,000-16,000 [Zaraysk] BC).
-24500	23 BC		Cold Spell	**Climatic Optimum. Gravette, East Gravette - Kostenki-Villendorf** community of cultures, 30,000-20,000 BC (including **Kostenki-Avdeyevo** culture, 26,000-16,000 [Zaraysk] BC).
-25500	24 BC		Warming	**Max. warming. «The Flood»**. End of **Aurignacian** culture. Blossoming **megafauna** in the North. **Gravette, East Gravette - Kostenki-Villendorf** community of cultures, 30,000-16,000 BC (incl. **Kostenki-Avdeyevo** culture).
-26500	25 BC		Cold Spell	**Climatic Optimum. Aurignacian** culture, **Gravette, East Gravette - Kostenki-Willendorf** community of cultures, 30,000 – 20,000 BC(including **Kostenki-Avdeyevo** culture, 26,000-16,000 BC).
-27500	26 BC		Warming	**Climatic crisis. Aurignacian** culture. **Gravette, East Gravette - Kostenki-Villendorf** community of cultures, 30,000 – 20,000 BC (including **Kostenki-Avdeyevo** culture 26,000 – 16,000 BC).
-28500	27 BC		Cold Spell	**Climatic Optimum. Aurignacian** culture. **East Gravette - Kostenki-Villendorf** community of cultures (30,000 – 20,000 BC).
-29500	28 BC		Warming	**Climatic Crisis, analogue of Subboreal**. End of **Selet and Streletskaya** cultures. **Aurignacian** culture, **East Gravette - Kostenki-Villendorf** community of cultures (30,000 – 20,000 BC).
-30500	29 BC		Cold Spell	**Climatic Optimum. Aurignacian** culture. **Streletskaya** culture, **Selet, East Gravette - Kostenki-Villendorf** community of cultures (30,000 – 20,000 BC).
-31500	30 BC		Warming	**Climatic Crisis, analogue of Atlantic**. Start of **Aurignacian** culture. **Streletskaya** culture **Selet, East Gravette - Kostenki-Villendorf** community of cultures (30,000 – 20,000 BC).
-32500	31 BC		Cold Spell	**Climatic Optimum. Streletskaya** culture. **Selet.**
-33500	32 BC		Warming	**Climatic crisis. Analogue of Preboreal. Streletskaya** culture. **Selet.**
-34500	33 BC		Cold Spell	**Climatic Optimum. Streletskaya** culture. **Selet.**
-35500	34 BC		Warming	**End of Leyastsiem glaciation. Analog of the Boreal. Streletskaya** culture. **Selet**.
-36500	35 BC	Leyasciem Glacial	Cold Spell	**Glaciation. Analogue of Young Dryas. Streletskaya** culture. **Selet.**
-37500	36 BC		Warming	**Glacier retreat. Analogue of Allered. Streletskaya** culture. **Selet.**
-38500	37 BC		Cold Spell	**Maximum glaciation. Analogue of Old Dryas. "Nuclear winter" of Paleolithic Age. Aurignacian of Kostenki** and its catastrophic end («layer of ashes"). **Streletskaya** culture. **Selet.**
-39500	38 BC		Warming	**Glaciers retreat. Analogue of Bolling. Spitsynskaya** culture («pre-**Aurignacian** »). Start **Streletskaya and Selet** cultures.
-40500	39 BC		Cold Spell	**Glaciation. Analogue of the Oldest Dryas. Kostenki 14 layer IVb** (preceded by «pre-**Aurignacian** »).
-41500	40 BC		Warming	**Glaciers are retreating. Kostenki 14 layer IVb** (preceded by «pre-**Aurignacian** »).
-42500	41 BC		Cold Spell	**The beginning of glaciation.** The beginning of the second wave of the **megafauna death**. End (43,000 **BP**) of the site **Kostenki 12/ layer V and Kostenki 12/ layer IV. Start Kostenki 14 / layer IVb.**
-43500	42 BC	Kostenki Interglacial 62 - 43 BP	Warming	**End of Kostenki interglacial. Kostenki 12/ layer V and Kostenki 12/ layer IV** (50,000 - 43,000 **BP**).
-44500	43 BC		Cold Spell	**Climatic Optimum. Kostenki 12/ layer V and Kostenki 12/ layer IV** (50,000 – 43,000 **BP**).
-45500	44 BC		Warming	**Climate crisis. Kostenki 12/ layer V and Kostenki 12/ layer IV** (50,000 – 43,000 **BP**).

Table 3. Climatic and Paleogeography Correlations

t of CE max.	Millenium	Period	Climatic Event (CE)	Paleogeography of the Caspian Sea		Paleogeography of the Black Sea	
				Matyushin	Yanina	Arhaeological data	Yanina
1	2	3	4	5	6	7	8
+500	3	Current Interglacial (Holocen)	Current warming		New Caspian Warm Transgression		Black Sea Warm Transgression
-500	2		Holocen Cold Spell 5			Korsun Regression	
-1500	1		Subatlantic Warming			Nimphea Transgression	
-2500	1 BC		Holocen Cold Spell 4			Fanagorian Regression	
-3500	2 BC		Subboreal Warmin			New Black Sea Transgression	
-4500	3 BC		Holocen Cold Spell 3	Turalin Transgression			
-5500	4 BC		Atlantic Warming	Makhachkala Regression			
-6500	5 BC		Holocen Cold Spell 2	Gousan Transgression			
-7500	6 BC		Boreal Warming	Gelaldin Regression			
-8500	7 BC		Holocen Cold Spell 1	Dagestan Transgression			
-9500	8 BC		Preboreal Warming	Mangyshlak Regression	Mangyshlak Regression		
-10500	9 BC	Upper Valday Glacial	Yung Dryas Cold Spell	Sargas Transgression	Upper Khvalyn Transgression		Novoevksinskaya Cold Regression
-11500	10 BC		Allerod Warming	Begdash Regression	Enotaevskaya Regression		
-12500	11 BC		Old Dryas Cold Spell	Khvalyn Transgression	Old Khvalyn Cold Transgression		
-13500	12 BC		Bolling Warming				
-14500	13 BC		Oldest Dryas Cold Spell				
-15500	14 BC		Warming				
-16500	15 BC		Cold Spell				
-17500	16 BC	Bryansk Interglacial	Warming		Akhtuba-Atel Regression of Bryansk Interglacial		Surozh Warm Transgression
-18500	17 BC		Cold Spell				
-19500	18 BC		Warming				
-20500	19 BC		Cold Spell				
-21500	20 BC		Warming				

Table 3. Climatic and Paleogeography Correlations (cont.)

1	2	3	4	5	6	7	8
-22500	21 BC	Bryansk Interglacial	Cold Spell		Akhtuba-Atel Regression of Bryansk Interglacial		Surozh Warm Transgression
-23500	22 BC		Warming				
-24500	23 BC		Cold Spell				
-25500	24 BC		Warming				
-26500	25 BC		Cold Spell				
-27500	26 BC		Warming				
-28500	27 BC		Cold Spell				
-29500	28 BC		Warming				
-30500	29 BC		Cold Spell				
-31500	30 BC		Warming				
-32500	31 BC		Cold Spell				
-33500	32 BC		Warming				
-34500	33 BC		Cold Spell				
-35500	34 BC		Warming				
-36500	35 BC	Leyasciem Glacial	Cold Spell		Akhtuba-Atel Regression of Old Valday Glacial Peak		Postkaragant Regression
-37500	36 BC		Warming				
-38500	37 BC		Cold Spell				
-39500	38 BC		Warming				
-40500	39 BC		Cold Spell				
-41500	40 BC		Warming				
-42500	41 BC		Cold Spell				
-43500	42 BC	Kostenki Interglacial 62 - 43 BP	Warming				
-44500	43 BC		Cold Spell				
-45500	44 BC		Warming				

References

Antonova, V.M., Khomenko, A.A. K voprosu o novochernomorskoi transgressii Azovo-Chernomorskogo basseina [On the Question of Novochernomorskaya Transgression of the Azov-Black Sea Basin]. Moscow State University. Proceedings of International Symposium on "Late Cenozoic geological history of the north of the arid zone." 2006, Azov, Rostov-on-Don. P.18-20. (In Russian).

Badyukova, E.N. Kogda poslednii raz soedinialis Chernoe i Kaspiiskoe moria? [When Connected the Black Sea and Caspian Sea Last Time?] Moscow State University. Proceedings of International Symposium on "Late Cenozoic Geological History of the North of the Arid Zone." 2006, Azov, Rostov-on-Don. P.21-24. (In Russian).

Badyukova, E.N. Odno iz dokazatelstv soedineniia Kaspiiskogo I Chernogo morei v konce pozdnekhvalynskogo vremeni [One of the Evidence Connecting the Caspian and Black Seas in the Late Upper Khvalyn Time]. / E.N. Badyukova / Geomorphology. - 2004. - N 2. - P. 23-31. (In Russian).

Badyukova, E.N. Vozrast khvalynskikh trsnsgressii Kaspiiskogo moria [Age of Khvalyn Transgressions of the Caspian Sea]. Oceanology - Volume 47, № 3, May-June 2007, P. 432-438. (In Russian).

Gorlov, U.V., Porotov, A.V. Stolyarova, E.V. K ocenke izmenenii urovnia Chornogo moria v antichnyi period po arkheologo-paleogeograficheskim dannym [To Assess Changes in the Level of the Black Sea in the Ancient Period on Archaeological and Paleogeographic Evidence]. / / Antiquities Bosporus .- № 7 .- 2004.- P.117-128. (In Russian).

Grosman M.G. Poslednee velikoe oledenenie territorii SSSR [The Last Great Glaciation of the USSR] "Znanie". Moscow, 1989. (In Russian).

Karafet, Tatiana M., et al, New Binary Polymorphisms Reshape and Increase Resolution of the Human Y Chromosomal Haplogroup Tree. www.genome.org on April 2, 2008.

Klyosov, A.A. Gaplotipy iuzhnykh i baltiiskikh russkikh slavian: chetvero plemen? [Haplotypes of the Southern Baltic and Russian Slavs: Four Tribes?] RA-DNA, 2009. Rozhanskiy, I. (2011) DNK-genealogiia i dokumentaknye rodoslovnye. Soiuz ili konflikt? [DNA Genealogy and Documented Pedigrees. Alliance or Conflict?] Bulletin of the Russian Academy of DNA Genealogy (ISSN 1942-7484), Volume 4, № 1, 4-19. (In Russian).

Fedorov, P.V. 1982 Posleledniкovaia transgressiia Chernogo moria i problema izmeneniia urovnia okeana za poslednie 15,000 let [The post-glacial transgression of the Black Sea and the problem of ocean level change during the last 15,000 years]. In Kaplin P.A., Klige R.K., Chepalyga A.L., eds, *Kolebaniia urovnia morei I okeanov za 15,000 let [Sea and Oceanic Level Fluctuations for 15,000 Years],* pp. 151-156. Nauka, Moscow. (In Russian).

Yurkovets, V.P. (2010) Klimaticheskie korreliacii [Climatic Correlations]. Bulletin of the Russian Academy of DNA Genealogy (ISSN 1942-7484), V.3, № 2, 301 - 325. (In Russian).

Yurkovets, V.P. (2011) Klimaticheskie korreliacii. Prodolzhenie [Climatic Correlations. Continued]. Bulletin of the Russian Academy of DNA Genealogy (ISSN 1942-7484), Volume 4, № 1, 66-80. (In Russian).

Yurkovets, V.P. (2011) … I snova otkrytie Kitaia [... And Again, the Opening of China]. Bulletin of the Russian Academy of DNA Genealogy (ISSN 1942-7484), Volume 4, № 3, 591-621. (In Russian).

Yakushev, D.I. Geoinformacionnoe modelirovanie prostranstvenno-vremennykh geofizicheskikh processov s poligarmonicheskoi strukturoi. Avtoreferat dissertacii na soiskanie uchenoi

stepeni doktora tekhnicheskikh nauk [Geoinformation Modeling of Space-Time Geophysical Processes with Polyharmonic Structure. Dissertation for the Degree of Doctor of Technical Sciences]. "LETI", St. Petersburg, 2008. (In Russian).

Yanina, T.A. Paleogeografiia basseinov Ponto-Kaspiia v Pleistocene po rezultatam malakofaunisticheskogo analiza. Avtoreferat na soiskanie uchenoi stepeni doktora geograficheskikh nauk [Paleogeography Ponto-Caspian Basin in the Pleistocene Based on Analysis of Molluscs. Dissertation for the Degree of Doctor of Geografical Sciences]. Moscow State University. Department of Geography. Moscow, 2009. (In Russian).

DISCUSSIONS
and correspondence on DNA genealogy, history, linguistics
(excerpts)

Anatole A. Klyosov
http://aklyosov.home.comcast.net

1. On the split of DNA-lineages of the Jews and the Arabs

Many were misled (and many still are) by a rather intense campaign in the literature that so-called J1-CMH appeared about 3300 year old.

It turned out to be a complete bogus. Since the first two papers in Nature in 1997-1998 this error was traveling from paper to paper. In fact, the authors (Hammer, Skorecki, Parfitt, et al) mixed up quite different populations, one relatively "young" (TMRCA ~ 1000 ybp), and one relatively "old" (TMRCA ~ 4200 ybp), and "produced" a phantom "common ancestor" with a phantom TMRCA in between. It was (and still is) a typical mistake by population geneticists, that is neglecting a possibility of various lineages in one dataset.

The so-called "J1 Cohen Modal haplotype") can be easily found in the present-day Arab haplotypes, which coalesce at about 9,000 years before present. There is nothing specifically "Cohanim" in it. Around 4200 ybp that lineage split into what later turned out to be the Jewish and the Arabic lineages (ref. see below).

The J1-"CMH" haplotype tree splits into two distinct parts, one, a tight cluster of CMH haplotypes, are predominantly "Cohens" and their descendants, with a common ancestor of 1070±170 years before present, and a loose, much older branch, which contains very few Jewish haplotypes, with a common ancestor of 4300±500 ybp. Apparently, those (mainly) Europeans and Arabs are descendants of the Arabs, for example, after the 7th century AD (Proceedings of the Russian Academy of DNA Genealogy, vol. 3, No. 4, April 2010, free download):

http://www.lulu.com/items/volume_68/8657000/8657872/1/print/8657872.pdf

Similar data were presented in [Klyosov, Human Genetics, v. 126, pp. 719-724 (2009)].

In the same manner, J2 Jewish and Arabic haplotypes split into two lineages, one exclusively Jewish, another exclusively Arabic, and the split occurred 4175±510 years bp. In other words, there were two "Abrahams", in J2 and J1 haplogroups. In R1b1b2 there was no such a split between future Jews and Arabs.

J2 Cohanim have a common ancestor at 3300±400 ybp (see ref. above).

The two lineages, it seems, lived together... well, always, in a sense. They live together now, as we know. Haplogroup J1 exists at least for 19,000 years

(see http://aklyosov.home.comcast.net
- Proceedings of the Russian Academy of DNA Genealogy, vol. 1, No. 1, 2008 (in English) (free download) (all Proceedings are in the above site at the end of the site).

and haplogroup J2 exists for at least 11500 years (Vol. 2, No. 3, 2009, "Iberian haplotypes and history of the Basques, Sephards and other populations of Spain and Portugal", pp. 390-421).

Before 4200 ybp both the J1 Jews and Arabs were, essentially, Bedouins. The origin of J2 Jews and Arabs is much less clear, either they lived there since time immemorial, or came from the Mediterranean, again, thousands years ago. Why their J1 and J2 populations would have lived separately? The "Cohen" CMH among the Arabs is traced as deep as 9,000 ybp.

More detailed subclade assignments in J2 and J1 Jews and Arabs are J2a and seemingly J1e (Proceedings of the Russian Academy of DNA Genealogy, Vol. 2, No. 9, 2009, pp. 1100-1116 and Vol. 3, No. 4, pp. 635-653 - in English).

There was also a rather detailed analysis of the Sharifs/Sayids (Vol. 2, No. 7, pp. 1182-1199).

We can only tell how many of descendants (fraction-wise) of those subclades live NOW. For example, Fig. 3 in Vol. 3, No. 4 (page 645) shows the J2 tree, which contains 131 of 67-marker haplotypes. Of them the "Abraham" split branch contains 50 haplotypes (21 Jews and 29 Arabs), that is 38% of all. Branches J2a4b, J2a4b1, J2b sit separately. They are also, of course, split from J2a, but, apparently, before 4000 years ago.

Some figures and calculations are also given in my paper ("Comments") in Human Genetics, v. 126, No. 5, pp. 719-724, 2009 (Klyosov A.A. A comment on

the paper: Extended Y chromosome haplotypes resolve multiple and unique lineages of the Jewish Priesthood).

The future Jews and future Arabs lived as a joint J1-J2 community (whatever "community" means here). Around 4200 years ago something happened between them, and it was VERY serious. Call it cultural, religious, or whatever, but they split very decisively, "cold turkey", and it was, as times showed, irreversible. Both J1 and J2 Jews and Arabs parted. This partition we see on the haplotype tree. This is it, facts stopped here.

Now, interpretations begin. One interpretation is that there was someone whom we now call "Abraham". Of course, we do not know his real name, but he was there. He was either father or a leader of both the (future) Jews and the (future) Arabs, as it is described in the Bible/Thora. He might have been J1 or J2, it does not matter. At his lifetime the (future) Jews and the (future) Arabs split. Both present day J1 and J2 consider him as their founder and honor him, both the Jews and the Arabs. In that sense, he belonged to both J1 and J2. Both haplogroups are equally Jewish (and equally Arabic).

Another interpretation is a similar one, however, "Abraham" was not a father as it is described in the Bible. He was just a wise leader, and he initiated that split between J1-J2 Jews and J1-J2 Arabs. Both Arabs and Jews honor him in their stories. The legend has made him "the father".

I prefer the first interpretation, since I prefer not to deviate from the Bible unless I know for sure that Abraham was not a "father". Since I do not know it, I prefer to stick to the Biblical version ("if it ain't broke, don't fix it").

2. On DNA genealogy and the Tenths (having DYS388=10)

An open letter from Bill Tucker to a community of the Tenths

First, some background information should be related. My involvement with Anatole Klyosov, Ph.D., came about through a backdoor entry. Administrator Lee Tucker received an email from Charles Purkins, kit # 10866. Lee contacted Co-administrator Tom Clark. Mr. Purkins and Mr. Martin Voorwinden have a haplotype similar to that of myself and that of our R1a1 ChasCitCo (CCC) Tuckers, but at a large genetic distance. Tom referred Mr. Purkins to me since I am a member of the ChasCitCo Tucker family.

Several emails were exchanged between myself and Mr. Purkins and Mr. Voorwinden. Mr. Voorwinden and Dr. Klyosov exchanged a series of emails. Eventually, Mr. Voorwinden refined an extensive R1a1 haplogroup data base that he had compiled to include only R1a1 haplotypes with marker DYS388=10

an also with marker DYS388=10 plus marker DYS 448=Null. All of our ChasCitCo Tuckers have markers DYS388=10 plus DYS 448=Null. Therefore, most of the haplotypes of our R1a1 haplogroup ChasCitCo Tuckers were used by Dr. Klyosov in a recent paper:

DNA-lineages and origin of the "Tenths" the North-Western European subfamily R1a1 with DYS388=10, Anatole Klyosov, Proceedings of the Russian Academy of DNA Genealogy, vol. 3, No. 6, pp. 983-1028 (2010).

The contents of Dr. Klyosov's paper (and his previous papers) also are applicable to our other R1a1 haplogroup Tucker families, Thomas England b. 1614 (TTT group) and also to our most recent family, Hanover Co., VA, Tuckers.

The R1a1 haplogroup is a predominant one in Russia. However in Western Europe and the United Kingdom, it is a very small percentage of the population. Our CCC Tuckers are even further defined. We have the DYS388=10 marker plus the DYS448=Null marker. Consequently, one may use Dr. Klyosov's papers to trace our path for about the last 5000 years, the time when our ancient ancestor departed the Balkans moving west!

If one wishes to fully understand and follow Dr. Klyosov's papers, particularly if one is a novice like me then one needs to print our his papers in the order as listed below for study. The earlier papers give more details and more fully explain his calculation procedures. They are basically the same papers, but they are progressively refined and condensed down.

Dr. Klyosov is generally at odds with the academic world. His calculations have greatly shortened the time periods that are generally used by anthropologists in their papers.

If interested, most persons with a basic background in mathematics can follow him. Much of Dr. Klyosov's calculations involve natural logarithms. Personally, I had to do dust off the cover of an old mathematical handbook.

For giving us this knowledge and new research tool, we are deeply indebted to Mr. Martin Voorwinden and to Dr. Anatole Klyosov.

Bill, Tucker, CCC, 20652
June 24, 2010

APPENDIX

Reference:. http://en.wikipedia.org/wiki/Haplogroup_R1a_(Y-DNA)

PAPERS OR ARTICLES BY ANATOLE KLYOSOV

- Klyosov (October 2009a), "A comment on the paper: Extended Y chromosome haplotypes resolve multiple and unique lineages of the Jewish Priesthood by M.F. Hammer, D.M. Behar, T.M. Karafet, F.L. Mendez, B. Hallmark, T. Erez, L.A. Zhivotovsky, S. Rosset, K. Skorecki", *Human Genetics*, doi:10.1007/s00439-009-0739-1, http://www.springerlink.com/content/e7701424635633p7/fulltext.pdf?page=1

- Klyosov (2009) "DNA Genealogy, Mutation Rates, and Some Historical Evidences Written in Y-Chromosome. I. Basic Principles and the Method", *Journal of Genetic Genealogy* **5** (2), pp. 186-216 http://www.jogg.info/52/files/Klyosov1.pdf

- Klyosov (2009), "DNA Genealogy, Mutation Rates, and Some Historical Evidence Written in Y-Chromosome. II. Walking the Map", *Journal of Genetic Genealogy* **5** (2), pp. 217-256 http://www.jogg.info/52/files/Klyosov2.pdf

- DNA Genealogy, Mutation Rates, and Some Historical Evidences Written in Y-Chromosome, by Anatole A. Klyosov

  From Nature Proceedings. Nature Precedings is a permanent, citable archive for pre-publication research and preliminary findings.

  http://precedings.nature.com/documents/2733/version/1/html

- DNA Genealogy, Mutation Rates, and Some Historical Evidences Written in Y-Chromosome. I. Basic Principles and the Method, by Anatole A. Klyosov

  http://www.worldacademy.org/files/DNA_genealogy_Part_1.pdf

  http://www.worldacademy.org/files/DNA_Genealogy_Part_2.pdf

Bill Tucker, CCC, 20652

MY RESPONSE:

Dear Bill,

Thank you for your description and reproduction of the papers and your (and other) notes. No problem with that. I support it.

I have noticed in your description the following comment:

>Dr. Klyosov is generally at odds with the academic world. His calculations have greatly shortened the time periods that are generally used by anthropologists in their papers.

Let me comment on it. First, I am not at all "at odds with the academic world". My list of publications include more than 300 papers in peer-review (academic) journals. Most of my life I have worked in academic institutions, and for many years I was Professor of Biochemistry at Harvard University. In the mid of 1990-s I left Harvard for the industry, founded a company which soon became public, and since 2000 I am Chief Scientist of that company. I work in drug design area, particularly drugs against cancer and other severe inflammatory pathologies. As you see, I am not "at odds with the academic world", and if you look up at, say, www.amazon.com, you will see titles of some of my books.

What I am at odds with, are some "population geneticists", who are terribly ignorant and unqualified in calculations of timespans to common ancestors of populations of people. It is not a problem that they cannot make those calculations and do not have a proper background, many people cannot and do not. However, unlike those many people, those "population geneticists" actually do calculations and publish wrong data. How they do it, it is truly shocking. Most of them came from population genetics of birds and animals, and continue to employ VERY rough, crude and plainly wrong methods of calculations, which were acceptable for birds (birds did not care, right?), such as plus-minus couple of million years (who cares again, birds cannot object it anyway).

You have seen how I treat those haplotype trees. I split the tree into branches, and calculate each branch separately. Nobody in "academic population genetics" does it that way. They just have no clue. It is not their specialty. They do not use the logarithmic method as a criterion that they deal with a dataset which "follows the rule" of proper calculations. Instead of splitting the dataset into branches, they take ALL haplotypes, mix them, throw them into the "blender", and "calculate" something which is practically senseless. It is like if they want to measure an "average length of a branch on the tree" (which does not make sense at all, since branches are of all sizes, short and long) they cut the tree, chop it up, including the trunk, and measure something absolutely irrelevant. After it, they multiply that "size" by three. Why three? They "explain" that since the chopping damages the branches and makes them shorter, they decide to multiply it by three to make calculations more realistic. Go figure.

That is what they produce. They call it "academic research". Of course they are at odds with me, because most of they "calculations" is just senseless, and I say so. A few of those guys practically control academic journals in the field, and

they insist that their "multiply by three" method should be used by all. I am not kidding, this is a level of the contemporary "population genetics" regarding calculations of those timespans to common ancestors.

This is actually drama, if not tragedy, in "population genetics" for the last 10 years. Dozens of papers report, again, are trash with respect to times of population appearances, to datings of historical events. Sooner or later this huge mistake will be resolved, of course.

I thought it might be of interest to explain the situation. That is why I established a Society (Academy) of DNA Genealogy, and publish our own journal. The DYS388=10 paper is published there. It is already became known on international forums.

BILL'S RESPONSE:

Thank you. I appreciate your support. I got off lightly! From previous study of several of your papers, I understood what you have explained. I appreciate your thorough explanation. Even I, can understand that it would not be correct to throw all the data into one barrel then analyze it. Circuit analysis in my day was a big part of my education as an electrical engineer. Always, everything was broken down into its most basic parts for an analysis.

On behalf of our Tuckers, again thank you for helping us. You have greatly enlightened us.

My best.

3. DNA genealogy and mutation rate constants

Dear Dr. Klyosov—

I've seen your name come up here and there in issues related to genetic genealogy, and I'm curious. I've been doing traditional genealogy for many years (as a hobby alongside my own scholarly work) and recently have had long-lost relatives contact me to tell me they've had their genetic material tested by the FTNDA (Family Tree DNA) folks, and have found all sorts of correspondences. I'm skeptical. The way the FTNDA people seem to do their work is that they test a big pool of folks, and then if you have your own DNA tested, they match your DNA to others whose DNA is similar to yours, and then tell you that your ancestral "home" is Ireland, Britain, Czechoslovakia, or wherever.

I'm really not sure I understand most of what I've read about this testing (I'm a historian, not a geneticist), and I'm also not sure I believe what the FTDNA folks say they can do. I found your name in an article which took Dr. Michael Hammer's latest article in Human Genetics to task, and since he's the medical adviser to the FTDNA folks, I thought that was interesting.

Is there an article (that is readable for a lay audience) in publication that can give me (and my relatives) the real scoop about what can be believed about DNA testing for genealogy (and what can't)? I'd sure appreciate a bit of guidance in this area, if you have the time.

Best regards,

(…)

MY RESPONSE:

Thank you for your letter. Frankly, I receive similar letters by dozens, almost every day, and certainly several a week. However, I respond to every one of them, since I do believe that DNA genealogy is a great emerging field of science, with a great potential for history, archaeology, linguistics, family studies. People are entitled to know more about it.

However, as in any new emerging field, there are plenty of distortions, fantasies, false stories, shear falsifications, legends, myths, etc. FTDNA has contributed a good deal to that mess (a mess for some folks like you), along with a wonderful work that FTDNA has been doing for the last ten years or so. This is life, not a black-or-white thing.

Since my principal professional field is chemical kinetics (which includes biological kinetics), that is reaction rates and their computations and interpretations, I saw that it was precisely applicable for mutation rates (in the DNA) and their quantitative descriptions. This was several years ago. I have applied principal rules of chemical and biological kinetics to haplotypes and haplogroups, and realized, that DNA genealogy is in fact a blend of the DNA sequencing and chemical kinetics. Patterns of how mutations appear in those DNA fragments, selected for the DNA research (haplotypes) are nicely described in terms of chemical kinetics. Well, there is no other option, since those mutations are in fact chemical/biological reactions. It turns out that nucleotides (building blocks in the DNA) are chemically (by the enzymes, that is biological catalysts in the cell, in particular) converted to another nucleotides, hence, mutations.

Of course, every time when someone comes up with a new concept, an "official" science reacts nervously. In fact, it is not science which reacts that way, but self anointed "chiefs", which, unfortunately, control scientific journal at a great extent. The thing is that they traditionally consider "genetic genealogy", as they call it (though there is no genes in haplotypes, hence, no genetics, strictly speaking) as part of "population genetics", which does not include chemical kinetics and its approaches. Genealogy is based on time-related things, not on present-day populations, and time-related things should be calculated professionally, not by absolutely primitive and plainly wrong "approaches". To make a long story short, their "calculations" often have ~ 300% error (!). However, they BELIEVE that their values are right, they do not want to hear on chemical kinetics, on correct calculations, verified by calibration using actual genealogies and historical events, and they continue to create a terrible mess. This is life again.

This is an introduction to your questions, which are quite thoughtful. As a result, the "academic science" has created a mess (along, again, with great contributions not related to time estimates), and DNA-testing companies have a "good" share of that mess. Among them are those "matches", which you have mentioned. In 90+% cases this is just a bogus. Those matches do not mean anything, as a rule. They are unpredictable, since they are ruled by silly statistics. Mutations can happen in any marker in anyone in any generation. So, a "match" just shows a result of a (meaningless) coincidence between two individuals, the DNA (haplotypes) in which mutated - accidentally - one way and not another. There are not so many ways mutation can happen. Suppose you have haplogroup R1b1b2 (a typical one in many Europeans and their male offspring). In the first 12 markers there are only two-three positions each marker could mutate in historical time. Some markers (about four out of those 12) mutate VERY rarely. Hence, we have around 8 markers which mutate "up" or "down" by a step or two. One mutation happens - on average - once in 500 generations, that is once in 500/12 = 42 generations in the whole 12-marker haplotype. R1b1b2 arrived to Europe ~ 4500 years before present, that is ~180 generations ago. As you see, there are only 4-5 mutations in 12-marker haplotypes - on average - that the Europeans experienced after their ancestors came to Europe. Imagine, how many "matches" will be among Europeans and their descendants? Well, millions, literally. Thousands, in case of rare mutations. That is why your haplotype, or mine, or anybody's else has an astronomical number of "matches" when more people test their haplotypes, and those peoples "match" each other in any thinkable place on Earth. It has nothing to do with "relatives", except all R1b1b2 are relatives, in a way, if to consider 4500 years in Europe, and ~16,000 years since R1b was formed (via a specific mutation).

The same goes for any haplogroup. Those "matches" is just a scam, unless it is explained that those matches typically means practically nothing. The driving

force of that misleading approach is that people want to see their "relatives" (not necessarily to find them in person), since location of those “relatives” can give them an idea "where I came from?" And DNA-related companies shamelessly exploit that people's interest. None of their fliers and explanations explain it in a way that I have just explained above. They just give percentages of a likelihood that the “relatives” live this or that time ago. Of course, the longer the haplotype (25-, 37-, 67-marker length), the less amount of "matches", however, it is still a result of silly statistics. My brother would no match me, if a mutation is occurred in his haplotype. Mutations in a 12-marker haplotype happen, as I have just explained, once per 42 generations, or, equally, once per 42 births of boys in a society. In other words, mutations are ticking every day even in a small city or a town. Talk about "matches".

>I'm skeptical. The way the FTNDA people seem to do their work is that they test a big pool of folks, and then if you have your own DNA tested, they match your DNA to others whose DNA is similar to yours, and then tell you that your ancestral "home" is Ireland, Britain, Czechoslovakia, or wherever.

Exactly. It is senseless. Of course, in a small number of cases there might be some more or less distant relatives, but it is unpredictable. On the contrary, there might be no match between very close relatives. It is not science, since science is based on reproducibility of facts and observations. Here it is statistics. Statistics can be reproducible, of course, but on a different level. Those companies deliberately mix statistics with individual cases. It is like "matches" of heads and "matches" of tails when one tosses a coin. However, overall it is 50%. Science.

>I found your name in an article which took Dr. Michael Hammer's latest article in Human Genetics to task, and since he's the medical adviser to the FTDNA folks, I thought that was interesting.

Michael Hammer has done a wonderful job in studying SNPs, in popularizing "genetic genealogy", however, he did much less honorable job when dealt with time estimates. His "Cohen Modal Haplotype" story was an error, concerning "calculations" of their common ancestor. It gave the same 300% error. In fact, the same "CMH" goes with the Arabs since at least 9,000 years before present (of course, there were no "Arabs" those times, but their ancestors were). This "CMH" has to do with the Jews and the Arabs equally. The current population of the J1-"Cohens" has a common ancestor only ~1000 years ago. This was my paper in Human Genetics about. At the same time, the same calculation with the same mutation rate constants gave time of a split between Jews and Arabs around 4,000 years before present. It does make sense, does not it? In fact, it is ~ 4,000 ybp for both haplogroup J1 and J2. This is amusing, in a way. Two “Abrahams”, eh? :-))

Regarding publications on DNA genealogy in academic journals, unfortunately, practically all of them are erroneous concerning estimates of time spans to common ancestors. I can recommend you

http://www.jogg.info/52/files/Klyosov1.pdf

which contains an extended introduction. If it is understandable (skip math after the introduction), move to the next one

http://www.jogg.info/52/files/Klyosov2.pdf

Try also this one (there is a small pdf symbol there):

http://precedings.nature.com/documents/4206/version/1

P.S. Here some of my excerpts from other letters:

Here is some basics. Mutations in the DNA are of two kinds: (1) forced from "outside" (radiation, first of all), and (2) inherent mistakes done by a copying enzymatic machinery. Haplotypes in Y-chromosome are SO small compared to the whole genome, that radiation effects in them are negligible. They have nothing to do with genes. Radiation can break the DNA, but it cannot change an allele from, say, 17 to 16 or to 18. Radiation cannot change a number of repeats (STRs) in haplotypes. Hence, we are down to only one factor, that is a random change of alleles both ways - "up" or "down" (such as in DYS388 from 12 to 11 or 13, from 11 to 12 or 10, etc.

It is a truly random, statistical process. It is a "molecular clock". Nobody has ever showed that those mutations are not random. On the contrary, there are plenty indications that the mutations are random.

Therefore, common laws of chemical kinetics are applied to those mutations. This is my professional field.

The most common equation in chemical kinetics describes how "base" haplotypes disappear from the dataset due to their mutations. In chemical kinetics it is c/co = Ae^-kt. If to translate it to the language of DNA genealogy, it is [ln(N/n)/k] = n, where N is the number of haplotypes in the dataset, n is the number of base (identical) haplotypes in the dataset, k is the mutation rate constant, and n is a number of generations to the common ancestor for the whole dataset. The mutation rate constants is calculated from datasets for which a number of generations to a common ancestor is known.

To make a long story short, I have analyzed many of those datasets, and found numerical values for mutation rate constants for 12-, 25-, 37-, and 67-marker haplotypes. They are all published. Everything else in the above equation is known.

Another way to calculate – it is by using an average number of mutations per marker in a dataset. The more mutations, the longer time ago the common ancestor lived. The mutation rate constants are the same.

Since mutations do not know their history, each allele can mutate again "up" or "down". That is, alleles can return to their preceding value. Hence, "reverse" mutations, or "back mutations". They are also calculated mathematically, using the same rules and the same mutation rate constants.

In other words, DNA genealogy is a blend of DNA sequencing (haplotypes) and chemical kinetics. It is still do not recognized by the "academic community" in the area of population genetics. It is mind boggling, those folks work and live in stone ages. They do not want to hear and see anything they do not understand.

>...from different angels, e.g. which markers are slow and fast movers and what is their significance.

There is not significant whatsoever. I use AVERAGE values. It is good enough approach. With mutation rates of the individual markers, science is not there yet, since there is a huge scattering in those numbers. It is hopeless for the time being. And then, after all, it does not matter. When you measure pressure in tires of your car, you do not care how different molecules can hit it strongly or weakly from inside of your tire. Some molecules can bang against the tire as if there is no tomorrow. Who cares? You measure an average pressure. The same is with 67-marker (or whatever) haplotypes. You calculate the total, cumulative number of mutations in the dataset. Each one can happen in ANY generation. However, only their TOTAL amount divided by a number of markers (or haplotypes in the dataset) makes sense.

>I get the feeling that in the Tenth's the mutation rate of the individual markers is not the same as in general accepted (e.g. by looking in the larger family trees like the Tuckers and others).

Absolutely incorrect. It is like to say that when a Japanese tosses a coin, a number of heads and tails would be different compared with when a Dutchman tosses it. Would you buy it??

Many people wanted to show it in different populations, ethic groups, haplogroups, etc., but always failed. Do not even try, you are wasting you time. Do not try to reinvent the wheel. It is all done a long time ago.

Finally, regarding mutations rates which are "not the same as in general accepted", I do not know who is that "in general". There is a good number of various sets of mutation rates published in the literature and on personal sites. Some of them reasonably correct, some are definitely not. I employ a certain criterion in order to make my choice, namely, a comparison of calculated dates with actual ones. That is how I have collected a series of mutation rate constants for haplotypes of various formats, such as 12, 17, 25, 37, 67 marker haplotypes and other, overall more than 30 of haplotypes of different formats. For those do not "accept" mutations rates constants from my series I have only one suggestion: let us compare our mutation rate calculations with actual historic events, personal genealogies, etc. I would gladly do so.

4. On time-wise distances between pairs of haplotypes

Here are typical questions picked at random in recent discussions on some Forums:

--- I have a 62 out of 67 (62/67) markers match with someone with another surname. I've scouted around and asked questions. I don't think it's very unusual to have this. There is probably a relationship, but it may be very far back in time, perhaps long before people acquired surnames.

--- The statistical range for a 37/37 match is ridiculously large - being as low as 1 generation and going somewhere above 8-10 generations on the high side - depending on what mutation rate you use and how high you want to make the probability. Dropping the match to 36/37 basically is the same range - but with the high side sliding out at least another 5 generations. A 37/37 match is very strong - so is a 36/37.

--- I matched with 33 of 37 markers with a researcher who has documented his line to a Robert Finney of New London, PA so I am assuming that I am a Finney and somewhere down the line a Finney had changed their name to Mackey. Please let me know if I am way off base with this hypothesis. Thank you for all of your efforts as they are greatly appreciated.

--- According to one calculator, a 62/62 match at 3 generations has a 50% chance of happening. It goes up to about 75% at 6 generations. A 62/65 match at 14 generations only has a 50% chance of happening. It goes up to 75% at 20

generations. The calculator won't let me deal with 62/67 but it seems safe to say that if it was able to deal with such cases then you only get to the 75% chance of a common ancestor once you've gone back to a time before surnames were adopted anywhere in Europe. I'm not sure how many people would consider a 75% chance to be good enough. It does mean that one in every four such MRCA results taken as giving the right timeframe would be over-optimistic.

-- A 62/67 match that doesn't agree on surnames doesn't seem a terribly big shock to me.

--- I was thinking with 36 or 37 markers matched that the MRCA would be about 3 generations away from the tester. Is that true? I'm not much up on TMRCA calculations, so any help would be appreciated.

****

Folks continue to express their interest how far their haplotypes are one from another, in years or in generations. In other words, how related are their haplotypes, based on mutational differences between them.

In order to answer those questions, one should know:

(1) Whether the two haplotypes which are being compared belong to the same haplogroup,
(2) Whether they do belong to the same haplogroup, but belong to different subclades,
(3) Whether they do belong to the same subclade, but to different branches within the same subclade.
(4) Whether the two haplotypes belong to the same branch on a haplotype tree.

The case (3) can in fact be the case (2) when the two branches represent two different subclades, not identified as subclades as yet.

In fact, only in case (4) it makes sense to calculate a time-wise distance between the two haplotypes. However, a margin of error for such a calculation would be a very significant, typically between 100% and 50% within 95% or 67% confidence interval.

Now we can answer the above questions posted in various Forums.

1) I have a 62 out of 67 (62/67) markers match with someone with another surname… There is probably a relationship, but it may be very far back in time, perhaps long before people acquired surnames.

5 mutations on 67 markers are translated to 900±440 years between the two haplotypes. If the two haplotypes belong to the same branch, that is derived from the same common ancestor (and the two haplotypes are equidistant from the ancestral haplotype), then a common ancestor of the two haplotype lived approximately 450±220 years ago, that is around 1560 AD plus-minus a couple of centuries. Surnames were already in use those times.

2) The statistical range for a 37/37 match is ridiculously large - being as low as 1 generation and going somewhere above 8-10 generations on the high side - depending on what mutation rate you use and how high you want to make the probability. Dropping the match to 36/37 basically is the same range - but with the high side sliding out at least another 5 generations. A 37/37 match is very strong - so is a 36/37.

Well, the 36/37 "match" translates to 1/0.09 = 11±11 generations before present. The 37/37 match basically translates to the same time span. The thing is that mutation or no mutation in a 37-marker haplotype is just a matter of a simple chance.

3) I matched with 33 of 37 markers with a researcher who has documented his line to a Robert Finney of New London, PA so I am assuming that I am a Finney and somewhere down the line a Finney had changed their name to Mackey. Please let me know if I am way off base with this hypothesis. Thank you for all of your efforts as they are greatly appreciated.

Yes, you are a little bit off base. 4 mutations on 37 markers are translated to 4/0.09 = 44 generations to a common ancestor without a correction for back mutations, or 46 generations with the correction, that is 1150±590 years between the two haplotypes. If the two haplotypes belong to the same branch, that is derived from the same common ancestor (and the two haplotypes are equidistant from the ancestral haplotype), then a common ancestor of the two haplotype lived approximately 575±295 years ago, that is around 1435 AD plus-minus three centuries. It is a bit early to rely on their last names.

4) According to one calculator, a 62/62 match at 3 generations has a 50% chance of happening. It goes up to about 75% at 6 generations. A 62/65 match at 14 generations only has a 50% chance of happening. It goes up to 75% at 20 generations. The calculator won't let me deal with 62/67 but it seems safe to say that if it was able to deal with such cases then you only get to the 75% chance of a common ancestor once you've gone back to a time before surnames were adopted anywhere in Europe.

A 61/62 “match” results in approximately (and on average) 8±8 generations difference between the two haplotypes. The 62/62 match has a pretty much the same probability. I do not quite understand what is “at 3 generations has a 50% chance of happening”.

5) A 62/67 match that doesn't agree on surnames doesn't seem a terribly big shock to me.

5 mutations on 67 markers are translated to 1100±540 years between the two haplotypes, and if the two haplotypes belong to the same branch, that is derived from the same common ancestor (and the two haplotypes are equidistant from the ancestral haplotype), then a common ancestor of the two haplotype lived approximately 550±270 years ago, that is around 1460 AD plus-minus a couple of centuries (see above). I do not know is it “a terribly big shock” or nor in terms of surnames existence. My ancestor, for example, was born in 1575, and he had the same surname as mine. It was in Russia, in the Kursk area. In the Isles, surnames became common after Henry VIII

6) I was thinking with 36 of 37 markers matched that the MRCA would be about 3 generations away from the tester. Is that true? I'm not much up on TMRCA calculations, so any help would be appreciated.

See above. The 36/37 “match” translates to a common ancestor who lived 11±11 generations before present.

5. R1a1 haplogroup in India

Your Indian R1/R1a1 has a deep, ancient history. I do not know who had estimated your R1a as 10,000 years "old" and on what basis, and why its location is necessarily in Kazakhstan.

I suggest you disregard it for the time being unless some DATA are presented by those who suggested it.

Back to India and R1a. There are two principal lines of R1a in India. One line is a really ancient one. It has appeared apparently in South Siberia 21000±3000 years before present (Klyosov, J. Genetic Geneal., vol. 5, No. 2, pp. 217-256, 2009), and nowadays its haplotypes spreads as a rather wide band from South Siberia (Altai)-North-Western China through India-Pakistan to as far to the South West as Oman and Egypt, with a gradient (of their common ancestors) down from 21000 years before present in North China (up to 25% population there is R1a/R1a1) via Pakistan (9000 ybp) and India (8000 ybp) to South Central Asia (6900 ybp) and to South Arabia and Egypt (6000-4500 ybp). They are "non-

IndoEuropean" R1a1, linguistically speaking. A good part of that population speaks Altaian-Turk group of languages, reflecting their historical path from South Siberia-Altai.

Another, "Indo-European" R1a1, has different haplotypes. They are easily recognizable and distinct from the "non-IE" R1a1 haplotypes. They have a common ancestor in Europe around 9,000 years bp and by all means traveled all the way from South Siberia, bringing their "Europeoid" anthropology. Funny, but Europeoids-Caucasoids appeared in fact in South Siberia, came to Europe, and were named "Europeoids" since they settled in Europe many thousand years back, first, apparently, on the Balkans. They brought (or worked out) Indo-European language. Those R1a1 were proto-Aryans, since eventually it was them who brought IE language and their R1a1 haplotypes to India around 3500 years ago (see below).

These R1a1 had populated Europe from the Balkans to the Isles, and from Scandinavia to Greece, and around 6000-5000 years ago spread to the East, to the Russian Plain (aka East-European Plain). A common ancestor of ethnic Russians-Ukrainians-Belorussians of R1a1 haplogroup lived on the Russian Plain 4800 years ago, and if to add to said populations also Poland, Germany, Scandinavia, and ALL other European R1a1 populations (altogether 14 R1a1 principal branches, including two branches of M458 subclade), a common ancestor would be of 5100 years of “age”.

Between 5000 and 4000 ybp R1a1 from the Russian Plain moved to the East, established Andronovo archaeological culture (North Kazakhstan, South Ural and more to the East), built settlements such as Arkaim in South Ural (3800-3600 ybp) and many others in the area, established "Avesta culture" in the South of Central Asia (~4000-3500 ybp), expanded to the Caucasus by 4500 ybp, spread over the Caucasus to Anatolia by 3600 ybp and confused linguists that much that the latter believe that "Indo-Europeans" appeared in Asia Minor - Anatolia along with their language.

Finally (in this context) R1a1 went from South Ural to India as the Aryans, and from South Central Asia (Avesta Aryans) to Iran by about 3500 years ago in the both directions. As a result, R1a1 haplotypes in India, Iran, and in Russia are practically identical to each other (there are some branches in Russia which are more inclined to the West, to Poland, for instance). Personally I belong to the Russian branch whose haplotypes are identical with the Indian "IE"-haplotypes up to 67 markers long. On a 67 marker R1a1 haplotype tree my haplotype sits on the same branch along with a bunch of Indian R1a1 haplotypes.

This is a story of R1a1 in a rather brief format. Last year I have published (along with a colleague of mine, Igor Rozhanskii) a detailed study of R1a1 of more than

a hundred pages long, with dozens of graphs and haplotype trees. Of course, this story can and should be updated, modified, corrected, however, based on DATA, not on "opinions".

6. On the Ballantyne's et al (2010) paper on father-son mutation rate constants

Ballantyne, K.N., Goedbloed, M., Fang, R., Schaap, O., Lao, O., Wollstein, A., Choi, Y., van Duijn, K., Vermeulen, M., Brauer, S., Decorte, R., Poetsch, M., von Wurmb-Schwark, N., de Knijff, P., Labuda, D., Vezina, H., Knoblauch, H., Lessig, R., Roewer, L., Ploski, R., Dobosz, T., Henke, L., Henke, J., Furtado, M.R., Kayser, M.(2010) Mutability of Y-chromosomal microsatellites: rates, characteristic, molecular bases, and forensic implications. Am. J. Human Genet. 7, 341-353.

(Excerpts from discussion letters)

MY LETTER:

The data in the paper (see above) are certainly very interesting to consider (and re-consider) in detail.

Now, the main conclusion. The Ballantyne's data are fine, except they are applicable only to 12 and 25 marker haplotype formats in the FTDNA standard [and, of course, to many various shorter haplotype formats, for which I have more than 30 mutation rate constants on my list, though for each one Ballantyne's data should be examined and verified], since some markers for the 37 and 67 marker formats are missing in the quoted paper. Also, the Ballantyne's data have some systemic problem, characteristic to father-son pair studies – statistics is not there. On the first glance, how could it be, since they have studied nearly 2000 father-son pairs?

Let me explain it. Indeed, they have studies nearly 2000 father-son pairs, and measured mutations in 186 markers. Sounds impressive, isn't it? However, let us look at concrete data. In the first 12 markers (according to the standard FTDNA nomenclature) there are 3 mutations in DYS393 among 1750 father-son pairs, 2 mutations in DYS390 among 1758… well, you got the idea. As you already see, statistic is not here, and the overall, average numbers will not be very reliable. Anyone who understands mutations in DNA genealogy, would have already raised his/her brow: DYS393 is at least 4-times slower marker compared to DYS390, however, in the quoted data they are just about the same.

However, let's continue. For the next 10 markers numbers of mutations are as follows: 7, 5, 3, 6, 0, 0, 6, 9, 1, 6. There are 48 mutations altogether in those 12 markers, which gives 4 mutations per marker in 1727±49 father-son pair (they varied for different markers, which is also not a very good thing, but a realistic, almost unavoidable one). It gives 4/1727 = 0.00232±0.00033 mutations per marker per generation, and 0.028±0.004 mutations per haplotype per generation (the margin of error is calculated taking into account that there were only 48 mutations in the dataset). My mutation rate for the first 12 markers (in fact, it is the same as the Chandler's one) is 0.00183±0.00009 and 0.022±0.001 per marker and per haplotype per generation, respectively. As you see, these figures (Ballantyne's and mine/Chandler's) are almost within the same margin of error. However, my mutation rates are calibrated for 25 years per generation. That means that the Ballantyne's data would exactly fit my data for 32 years per generation (0.028x25/32 = 0.022). This makes sense. You see, the systemic problem with father-son data is that they can work only with "generations", while history works with years, NOT with generations. Therefore, the father-son data need to be calibrated anyway, as in this case with my data, calibrated elsewhere.

Anyway, let's remember, that for 32 years per generation the Ballantyne's data coincide with my mutation rate constant for 12 marker haplotypes (and with John Chandler data taken for 25 years per generation, though in his case it was for an unspecified "generation", as in the Ballantyne's case).

If we calculate their data more precisely, per each transition separately, we get 277.157 mutations per (theoretical, extrapolated) 10,000 pairs father-son, that is the mutation rate of 0.0277±0.0040 per haplotype, which is the same value as above, and all the above considerations are equally applicable.

If we move to 25 marker haplotypes, than a number of mutations was 14, 4, 0, 0, 3, 2, 0, 19, 12 (DYS459a,b were combined, as well as 464a,b,c,d), that is 54 mutations were added, making the 25 marker format having 48+54 = 102 mutations per 1704±86 father-son pairs. This will give 102/1704 = 0.0600±0.0059 mutations per haplotype, that is 0.0024±0.0002 mutations per marker. My calibrated mutation rates give 0.046 and 0.00183 mutations per haplotype and per marker, respectively. If to calculate the Ballantyne's data more precisely, they would give 594.915 mutations in 25-marker haplotypes per (theoretical) 10,000 father-son pairs, that is the average mutation rate will be 0.0595 per haplotype, which is practically the same value of 0.060, calculated above. Again, the Ballantyne's series is on the higher side. However, they would exactly fit my mutation rate at 32 years per generation, 0.0595x25/32 = 0.046.

First conclusion: the Ballantyne's mutation rates would be O.K. at 32 years per generation, and in this case they would fit my mutation rates for 25 years per generation.

Unfortunately, the Ballantyne's study misses markers from the standard 37- and 67-marker haplotypes, and the respective mutation rates cannot be calculated.

Second conclusion: My mutation rates are correct for 25 years per generation, as I use them for calculations.

Now, a few words about the Chandler's data. His 12-marker average mutation rate is practically identical with mine (0.022 mutations per haplotype per generation), but his 25-marker panel is grossly off being too high. He overestimated, as I see it, DYS464, since did not take recLOH into consideration. Hence, he brought in toooooo many extra mutations. The same was with his 37 marker haplotypes, his data are not applicable for calculations. I have described it in my paper in J. Gen. Geneal., 2009 (vol. 5, pp. 186-216). For the 25 marker format the Chandler's average mutation rate is 0.0695 per haplotype per generation (in my case it is 0.046 at 25 years per generation, for Ballantyne's it is 0.0595 at 32 years per generation). Furthermore, according to the Chandler's data, only DYS464a,b,c,d contribute 0.00566x4 = 0.02264, that is 33% of the whole 24 marker panel (!). Obviously, their "weight" is grossly exaggerated on the Chandler's list. Anyway, the Chandler 25 marker mutation rate fits mine and Ballantyne's ones at 38 years per generation (0.0695x25/38 = 0.046, 0.0695x32/38 = 0.0585; 0.0695x32/37.4 = 0.0595).

In other words, the Chandler's data work for 12 marker panel with 25 years per generation, but for 25 marker panel with 38 years for generation. Therefore, the Chandler 25 marker panel in erroneous. Even more erroneous is his 37-marker panel, which gives an average mutation rate for the panel of 0.182. If fact, it is 0.09, that is two times less. It would work for Chandler only for 50 years per generation. Nice.

RESPONSE:

Thanks for your insights. When I do my TMRCA I omit all of the multi-copy markers. Thus any errors that John made in calculating the mutation rate for DYS 464 don't affect my calculations. I am currently using 30 years as the generation interval in my program, which is very similar to Ken's Generations 5 program. It is certainly difficult to know the generation interval in ancient times with precision. In any case, I find this topic extremely fascinating and have certainly appreciated your insights on the DNA list. I still think that the issue of trying to figure out which markers to use for specific calculations is a thorny one. I feel quite comfortable including the faster markers in situations where the

TMRCA is less than 1000-2000 years, but figuring out which markers to include with in situations where the TMRCA is over 10,000 is more challenging.

MY RESPONSE:

As you well know, discussions on the mutation rates can be endless, because a number of variants - which markers to pick and in which situations and which of them omit and what a generation length to pick - is endless. In a situation such as this one the problem is solved in four stages.

Stage one: someone should move ahead and make a decision (for himself, at this stage). I made it for myself, and for the last two years I employ a firmly defined set of mutation rate constants for 6-, 12-, 17-, 25-, 37-, 45-, 67 marker haplotype formats and for twenty-five more formats used in the literature. All of them form a cross-linked net of figures. All of them are published. All include fast- and slow-markers, which is the right way to do. By this way statistics is the best, and time intervals for analysis are the widest - from a few generations to thousands of generations.

Stage two: to verify these mutation rate constants with real, practical examples, both private genealogies and historical events. I ran hundreds of them, and came to a conclusion that my set of mutation rate constants works the best. To do it properly, I always make corrections for back mutations, introduce corrections for asymmetry of mutations (in some cases, when needed), make corrections for recLOH, resolve branches using haplotype trees (the must), use the logarithmic method for cross-verification of data, where possible (VERY useful and informative), and set a generation length (25 years) firmly as an integral part of my set of mutation rate constants. As you well know, you can not just say - "I use 30 years", or "32", or "20", or "25" without affecting your mutation rate constants, because in reality we always determine a product KT, where K is the mutation rate constant and T is a number of generations to a common ancestor. You just cannot separate these factors without changing the outcome. A generation length is a part of the equation, you cannot say "I currently use this" without a proper, mathematical justification. That is why father-son data are useless unless you have a mathematically (!) determined the generation length. A conclusion: I have done Stage Two. It is finished.

Stage three: you run multiple examples and convince people, including the genealogists, that your set of mutation rates and the approaches are correct and it does work. A firm rule: you cannot change game rules with different genealogies and other data, you cannot remove markers as you wish in one case and bring them back in other cases. I analyze those data and genealogies almost every day, and publish the most representative and interesting in our Proceedings in a separate section "Personal cases". More than 60 cases are already published, and

this is a small fraction of all. The rules are always the same. My firm conclusion: my system does work.

Stage four: establishing the consensus, that this set of data is good to work with. Skeptics, of course, will be, as always, but this is life.

A few secondary remarks. DYS464 are very useful markers in composing of haplotype trees, because those recLOH is a branch-forming and branch-resolving event. That is why I never remove DYS464. Their recLOH are very visible, and I use them as one mutational event, one-step (despite it looks like 2-, 4-, 6-mutational event). It does not introduce a significant mistake even in some dubious situations, it is always within a margin of error.

7. R1a1-M458 in Italy and principles of DNA genealogy

AN INCOMING LETTER:

I came across a number of your highly interesting articles on genetic genealogy when researching a German client's ancestry. You have him in the map of the "Central European Branch" of R1a1a7.
I understand that you have a lot of R1a data on file, so I hope you might be able to point us in the right direction, or "cluster".

We were able to trace, in what is sort of an educated guess, my client's ancestry back to a family, which is first documented around 1100 AD in Northern Italy, north of Milan. The first name bearers there have clearly Germanic given names, as is the case for at least half the early and high medieval population there. So far, we have been unable to find any other male line apart from my client's. And it is extremely improbable that that we will ever find some appropriate line. Also, so far, we have been unable to find a really good explanation for the family name. The area shows a number of medieval family names similar to this type and, seemingly, referring to localities. I recently came to the conclusion that the name might, theoretically, also have some Slavic roots, but lacking any knowledge of Slavic languages, I am not sure about that.

In any case, we were rather surprised to find his haplogroup being R1a1a7 (M458+, SNP tested), instead of some "Germanic" or "Roman" or whatever group. Of course we cannot exclude illegitimacy on the way, but let us assume for the moment that paper trail and DNA genealogy are

consistent - otherwise, all the fun would be lost.

More specifically, the client belongs to what you have defined as the "Central European Branch". The map clearly shows that he is, geographically, quite far from everybody else, and whatever I do, the closest I can get in terms of MRCA is a little over 30 generations, and those I've contacted weren't able to trace their genealogy farther back than 18th or 19th century, and usually end up somewhere in Poland.

According to your papers, the age of the "Central European Branch" of R1a1a7 is something around 2500-2700 years. I think Peter Underhill is far too high with his guess - however I'm just a historian and barely able to understand all those calculations, let alone the implications of population dynamics.

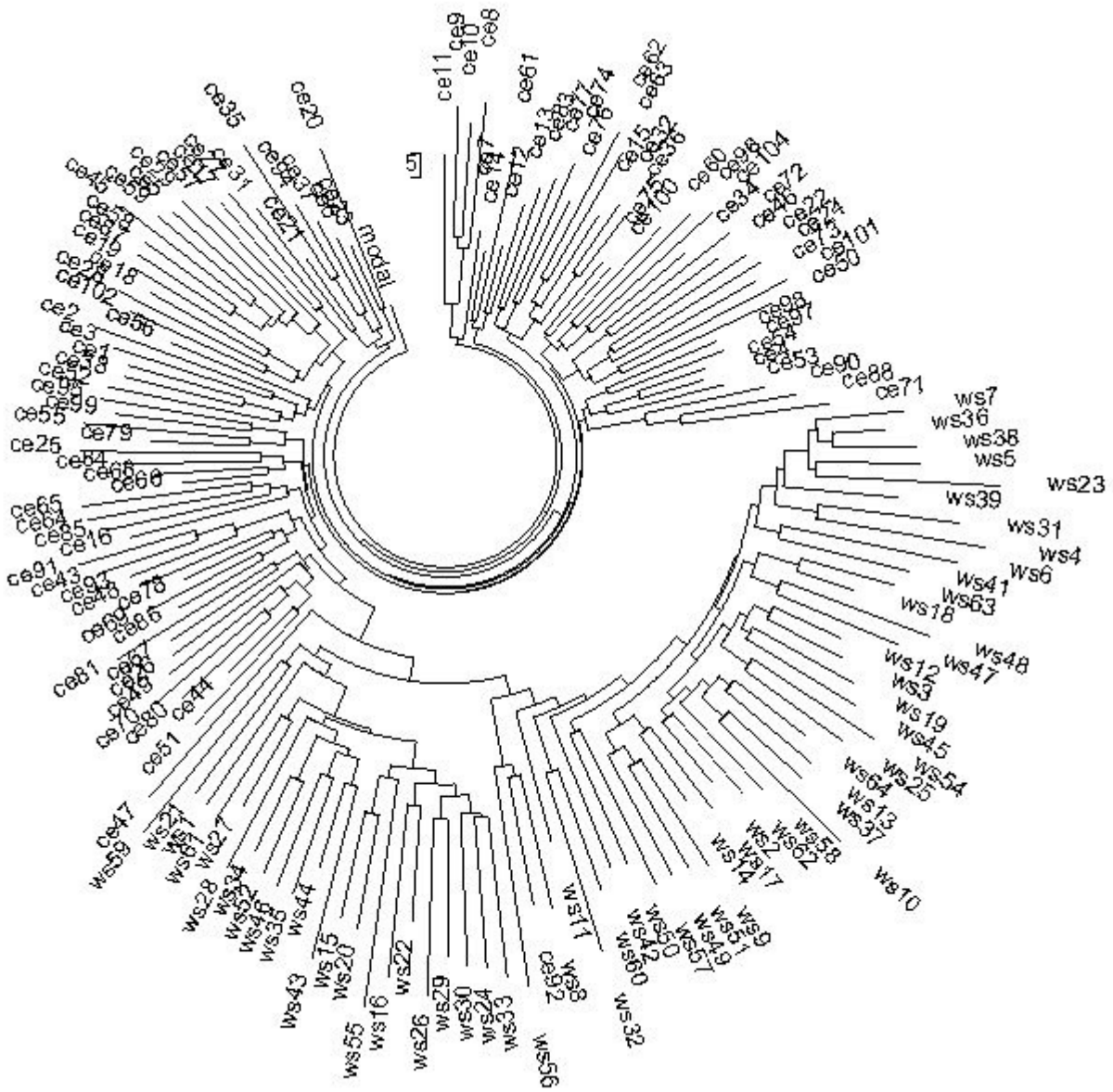
We know that Lombardy - the area where my client might be associated – was settled by the Germanic Longobards during the 6th century AD. We also know that the Lombards were accompanied by a number of non-Germanic tribes they had picked up during their journey through Eastern Europe; it would take a bit of additional research to get a better idea which tribes exactly were involved in the migration. So at the moment my best guess is that my client descend from a member of one of those non-Germanic tribes, who went with them and settled in Lombardy. I just wonder why there weren't more of them (well; maybe they haven't yet been tested), and/or, how I could discern the "Germanic" DNA in that area.

Also, I cannot find any detailed studies about the distribution of haplogroups in Northern Italy, migrational patterns in that area, or anything referring to aDNA except the articles about the German and Central Asian cases.

So - what do you think? Where should we search? Is there some sort of more specific cluster you could put that haplotype in?

MY RESPONSE:

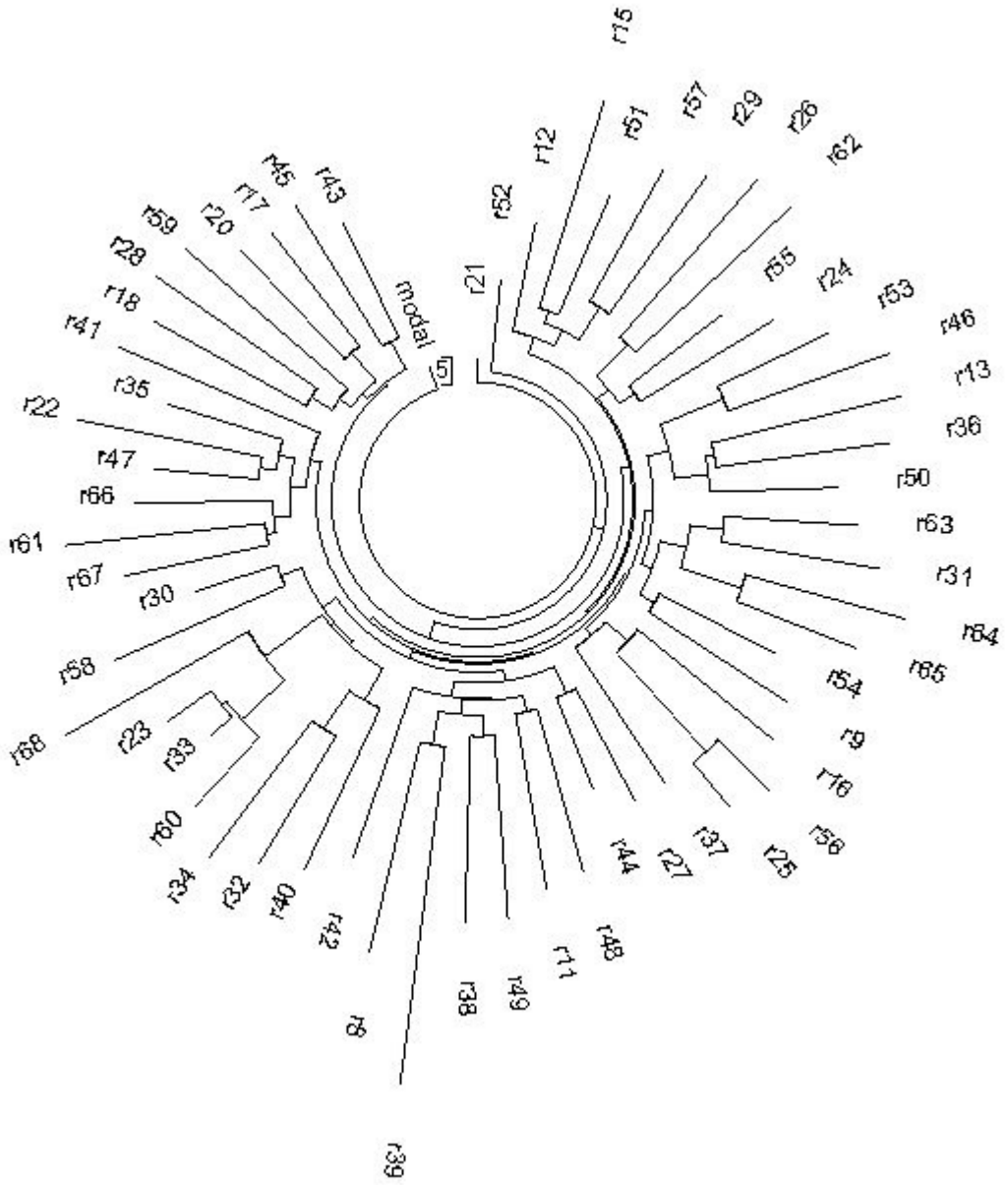
Attached are two haplotype trees, one a 67 marker one, in which your client's haplotype under a name ce96 sits in the upper right "corner"; another is a 69 marker tree (including DYS464ef), in which he is under name r57.



A 67 marker haplotype tree of R1a1-M458 subclade. Haplotypes marked CE belong to the Central European branch, and WS are those which belong to the Western Slav branch (Klyosov and Rozhanskii, 2010).

The first tree combines so-called "Central European" and "Western Slav" branches of R1a1 haplogroup. The second tree shows the "Central European branch", those who include so-called recLOH mutation, which forms a separate sub-branch. Explanations are partly given in

http://maps.google.ru/maps/ms?ie=UTF8&hl=ru&msa=0&msid=114781513110833464918.0004689cce3eb079d8da4&ll=54.876607,14.677734&spn=36.666387,76.376953&z=4



A 69 marker haplotype tree of R1a1-M458 subclade (including DYS646e and 464f), the Central European branch (Rozhanskii and Klyosov, 2010).

The first tree embraces individuals with M458 SNP mutation in their Y-

proto-Slav, if you wish) origin. Their ancestor moved (migrated) to the West in the middle-to-end of the 1st century BC from the East European Plain (its Western part).

His (ce96) closest neighbors in the tree are ce104 and r15, ce100 and r12, and also ce75, ce60, ce34, r29, r51. I am forwarding to you their names and the FTDNA and YSearch ID indexes.

Peter Underhill was, unfortunately, totally incorrect with his calculations. They are based on wrong assumptions. Even more unfortunately, the whole area of "academic science" follows this wrong way for the last six or so years. A damage which was done to the field is hard to overestimate. They are "population geneticists", and mutation rates and the respective calculations are certainly not their forte.

Personally I have not studied much of Northern Italian haplotypes and haplogroups (I know the situation in general, and have studied R1b1b2 group there, but not really R1a1, which there are only a handful of representatives there). It might be that R1a1 in Italy were those "Barbarians" (though they sometimes made their way to become Patricians and Kings of Italy, as we all know). We also remember a story on a sword and a scale..., and words “Vae victis” (woe to the conquered). Anyway, the story of R1a1 in Rome and Italy in general is still awaiting their historians. If you want to close on it, I might be of help. Start with ethnonyms Venetian, Veneto, Veneti, etc.

RESPONSE:

Oh. You are right about Venetian, Veneto, Veneti. One should follow that further. The Venetians of course are settling a little east of the area around the Lago di Como, but this really looks interesting - most especially because of the possible connection to the people called Wenden in German.

I tried to calibrate the data by comparing the Altai R1a1a* with the age of the archaeological findings there, but of course I am an absolute greenhorn in those calculations. On the other hand, those thoughts about population bottlenecks and slower growth also might apply in some cases.

MY RESPONSE:

Regarding the Altai R1a1 data, very often it is hard to compare them with archaeological data, due to lack of our knowledge. Often the DNA data go deeper in time, since archaeology cannot claim that they have reached the bottom of history (of course). Often the DNA data are much more shallow due to

population bottleneck. Rarely (indeed) archaeological and DNA data match each other. That is why they badly need each other.

"Slower growth" is irrelevant in this context, but population bottlenecks are ALWAYS important.

Regarding your questions on barbarians as Italian kings, of course you, being a historian, know what I am talking about. *De jure*, Odoacer was not an Emperor of Rome, but *de facto* he was. Look up into Edward Giibbon's "The History of the Decline and Fall of the Roman Empire", vol. IV. Odoacer, "the bold barbarian", a Germanic chieftain, deposed Romulus Augustus on September 4, 476, and become the ruler of Italy. He was granted the title Patrician only because he was not allowed the title of Emperor, he returned the Imperial insignia to Constantinople and ruled as King of Italy being blessed by the Eastern Empire. As Gibbons notice, "Odoacer was the first barbarian who reigned in Italy, over a people who had once asserted their just superiority above the rest of mankind". And then "the patient Romans were insensible prepared to acknowledge the royalty of Odoacer and his barbaric successors".

Regarding a sword and a scale... look up in Plutarch, LIVES, vol. III, Camillus. Remember Brennus, who invaded Rome ca 390 BC? Plutarch: "Brennus in a scoffing and insulting manner pulled off his sword and belt, and threw them both into the scales; and when Sulpicius asked what that meant, "What should it mean", says he, "but woe to the conquered" (Vae victis).

RESPONSE:

I'm still trying to understand the mathematics behind all that DNA genealogy, and to me it seems that the biggest problem is the mutation rate applied. I understand that research is still in progress, and that results can vary extremely (e.g. Underhill/Zhivotovsky vs. "rest of the genealogical world"). I also understand that father-son rates do not seem to apply directly because they ignore other, "demographic" influences. I also understand that it may be relevant whether one includes markers like DYS464 or not. Just now there is no access to the Chandler article (the jogg.info website seems to be down temporarily), so I am not sure if he offers father-son rates or "effective" ones. In any case, the Chandler rates are much higher than yours, and if I apply the Chandler rates, I get TMRCA results on R1a1a7 with recLOH mutation which are way too young - just a little over 50 generations, which simply doesn't work. The R1a website puts that group's age at 2400 years (or 96 generations). My own data set includes a little over 60 67-marker-haplotypes, all supposedly R1a1a7, all with markers DYS464a-f (excluding those with a-g).

How did you calculate the mutation rates for the FTDNA marker series in Table

1 of your "DNA genealogy, mutation rates, etc." article? And how do you get from columns 2 to 3 in Table A of the same article?

MY RESPONSE:

The mathematics behind "all that DNA genealogy" is very straightforward, and based on theory of probability and chemical kinetics, with some twists (though completely mathematically justified) related to back mutations, symmetry of mutations, etc. All separate parts of my approach are well known in those areas of science, though nobody seemingly brought them together just because there was no such a need. Now, with the appearance of DNA genealogy, such a need became a necessity. Hence, I did it.

Since "population geneticists" are not familiar with chemical kinetics and its apparatus, they failed with their approaches to mutation treatment. In some simplest cases their approaches work, for instance, where there is only one lineage in the dataset, and a common ancestor lived quite recently (typically, no more than 1000 years ago). A dataset with two and more lineages they simply through into a "blender", grind and apply some funny "mutation rates", which are about 2.0 – 3.6 times less that actual mutation rates (they use 0.00069 mutations per marker per generation, however, in reality they are between 0.00135 to 0.0025, for a selection of 30 different haplotype formats, as you can see in Table 1 in my paper which you have cited). As you see, they (a) do not resolve branches, and, mind you, each branch has its own common ancestor, (b) they did not care about which haplotype format they employ, and for all (!!) they use the same "magic" 0.00069 mutation rate, (c) they do not care how old is the dataset (in terms of a timespan to the common ancestor, hence, they do not introduce any correction for a back mutation, which in fact vary from 1.0 to 4.8 times (see Table A in the cited paper).

As a result, they get absolutely phantom timespans to common ancestors, and make "historical" conclusions taken from their blend. It is mind boggling, how unprofessional and ignorant those people can be, and they RULE academic science! As a result, that "academic science" of the last ten years or so became so damaged, when we talk on dating, chronology, and historical "conclusions" based on those "calculation". This is, first and foremost, such names as Zhivotovsky, Underhill, Cruciani, Kivisild, Behar, and other "reputable" names in population genetics. On the other hand, they do a great job (except Zhivotovsky, of course, who is a mathematician) on typing haplotypes and haplogroups, and collecting good databases. I would not have mentioned names, this is not my style, but after their "Response" to my critique in Human Genetics all bets are off.

This is a brief summary of a current situation in DNA genealogy. Those folks do not want to listen to my reasoning in that area (the mentioned Response in "Human Genetics" on my critique was a good example), because, probable, they are weak to admit that they have produced too much wrong data for the last several years in that area. As to my work, currently I productively collaborate with very good genealogists, and they are astonished how well my calculations fit to their documentary evidences.

Regarding father-son mutation rates. They are supposed to be the best way to lay a ground for mutation rates to be used in DNA genealogy. Unfortunately, it is only "supposed to be". Reality is more complicated, and for precise (or appropriate) data we should have hundreds of thousand, or even millions of father-son pairs. So far it is unthinkable and astronomically expensive. The problem is that mutations are seldom, and statistics is not there. As a result, father-son mutation rates from different studies vary from about 0.0016 to 0.0046 mutations per marker per generation. This uncertainty, of course, is not applicable for calculations in DNA genealogy. The good thing is that mutation rates which I use, for different haplotype formats ranged from 0.00183 (12- and 25-marker haplotypes) to 0.0020 (17-marker and 45-marker haplotypes) to 0.00216 (67-marker haplotypes) to 0.00243 (37-marker haplotypes) mutations per marker per generation (of 25 years), that is fit well within the father-son (wide) range. These mutation rates were obtained from calibration of a number of mutations in haplotypes in extended genealogies with known timespan to a common ancestor. They are confirmed in dozens of "conventional" genealogies (I have published a number of studies of this kind) and in a good number of historical events, dates for which are more-or-less known. I am now working with professional genealogists on a joint paper to show how "documentary" genealogy and my calculations nicely fit each other. I do not have a slightest doubt that my mutation rates and calculations are incorrect. Of course, they can be tweaked a little bit (however, it would be rather silly, since it would be within a margin of error anyway), but principally they are correct. Statistics of mutations themselves introduce a larger variation compared with a (potential) variation of mutation rate constants.

Regarding your remark that "I also understand that father-son rates do not seem to apply directly because they ignore other, "demographic" influences", they can be applied (see the preceding Discussion in this issue – AK), and "demographic" influences have nothing to do with it. Typically, all those "demographic influences" is just a smokescreen for the skeptics. Of course, there "population bottlenecks", migrations, etc., but all of them can be objectively analyzed using the same approaches.

You have also noticed "I also understand that it may be relevant whether one includes markers like DYS464 or not". It is not a problem again. So-called

recLOH factor is identifiable and does not provide any problem. On the opposite, it is a very valuable marker(s).

>*… I am not sure if he (Chandler) offers father-son rates or "effective" ones. In any case, the Chandler rates are much higher than yours, and if I apply the Chandler rates, I get TMRCA results on the "light blue pins group" (i.e. R1a1a7 with recLOH mutation) which are way too young - just a little over 50 generations, which simply doesn't work.*

Chandler offers neither father-son nor "effective" ones (though the latter needs a definition). He has collected thousands haplotypes from databases and counted mutations. In fact, his 12-marker average mutation rate is practically identical with mine (0.022 mutations per haplotype per generation), but his 25-marker panel is grossly off as too high. He overestimated, as I see it, DYS464, since did not take recLOH into consideration. Hence, he brought in too many extra mutations. Same was with his 37 marker haplotypes, his data are not applicable for calculations. That is why you are absolutely right, it does not work.

And you know why? You are the first one who said it (except me, of course, since I have published it and had lengthy discussions with the editor). Is it not amazing? Dozens and hundreds (probably) people mentioned the Chandler's mutation rates, and nobody has bothered to examine them and say that they are unrealistic, as you did. It is again mind boggling. If someone would have taken a dataset and applied the Chandler's mutation rates to 12- and 25-marker panels, he/she would see that results will be grossly different. They would not fit each other. Again, it seems that nobody bothered to do it.

>*How did you calculate the mutation rates for the FTDNA marker series in Table 1 of your "DNA genealogy, mutation rates, etc." article? And how do you get from columns 2 to 3 in Table A of the same article?*

A good question. It means that you indeed think about details. I like it. First, on Table A. Column 2 shows an average number of mutations per marker, how they are observed and calculated directly. For example, if a dataset of 100 of 25 marker haplotypes contain 400 mutations from the base haplotype, then 400/100x25 = 0.16 mutations per marker. At the mutation rate of 0.002 it would give you 0.16/0.002 = 80 generations, that is 80x25 = 2,000 years to a common ancestor.

However, one should know that at 24 generations and deeper one should introduce a correction for back mutations. At 80 generations it is about two centuries. Not a big deal, but still some. Since "forth" and "back" mutations must have the same mutation rate (a copying enzyme does not know to which side it makes a mistake when copying), then a plain probability tells us how to make that correction. This is formula (4) in the paper.

At a small number of (observed) mutations per marker (less than 0.046, that is 23 generations and less), exp ~ 1, and the formula gives no corrections (it is too small). “2” in the denominator and “2” in the nominator just cancel each other. At 0.16 of mutations per marker (see above), that is “apparent” 80 generations, or 2000 years, exp = 1.1736 (check yourself, take a1 =1 for a simple case, that is mutations are symmetrical). With exp = 1.1736 a corrected number of mutations per marker is 2.1736/2 times higher, that is 80x1.0868 = 87 generations. Check it at the bottom of the first page of the Table A, next to 0.16 (first column) and 80 (second column). It is 87. The last column shows 2175 years (the corrected value). Simple?

RESPONSE:

Ah. That explains a lot. Thank you very much.

Also, as you had explained, for the panels with 25 markers and more, logarithmic approach doesn't work.

My RESPONSE:

I understand that you want to play around with different combinations of markers; everyone does it or did it or will do it. It is a natural phase for a novice, however, I have done it years back and do not want to do it again. It does not make sense for me to take the Chandler rates with or without certain markers, this way leads nowhere.

I have a clearly defined set of mutation rate constants for more than 30 different haplotype formats, so why do I need the Chandler's table? Everyone has tried to work only with the slowest markers, etc., this is not a good way ether, since you lose statistics – unless you work in a really ancient timeframe. Only the fast markers are not good and only the slowest markers are not good. A good system should contain both, that is I remove nothing from those 67 markers.

Hence, you should ask yourself - either you are just playing with different combinations of markers to get a sense of it, or you want to create your own system for calculations. In the last case you are joining a huge crowd of irreproducible approaches and irreproducible data, in which other people do not have any interest.

>*Also, as you had explained, for the panels with 25 markers and more, logarithmic approach doesn't work.*

No, it is not true. I have used logarithmic method not only with 25-, but also with 37- and 67-marker haplotypes. It all depends on a size of a dataset and on how "old" it is. For example, for a hundred of 25-marker haplotypes with a common ancestor of 800 years "old" as many as 24 haplotypes with be "base" (identical to each other). This will be perfect for 25 marker haplotypes. In fact, logarithmic method work fine with all complications of DYS464, since it does not consider a number of mutations, recLOH, etc. It just needs identical haplotypes.

LETTERS FROM THE READERS: PERSONAL CASES

Anatole A. Klyosov

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

Part 28

Below is a selection of letters in English, published in previous issues of the Proceedings.

LETTER FIFTY-FIFTH

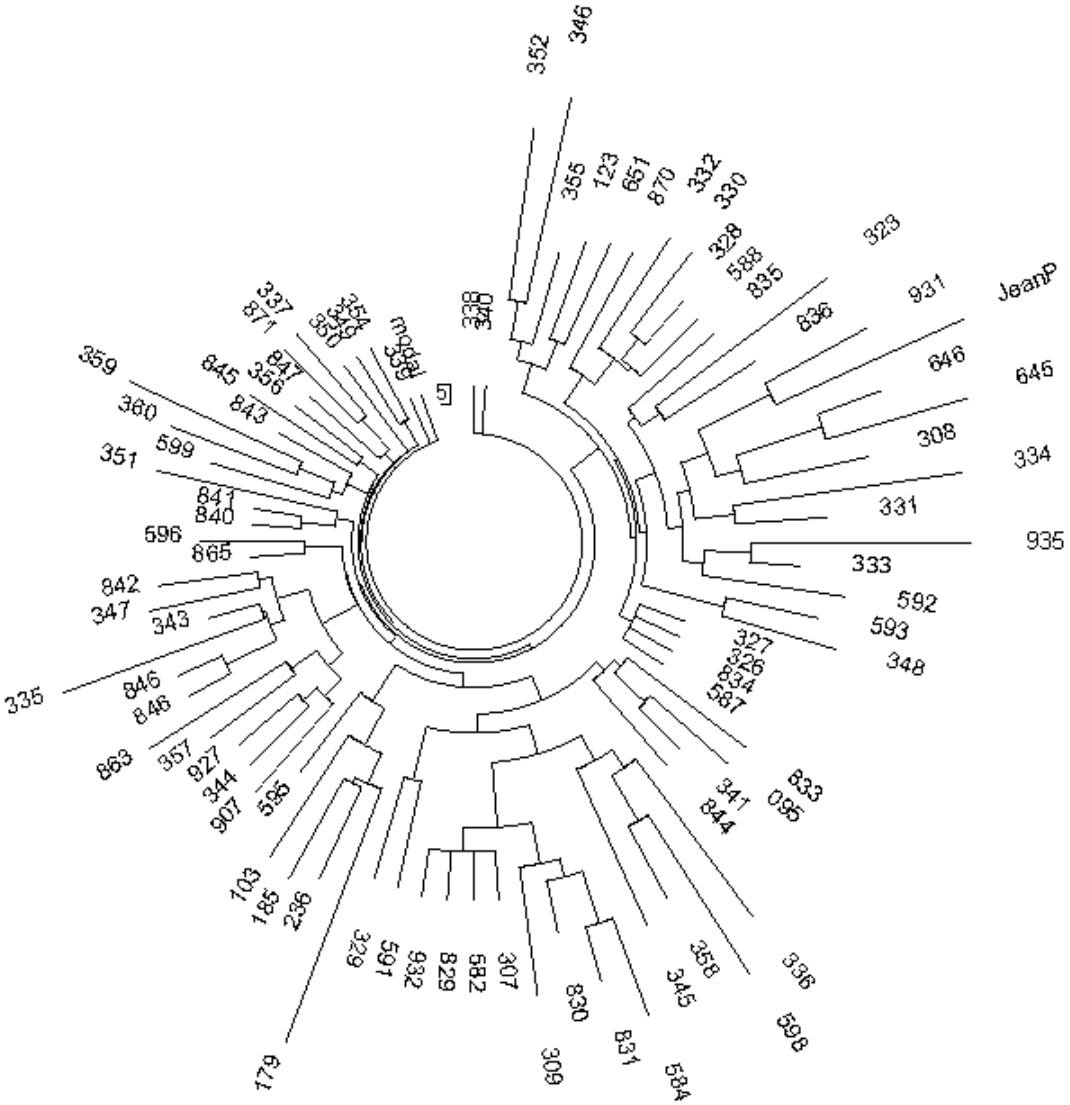
I live in Northern France, my ancestors (since at least 13th century) lived in Antwerp, and I am a participant in the Flanders Project. My haplogroup is R1a1-M17, and my 37 marker haplotype is attached (it is not shown here – AK). R1a1 is a reasonably rare among Flanders, and its share there is merely 3.9% (31 out of 789). For the record, other haplotypes have the following presence among the Flanders:

R1b-M343 (485)	61%
I1-M253 (93)	12%
I2b-M223 (48)	6.1%
E1b1b-M215 (36)	4.6%
R1a-SRY 10831.2 (31)	3.9%
J2a-M410 (26)	3.3%
G2a-P15 (24)	3.0%
I2a-P37.2 (15)	1.9%
J1-M267 (10)	1.3%
J2b-M12 (6)	0.8%
T-M70 (6)	0.8%
L-M11 (4)	0.5%
Q-M242 (4)	0.5%
I2-P215 (1)	0.1%

Looking at my haplotype, can you tell whether I descended from the Vikings? I believe that DYS 389 = 14-31 are located predominantly in Northern Scandinavia.

MY RESPONSE:

Your haplotype belongs to the "Old Scandinavian Branch" of R1a1 haplotypes. Here is a 25-marker haplotype tree of Scandinavian haplotypes, and you can see that your personal haplotype, marked as "JeanP", comfortably sits in the fluffy (that is, ancient) branch on the right-hand side.



The whole tree contains two principal branches – a "young" one on the left, which contains YCAII=19,21, with a common ancestor who lived 1900±400 years before present (Rozhanskii and Klyosov, 2009), and an "old" branch of the right, with YCAII=19,23, and a common ancestor of which lived 4100±700 years before

belong to an ancient lineage of the R1a1 tribe.

Unfortunately, you obtained your haplotype from a company (not the most known FamilyTreeDNA) which uses a different nomenclature for alleles, and I could use only the first 25 markers in your haplotype. Your company uses different numbers for GATA, DYS607, CDYa,b, DYS442, and they are not compatible with the most widely used nomenclature. However, it is not a problem in your particular case, since your branch on the tree is easily recognizable even in the 25-marker format. I had those more detailed trees in 37- and 67-marker formats, and published them last year (see ref. above).

This is the ancestral haplotype for your branch of 4100 years "old":

13 25 15 **11** 11 14 12 12 **10** **13** 11 **30** -- **15** 9 **10** 11 11 **24** 14 20 **32** 12 **15** 15 **16** – 11 12 19 23 16 16 18 19 35 38 13 11 – 11 8 17 17 8 12 10 8 11 10 12 22 22 15 10 12 12 13 8 13 23 21 12 12 11 13 11 11 12 13

In **bold** nine alleles are marked which are mutated in your haplotype in the first 25 markers. There are rather many of those mutations because 4100 years from your ancestor is a long time period. In fact, your haplotype developed more mutations compared with other folks on your branch, because your haplotype is already a half-way to the "Viking" haplotype. I will explain it below.

The history of your lineage for the last ~ 5000 years is as follows. Your branch split from the "Base Haplotype of the Russian Plain", whose common ancestor lived on the Russian Plain around 4800 years before present. He has the following haplotype:

13 25 **16** 11 11 14 12 12 10 13 11 30 -- 15 9 10 11 11 24 14 20 32 12 15 15 16 – 11 11 19 23 16 16 18 19 **34** **39** 13 11 – 11 8 17 17 8 12 10 8 11 10 12 22 22 15 10 12 12 13 8 **14** 23 21 12 12 11 13 11 11 12 13

As you see, it is "older" by only four mutations in all 67 markers. It is translated to 700 years difference between them, hence, the difference between 4800 and 4100 years before present.

Your ancestral "Old Scandinavian" haplotype (4100 ybp) derived "Young Scandinavian" ancestral haplotype 1900±400 years before present. It already had YCAII=19,21. Many Vikings had that haplotype (with some mutations) and brought it to the Isles.

The "Clan Donald" family descended from the "Young Scandinavian" branch around 650 years before present.

The current data show that the "Old Scandinavian" branch was brought to Scandinavia in the middle of the first millennium AD from either the Russian Plain, or from Central Asia, where this lineage was brought by the Aryans on their way from Europe to the East 4500-4000 years ago, and then to India and Iran in about 3500 years before present. Common ancestors of the R1a1 family in India and Iran lived around 4050 years ago, and their ancestral haplotypes are practically identical with that on the ethnic Russians.

LETTER FIFTY-SIXTH

Some days ago I have found and read your work "Proceedings of the Russian Academy of DNA Genealogy ", Volume 1 and maybe I have found there right information I was looking for. My haplogroup is J2 (tested), haplotype is attached. I am trying to understand from where my the paternal line might come from. I was considering the Neolithic farmer option together with the Phoenician, the Arab, the Greek, and the Jewish one considering also the history of the place where I live: South-East Sicily. Since I have an odd value for DYS426=10, that seems to be very rare for J2 haplogroup, I started to search for information, and I found in your work cited above at page 83 the following: "Regarding haplogroup J2, there are only 14 individuals among total of 1,302 with a mutated DYS426. None of them is Jewish.

So my question is: given your results and having DYS426=10 in J2 can I exclude at least the Jewish option? Please note that the Greek is very likely in my opinion. Can I consider this value a proof of a non-Jewish origin? May be I am an Arab? I read that in the ancient Roman empire almost 10% of the total population of about 100 millions were Jews and only this fact in my opinion should suggest that many genetic lineages have been lost forever or entirely migrated to non-Jewish population due to pogroms or conversions into Christian or Muslim religions or even due to Spanish Inquisition, so how can we know for sure who has Jewish ancestors and who has not based only on the modern partial data? For example, with the Spanish Inquisition only for Sicily very many Jews have been erased, partially emigrated and partially converted causing a lost of many lineages in the second case. More or less the same arguments could be valid also for Arabs but they are much more numerous.

MY RESPONSE:

In my opinion, you have started from the wrong end. Forget about "Neolithic farmers" and Phoenicians. You are not there yet.

Please consider three simple rules:

1. Your DYS426=10 means nothing. Yes, it is a very rare mutations. However, it can happen with anyone in any generation. Mutations are random.

Let me give you an example. Mutation in DYS426 happens once in 10,000 generations. It is rare indeed. However, it means that it happens once per every 10,000 births. In a large city (or a country) with 1000 births every day, this mutations happens every 10 days. There are millions of men with such a mutation.

It means, you cannot look at an isolated mutation (even in DYS426) and get any distinctive sense in it. You might be first in your lineage who got it, and your father does not necessarily to have it.

2. 12-marker haplotype does not give you any serious information in terms of your ancestors, and even less in terms of their location. If you are serious, you have to have AT LEAST 37-marker haplotype. Better, if 67-marker one. It will place you on a certain branch in a haplotype tree. This branch will be a collection of your (more or less distant) relatives from all over the world. THEY will give you an idea where your ancestors came from. A single haplotype does not tell a story.

3. You HAVE to determine a deeper clade, rather than just "J2". "J2" tells you almost nothing. J2 are all over the world. A deep clade, such as J2a4h1 will tell you much better where your ancestors came from.

A combination of items 2 and 3 will give you an assignment that you need. After making sure that the branch you belong to belongs to a certain clade, you can calculate (not you, but me) time to your common ancestor for the branch. It can be 500 years, it can be 5,000 years. If it, say, 500 years, what "Neolithic" you are talking about? What "Phoenicians"?

As you see, you do not have practically ANY information to consider as yet.

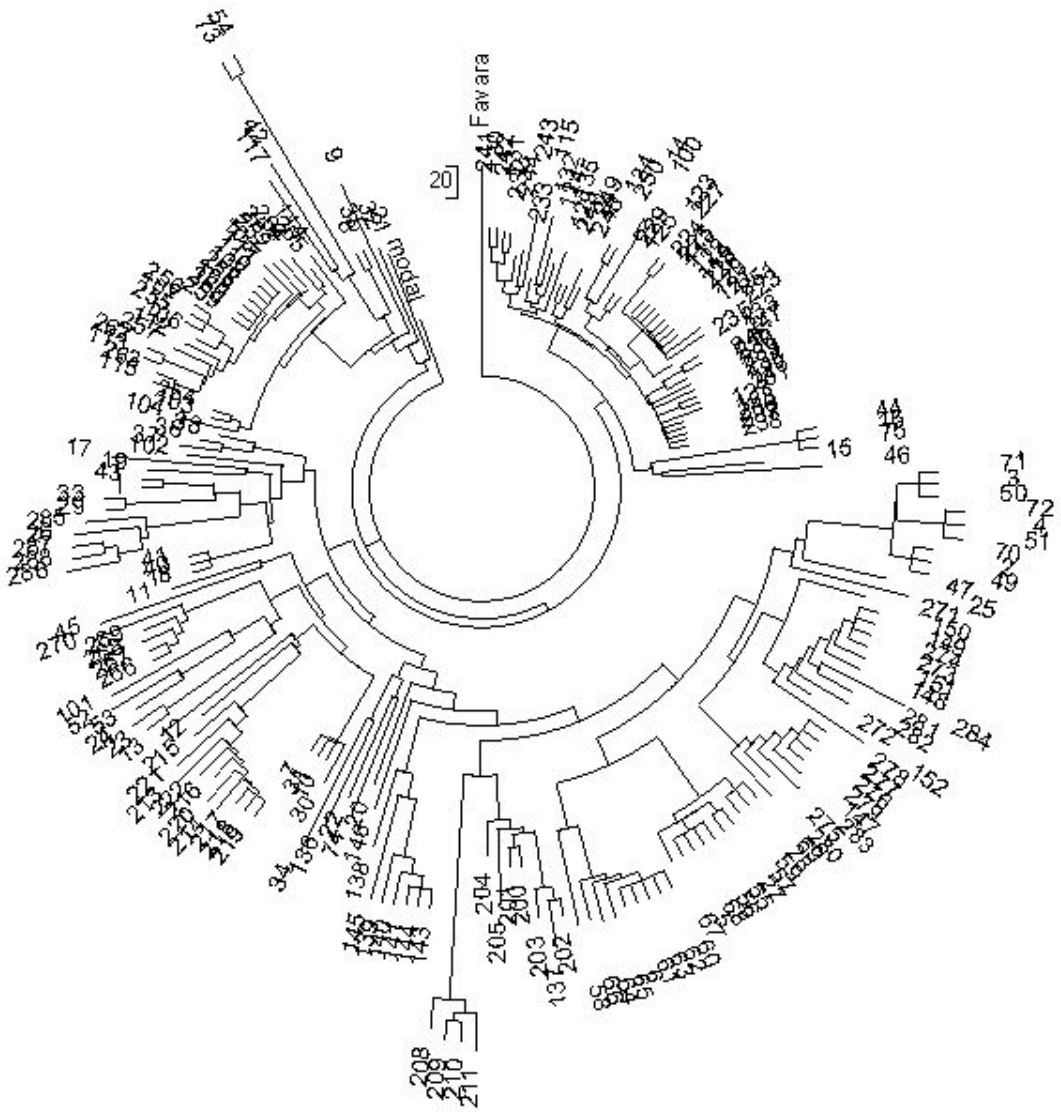
To your direct question regarding the Jews, I do not know any Jews or Arabs who have DYS426=10 in haplogroup J2. There MUST be, but they did not test their DNA as yet. But, as you know now, it means just nothing.

After you accomplish items 2 and 3 done, give me your 37- or (better) 67-marker haplotype, and your deep clade. I will calculate a timespan to your early ancestors and tell where they came from.

I am certainly not in a good position to second-guess your origin from different things which you have meant. However, I can tell that based on 25 marker

haplotypes your personal haplotype does not fit any know lineage of the Jews or the Arabs.

Of course, 67 marker haplotypes would give you the best resolution. Imagine, besides your current 25 markers, there will be 42 more which would place you to a branch you belong to with a much higher certainty.



A 25-marker haplotype tree of the Jews and the Arabs of haplogroup J2. Haplotype numbered below 100 belong to Arabs, of 100 and above belong to the Jews.

Here is an example what can be done with your 25 marker haplotype. If you read the paper

you will see J2a4 (in fact, J2a4b and J2a4b1 branches) on haplotype trees. Those are Jews and Arabs. You might have thought that you have some Jewish or Arabic Y-chromosomal background. See the attached haplotype tree with your name on the top. You do not belong to either Jewish or Arabic lineages. As you see, your haplotype is an "outcast" there. You even climbed on a top of the tree to escape both of them.

A CONTINUATION by the reader:

Your findings are very useful to reject some of the first hypothesis I did, but now I'm getting a little puzzled since some clues in my paternal family seem to point to a "conversion" origin. Now I wonder: "converted from what"? Below some of the clues I considered till now:

- Paternal grandfather kept and hid a book written with an alphabet that was neither Latin nor Greek, the book was later seized by one of the priests of the town after his death (1950) saying that this was not a Christian book and thus dangerous. Was it the Koran? Was it something else?

- Despite to some local traditional recipes in my father's family they did not use animal fats for cooking or they did not cook milk and meat together, considered distasteful and harmful to health.

- My surname is an Arab word and even a name of a town and of some other places (a river, a lake) around here. Even in Sicily usually converted families took names of places.

- Sometimes my father joked about the tradition of using red strings against bad luck or the evil eye.

Meanwhile I'm waiting for my 37-marker haplotype. Do you believe that 67 markers could be really useful, or I'm too far from any meaningful match?

A CONTINUATION by the reader:

I have tested my Deep Clade, and obtained J2a4
(L24- L25- L26+ L27+ M172+ M319- M339- M419- M47- M67- M68- M92- P81-).

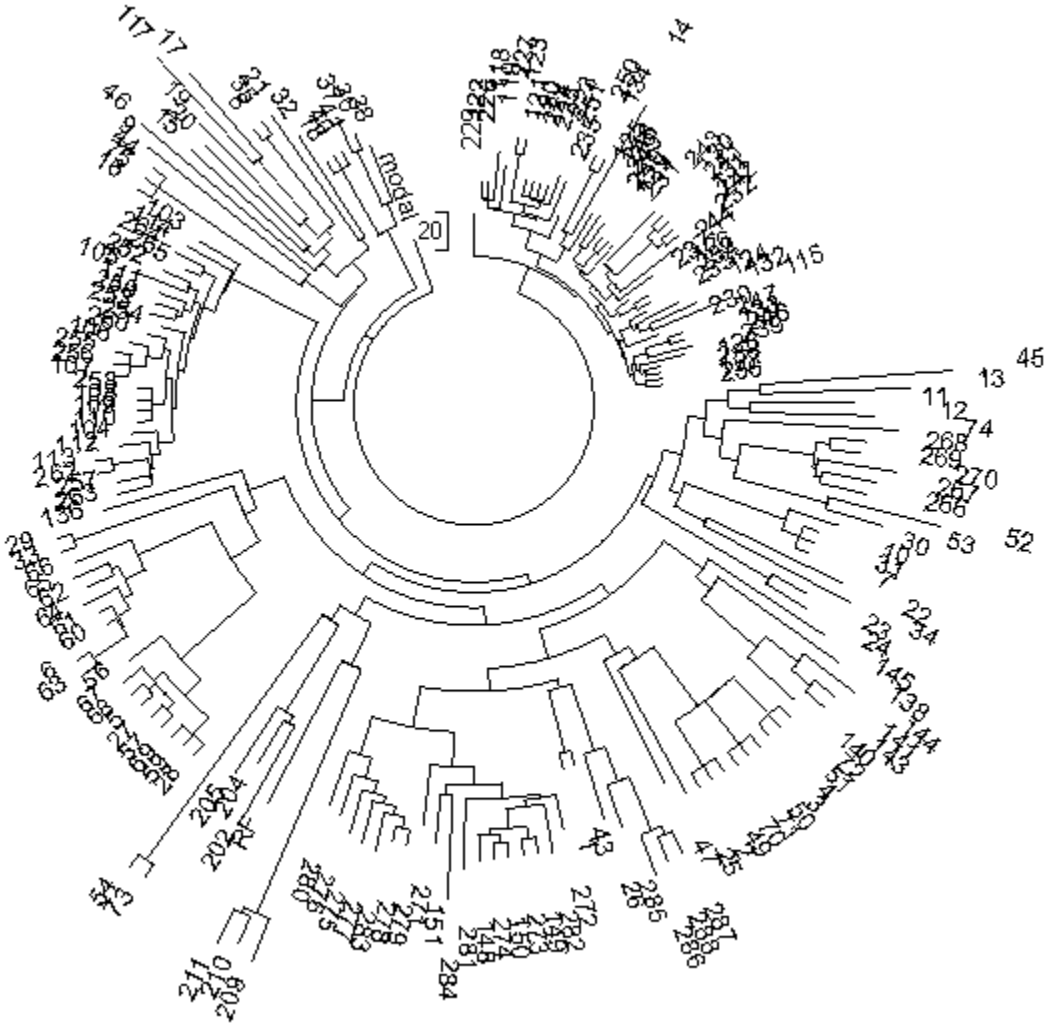
and obtained my 37 marker haplotype (attached).

FTDNA has informed me: "There are currently 186,698 samples in the database. This represents 36, 937 unique twelve marker haplotypes, 68,251 unique twenty-five marker haplotypes, and 78,362 unique thirty-seven marker haplotypes...", that means that they find an average of 1 unique thirty-seven marker haplotypes in each two samples, more or less. This does not seem encouraging to me. I'm starting to think that to have some interesting match I'll have to wait for some years.

MY RESPONSE:

Please forget about the “matches”. For one meaningful match there might be a dozen (or more) absolutely accidental matches, due to a silly statistics. There are thousands and tens of thousands of people in your haplogroup, and there are only a few possible haplotypes with their mutations. Of course, there would be

dozens and hundreds of haplotypes accidentally matching. What can it tell you?



A 37-marker haplotype tree of the Jews and the Arabs of haplogroup J2. Haplotype numbered below 100 belong to Arabs, of 100 and above belong to the Jews. Haplotype marked RF is at 7 o'clock.

Your 37-marker haplotype tree is shown above. You can see a power of 37-marker data compared with 25-marker ones. In the 25-marker haplotype tree above your haplotype was sitting in the middle of nowhere (see the tree in the preceding letter). In the 37-marker tree (see above) the addition of 12 markers made an important change, and your haplotype (marked RF) now sits in the lower left "corner" being surrounded by several Jewish haplotypes and next to two Arabic haplotypes. It is not a very clear branch. Three Jewish haplotypes (209, 210, 211) are almost identical, hence, descended from a very recent common ancestor, who lived 850±300 years before present. The next set of Jewish haplotypes on the other side descended from a common ancestor who lived

However, these two sets descended from a very ancient common ancestor who lived approximately 7225 years before present (30 mutations between their 37-marker haplotypes). Of course, there were no Jews or Arabs those times.

It seems that your ancestors lived on the Mediterranean (where J2 came from) since times immemorial. When you have your 67 marker haplotypes, I can add something to the story.

LETTER FIFTY-SEVENTH

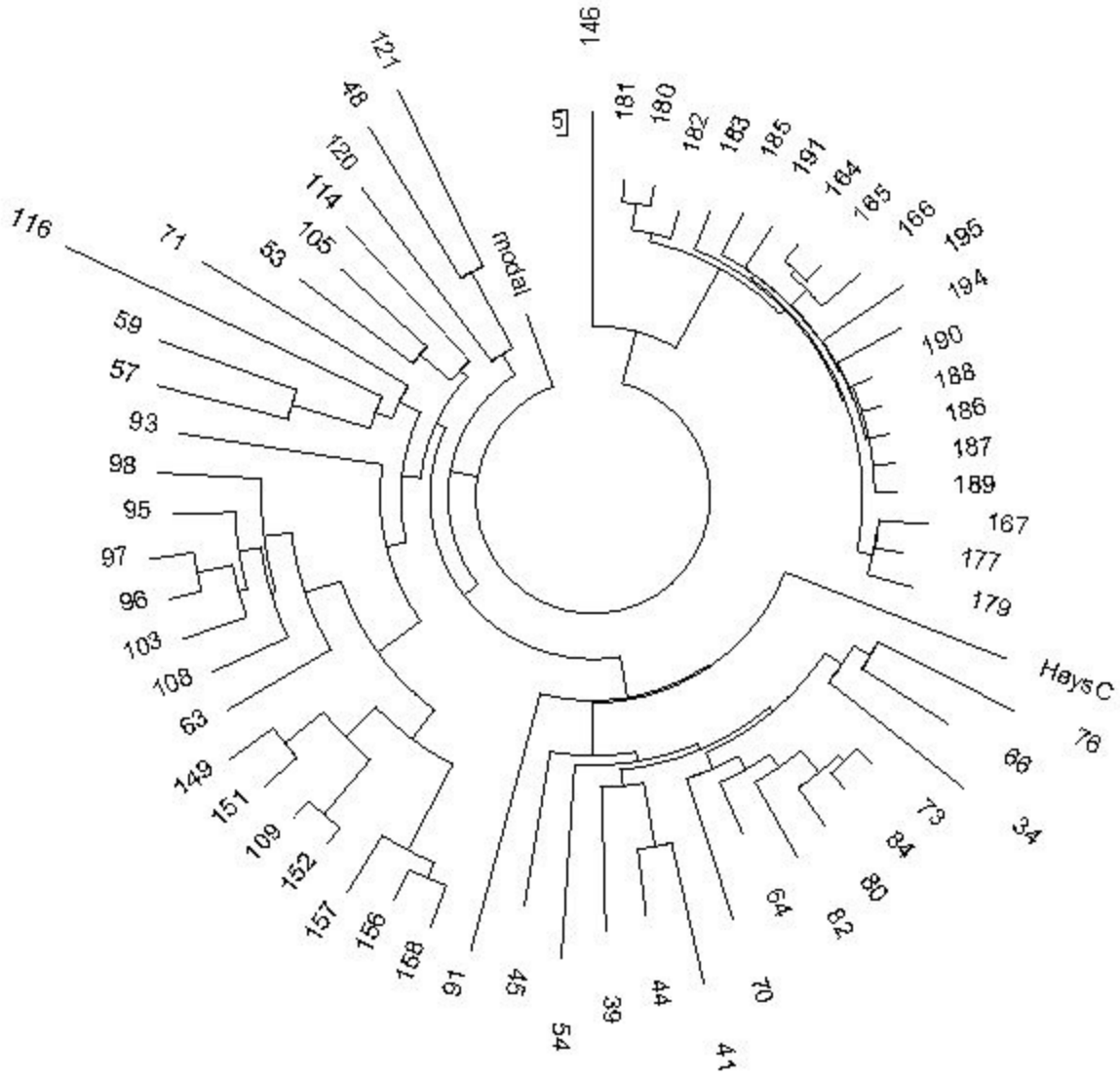
I have a haplotype which you have defined as one of the Tenths in your recent paper, since it contains DYS388=10. However, it begins with "14", which I do not see in your extended list of haplotypes of the Tenths in your publication (Proceedings of the Russian Academy of DNA Genealogy, 2010, June, No.6)

http://www.lulu.com/items/volume_68/8895000/8895749/1/print/8895749.pdf

I'm trying to determine if my ancestry is Irish or Scottish. From what I can tell my DNA is leading more toward Scot since my haplogroup could be Norse. Does that sound correct?

MY RESPONSE:

The 67-marker haplotype tree of the Tenths is shown below. You can see your haplotype in the lower-right-hand side of the tree. As you see, it sits rather lonely, albeit adjacent to the branch which I described in detail in my paper, cited by you. The branch is rather "young", with a common ancestor who lived 1275±180 years before present, that is around the 8th century AD. It is a so-called "young branch of the Tenths", with a common ancestor of 1425±180 years before present, and which was considered in detail from viewpoints of geography and history in my (with I. Rozhanskii) publication (in Russian) in the same journal, but No. 6, 2009. It will be translated to English and published. The paper is more than 100 pages long, and describes all known branches (nearly 20) of R1a1.



There are two principal hypotheses regarding the origin of these haplotypes in the Isles.
(1) Norwegian settlers in Northern parts of Ireland and Scotland, and (2) Norman invaders lead by Wilhelm the Conqueror in 1066. There were many Viking descendents among his troops, who settled in Normandy and the Low Countries by that time.

However, there are more details in your personal haplotype. It differs from the "Young Tenths" base (ancestral) haplotype by 14 mutations in 67 markers, which translates to 2700 years cumulative difference between the common ancestor of the "Young Tenths" and yourself. It is clearly more than 1425±180 years separated him from our times. In other words, a common ancestor of HIM and yourself lived (2700 + 1275)/2 = 2000 years bp, or (2475+1425)/2 = 1950 ybp,

words, it was the beginning of the Common Era.

I do not think there were Scotts or Irish in your roots that time. They were Scandinavians or even members of those tribes who came from the East. You can call them Scythians, or Sarmatians, or protoSlavs, or protoRus, or whoever. It would be the same thing in this context.

A RESPONSE TO ME:

I was reading Samaritan history. "In the Roman provinces, Sarmatian combatants were enlisted in the Roman army, whilst the rest of the population was distributed throughout Thracia, Macedonia and Italy. "

Isn't possible that my ancestor could have become a Roman soldier and was sent to Britain to fight or guard Roman Interests in Britain? Eventually staying and starting a family of his own?

I haven't seen this theory anywhere else.

MY RESPONSE:

Of course it is possible. However, it is as possible as lots of other possible stories. Any theory should have something to be based on. If, for instance, there is any indication that the Sarmatians (or Italians on that matter) have had DYS388=10 haplotypes, this would such a (tentative) base. However, DYS388=10 were not found in the South at all. They are very much localized in the North-West of Europe, albeit similar with haplotypes of the Russian Plain. Clearly, they are from the same barrel.

Aside of it, we know nothing more.

LETTER FIFTY-EIGHTH

I am presently reading your two part article in the JOGG "DNA Genealogy, Mutation Rates, and some Historical Evidence Written in The Y-chromosome".

My grandfather was an R1a1 and from Ireland historically. We have no Idea how long, but family history says we are Irish. I would be curious if you have an idea were his haplotype comes from? To give you more detail, my grandfathers family has been in Georgia, USA for several generations back to the early 1800's maybe late 1700's and before then they were from Ireland. After researching modal haplotypes I see that he match's the Eastern European modal closest, and I

suspect his family must have been from Germany or Poland. We have no family history of this, the family history ends in Ireland.

I have asked our people, and they created a modal from the others that math my grandfather. They told me that I use as a modal an old Ysearch entry, with only 37 markers, constructed by a Scandinavian project, and it has little weight. According to them, my nearest neighbors in YSearch on 67 markers are five individuals from USA and England, origins or others unknown, and the closest distance is 5, then 12, 12, 14, and 15 in 67 marker haplotypes. Therefore, they told me that "there is no reason to think you have recent Slavic ancestry", and "the six of you appear to form a reasonably coherent cluster. When I enter you six into expert Ken Nordtvedt's spreadsheet, you six have an intraclade TMRCA of 65 generations, or about 2000 years. I can only guess that your cluster first formed among one of the peoples that later invaded England: the Angles, Saxons, or Jutes". Also they told me that my haplotype is positioned with respect to the following Polish clusters as:

Vs. J : 93 generations or 2800 years
Vs. A : 109 generations or 3300 years
Vs. I : 110 generations or 3300 years
Vs. G : 112 generations or 3400 years

Eventually they told me that I belong to the "Anglo-Saxon Modal" cluster. It was based on a table of seven haplotypes – one from England, four from the USA and two unknowns.

What would you say?

MY RESPONSE:

Please forget about "modals" and those "clusters" There is no such a thing as a "match" with any particular haplotype. Those are just illusions. Haplotypes mutate, and mutate in "either side" with each of their 67-marker (in your case). That guy's haplotype you compare your haplotype with -also mutates "either side" pretty fast. Each marker in a 67-marker haplotype mutates once - on average - every 200 years. Hence, after only 2000 years you and that fellow you compare with, mutate - each - 10 times. What "match" are you talking about??

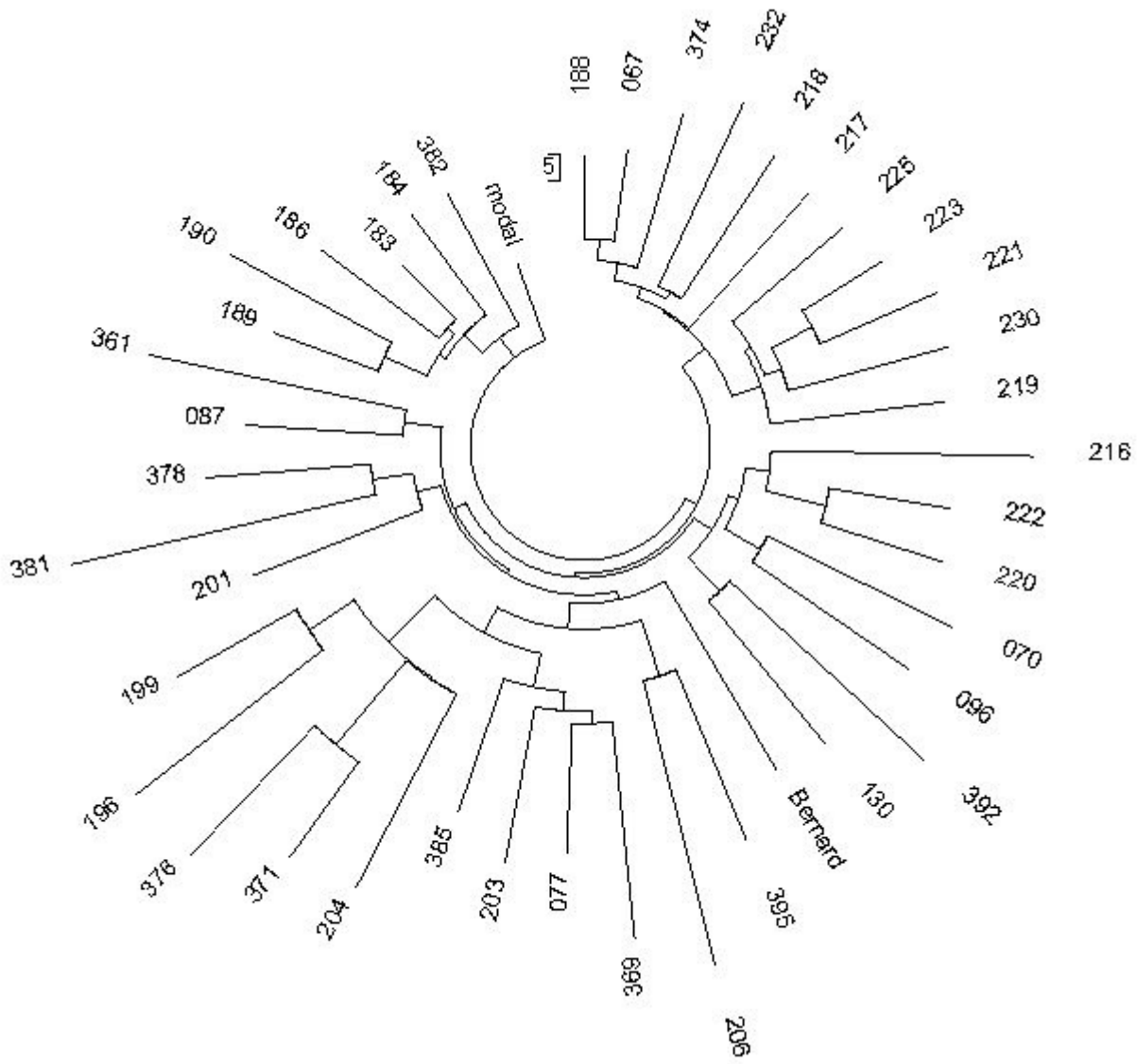
All those "matches" people around are talking about is just a scam. They are, as a rule, just an accidental overlapping. A few haplotypes out of thousands always overlap.

Therefore that first table with “genetic distances” between 5 and 15 does not make sense at all in your situation.

When your adviser compared your haplotype with some mysterious "J" (a Jewish "cluster"??), "A", "I" or "G", it had a bit better sense, since it was a comparison not with individuals, but with "clouds" of haplotypes, with more degrees of freedom. However, the biggest problem with that "Polish project" is that they took only Polish haplotypes, which represent only fragments of the European field of R1a1 haplotypes. The last “Anglo-Saxon” thing is also practically meaningless, since it was again based on “distances” between 5 and 11 for a few people, mainly from the USA.

Back to your actual question. Your haplotype belongs - with a high probability - to the Balto-Carpathian branch of R1a1 haplotypes. You can meet haplotypes of that branch in many parts of Europe, but their most compact locations is indeed in Poland, Lithuania, Carpathian region, and crosses Europe west to Ireland. The map of the distribution of this branch is shown below (Rozhanskii and Klyosov, 2009).

The Balto-Carpathian haplotype tree is also shown below, with your haplotype sitting rather comfortably in it. As you see, its length (which shows a time distance from a common ancestor of the whole tree) is about the same as the average tree haplotypes. The common ancestor of the tree (that is the Balto-Carpathian branch of R1a1) lived 2625+/-290 years before present. This was a mid of the 1st millennium BC, when R1a1 were re-populating Europe, moving from the Russian Plain. Most of European haplotypes and their branches (more than 10 of them) have appeared in Europe around that time.



Your haplotype has a mutational difference from the common ancestor of the BC branch (mind you, not from another guy, but from the common ancestor, which haplotype is firmly established) by 12 mutations per 67 markers. It corresponds to approximately 2275 years, which is indeed within the timeframe of the branch existence. You cannot belong to the Jewish R1a1 branch (J), because your mutational difference from its common ancestor is 23 mutations per 67 marker. It is 4700 years difference. There was no Jews 4700 years ago in the first place, much less in R1a1 haplogroup (it is 1100+/-150 years old).

LETTER NINETY

I am from Finland and I am a chairman of a local ancestral society. We have made DNA tests for five men with the same last name and who live in the same area. Three of them are I1 (one of these three has a deep clade test made and he is I1d P109), one is R1a1a-M198, and one is N1c1-M178.

We here are surprised how come that they all have the same name and lived at the same place but have different DNA lineages. Can you tell us more about this?

Best regards…

MY RESPONSE:

The situation you are describing is a very common one as soon as you start considering haplogroups and haplotypes. Of course, you are familiar with NPE, that is a Non-Paternity Event. On some accounts, it constitutes about 5-10% of all birth events. "Classical" genealogy can only guess or suspect, however, DNA genealogy faces them upfront.

Another explanation is that the bearers of other two (R1a1 and N1c1) lineages came from elsewhere and did not belong to the same family in the first place.

Again, the situation is not unusual. They might all belong to the same family, according to documents and/or church records. However, in fact, they were born to three different fathers, one of them belonged to R1a1, another to N1c, and third to I1 haplogroups.

Let us consider WHEN a nearest common ancestor of the three "I1" folks lived. Luckily, all of them have 67 marker haplotypes. They all three summarily have 7 mutations in three haplotypes. This gives 7/3/0.12 = 19 "conditional" generation ago, that is around 475 years ago. Here "generation" is just a mathematical value equal to 25 years. It is not a real generation. 0.12 is the mutation rate constant for 67 marker haplotypes.

That is, their common ancestor (of haplogroup I1) lived in about 1536, in the 16th century. Since then, their DNA-lineage was evolving, and by now it accumulates as many as 7 mutations in their haplotypes.

CONTINUATION:

Thank you, this cleared something out. We think that all the these I1 and R1a1 ancestors have been born in that area, but N1c1 ancestor moved from elsewhere. We are planning to run more DNA tests.

This is a very interesting information that a common ancestor to all these three I1 persons lived around 1536. I got information from one Finnish genealogist two of them had a common ancestor born at about 1575. He counted those mutations (I don't know the pattern) and told that common ancestor born 390 years back. He counted the time back from the tested person date of the birthday year.

Do you know what separates DYS values to each branch?

MY RESPONSE:

If the N1c1 ancestor had moved to that area from elsewhere, it explains, of course, that he was from a different DNA lineage. However, it does not explain why he has the same last name. People adopt different names, though.

Regarding 1536 or 1575, that is 475 or 390 years back, it is practically the same thing from a point of view of DNA genealogy. I did not bother you with margins of error calculations (though, as you might notice, I calculate them every time in scientific publications), since I have described what those 7 mutations mean. In fact, simple statistics tells us that when you have 7 random mutations, it can easily be 7±2, which means 19±5 "generations", that is 475±125 years back. It means in turn that the common ancestor of the three I1 individuals was born between 600 and 350 years ago. That 390 years back is within the margins of error.

As you see, when a number of tested people is small, and their number of mutations is small too, you will get a large margin of error. Still, it helps to built a general picture, but would not place a common ancestor exactly when he actually lived. It would always have a margin of error. As you see, to calculate, as you put it "He counted the time back from the tested person date of the birthday year", from the birth year or from this year (2011) practically does not make a difference, with a margin of error of 125 years.

> *Do you know what separates DYS values to each branch?*

I think I do know, since I work with branches of various haplogroups. It is because a pattern of mutations, particularly in so-called "slow markers" is involved. Those changed (mutated) haplotypes form a different pattern, which sometimes can be easily seen. In other cases one needs to compose a haplotype tree to see a different branch. European I1 are "young", they are "only" ~3500 years "old", and they do not split into different distinct branches. If they do, they are very "weak" branches, not as in I2, or R1a1, or N on that matter.

Best regards.