Scientific
Research
Publishing

Kilin-Klyosov TMRCA Calculator for Time Spans up to Millions of Years

Anatole A. Klyosov, Vladimir V. Kilin

The Academy of DNA Genealogy, Boston, USA, and Moscow, Russia
Email: aklyosov@comcast.net

Abstract

A TMRCA (Time to the Most Recent Common Ancestor) calculator has been developed, with a capacity to handle up to 10,000 haplotypes simultaneously, for haplotypes being in any format within the 111 markers in the FTDNA (Family Tree DNA, a leading company in systematics of the haplotypes) nomenclature, for haplotypes being in any combination with respect to number of their markers, and for the TMRCA values from a few hundred years to millions of years. The calculator shows the TMRCA data calculated separately and simultaneously in the 6-, 12-, 25-, 37, 67, and 111-marker formats by the linear method, and for haplotypes of any format, such as 7-, 8-, 9-, 10-, 17-, 19-, 23- and any other format by the quadratic method. The calculator also shows a number of mutations (in the whole given dataset of haplotypes), so the TMRCA values can be verified manually, if desired so. The calculator automatically makes corrections for back mutations (in the linear method; there is no need for corrections in the quadratic method), and considers multi-marker mutations and zero alleles, counting them correctly as one mutation. The calculator can be navigated to exclude markers which show an excessive dispersion, which likely is an indication of "admixtures", which do not belong to the given set of haplotypes. The paper provides a number of examples of TMRCA calculations for datasets of different haplogroups, and shows that the mutation rate constants are the same in different haplogroups. The papers provides a comparison of mutation rate tables by Chandler (2006), Ballantyne et al. (2010), Heinila (2012) and an anonymous investigator (2014) with the mutation rate constants determined and examined in this study. It is shown that the above authors noticeably and significantly overestimated their mutation rates, which often lead to unrealistic TMRCAs.

Keywords

Haplogroups, Haplotypes, DNA Genealogy, TMRCA, Mutation Rates

1. Introduction

The TMRCA values (Time for the Most Recent Common Ancestor) provide valuable data for DNA genealogy.

In one common approach, they are calculated by counting a number of mutations in a given set of haplotypes (in the same format, such as 12 marker haplotypes, or 17 marker, or 25, 37, 67, 111 marker haplotypes, or any other format) from a so-called base haplotype, which is assumed to be the ancestral haplotype for the haplotype set. The total number of mutations is divided by the number of haplotypes in the set and by the mutation rate constant. For example, if a set of one hundred of 67 marker haplotypes shows 250 mutations total, then we have $250/100/0.12 = 20.83$ conditional generations (25 years each), that is typically rounded to 21 conditional generations (since a generation is a discrete figure), and get 525 ± 60 years to a common ancestor. Here 0.12 (a number of mutations per conditional generation, calibrated for 25 years in the generation) is the mutation rate constant for the 67 marker haplotype, and the margin of error is calculated following common rules of statistics (Klyosov, 2009a; Klyosov & Rozhanskii, 2012a). If the number of 20.83 is not rounded, one obtains the TMRCA of 521 ± 60 years, that is practically the same number within the margin of error. The method described above was coined the linear method. If one does not like 25 years as a conditional generation and wants to employ, say, 30 years, then the mutation rate constant is 0.144 (mutations per conditional generation of 30 years), and the TMRCA is exactly the same: $250/100/0.144 = 17.36$ conditional generations, that is $17.36 \times 30 = 521 \pm 60$ years. Obviously, the linear method is applicable only to a set of haplotypes in the same format, since a number of mutations is divided by a number of haplotypes.

Another approach, coined the quadratic method (Klyosov, 2009a), or ASD (average square distance, Goldstein et al., 1995), can be applied to haplotypes in various formats in the same dataset, and does not require a correction for back mutations. However, it is too tedious for manual calculations, since it requires calculations of average square distances, as its name indicates, pair wise between all alleles for each marker in the dataset. There is also a logarithmic method for the TMRCA calculations (Klyosov, 2009a) which we will consider very briefly in this paper.

Generally, a high level of skills for the TMRCA calculations is required to be correct, as much as this term is applicable to statistical mathematical operations. There are seven major difficulties confronting those who think it is easy. **First**, the dataset (a series of haplotypes) should be subdivided to separate lineages, or branches, each of them had their common ancestor; otherwise, a "phantom" common ancestor, which is a product of various branches, will be "dated". **Second**, one should know the mutation rate constants for the haplotype dataset chosen (or available) for calculations. For the linear method the cumulative rate constants should be known (that is for the 12-, 17-, 25-, 37-, 67-, 111-marker panels, or for any other haplotype format). For the quadratic method in which calculations are conducted along each of the marker, the individual mutation rate constants should be known, ideally for each of the 111 markers. Those (both the cumulative and individual mutation rate constants) are largely unknown in the scientific community, or questionable and highly debated. The name "constants" is employed not because all of them are identical across all the markers, but because the mutation rate constant for any given marker remains the same for different haplogroups, in the course of mutations in the same lineage, and does not depend on a size of the dataset. An analogue is radioactive decay in the course of which the half-life time (hence, the decay rate constant) remains the same and does not depend on a size/weight of the sample.

Third, the generation length should be settled, otherwise the TMRCA values remain to be uncertain. Assumptions do not help there, such as typical "assuming a generation time of 30 years" (Poznik et al., 2016, as an example of the recent publication). **Fourth**, corrections for back mutations should be introduced when the linear (and the logarithmic) method is employed, and they are largely unknown for population geneticists. **Fifth**, it is not easy to count mutations in haplotypes, since a number of markers are multi-copy, and they mutate in pairs (DYS385, DYS459, YCAII, DYS413) or quadruples (DYS464). **Sixths**, some alleles reproducibly show zero values, and many users do not know how to count them. **Seventh**, compared with a few years ago, datasets now often contain many hundreds and thousands of haplotypes, which makes it practically impossible for manual counting of mutations.

The first four difficulties have been resolved in our studies published earlier (Klyosov, 2009a, 2009b, 2009c; Rozhanskii & Klyosov, 2011; Klyosov et al., 2012, 2012a, 2012b), exemplified with manual calculations. Fifth and sixths difficulties have been generally resolved in the literature, though very few users employ multi markers and zero alleles in TMRCA calculations. In this publication, we resolve that and the last complication, by offering an automatic calculation method as a multifunctional calculator with a capacity up to 10,000 haplotypes in a format from a single marker to the 111 marker haplotypes, and calculations are conducted simultaneously by the linear method in the 6-, 12-, 17-, 25, 37-, 67-, and 111-marker format (the FTDNA nomenclature) along with the 22 "slow markers" selected from the 67 marker panel, as well as by the quadratic method in the any

format up to the 111 marker haplotypes along with the 22 "slow marker" haplotypes. Therefore, at least ten TMRCAs could be obtained simultaneously at the same display along with margins of error for each of them, and reliability of the TMRCAs can be compared in real, practical terms.

2. Kilin-Klyosov Calculator

2.1. Description of the Calculator

The Calculator can be downloaded from either of two locations: http://www.anatole-klyosov.com/ (in the section "DNA Genealogy TMRCA Calculator"), or http://dna-academy.ru/kilin-klyosov/ (the first textual link from the bottom). Both downloads are rather slow, due to a size of the file (37 Mb).

The calculator is multifunctional, and runs concurrently two methods, quadratic and linear.

The first one shows the TMRCAs in the first two columns, indexed as KKK111 and KKK22. KKK22 is essentially a part of the KKK111, however, it is based on 22 slowest markers, listed below. The KKK22 method is applicable only to the 22 marker haplotypes. For all other marker datasets the KKK22 column should be ignored. KKK111 is applicable to any haplotype format, and haplotypes can be totally assorted with respect to their set of markers and length of haplotypes. Calculations in the KKK (quadratic) approach are based on the discrete random walk numerical model, known in mathematics.

The linear method is employed for 6, 12, 17, 25, 37, 67, and 111 marker haplotypes, as well as for the slow 22 marker haplotypes. All haplotypes in the dataset should have the same format, since the method employs the specific mutation rate constants for specific length of the haplotypes, as follows (the first column shows a number of markers in the haplotype according the FTDNA nomenclature, the second column shows the mutation rate constant, in mutations per a conditional generation of 25 years):

6 marker	0.0074
12	0.0200
17	0.0365 (with DYS385)
25	0.0460
37	0.0900
67	0.1200
111	0.1980
22 slow markers	0.0054

In the linear method employing "non-standard" haplotypes, such as 5-, 7-, 8-, 9-, 10-, 17 (with DYS426 and DYS388), 18-, 19-, 23-, 39-, 43-marker, as well as any others, the cumulative mutation rate constant can be calculated by summing up the individual mutation rate constants in the calculator (line 7), or in **Figure 1** below, and a number of mutations in the haplotype data set can be seen and summed up in the "Abs Deviation" line

DYS393	DYS390	DYS19	DYS391	DYS385a	DYS385b	DYS426	DYS388	DYS439	DYS389i	DYS392	DYS389ii		
0.00059	0.0022	0.00179	0.0022	0.0028	0.0036	0.00009	0.00022	0.003	0.0008	0.0004	0.00231		0.02

DYS458	DYS459a	DYS459b	DYS455	DYS454	DYS447	DYS437	DYS448	DYS449	DYS464a	DYS464b	DYS464c	DYS464d		
0.0062	0.0004	0.00096	0.0001	0.0003	0.0032	0.00064	0.0014	0.0068	0.001	0.0011	0.0025	0.0014		0.046

DYS460	Y-GATA-H	YCA11a	YCA11b	DYS456	DYS607	DYS576	DYS570	CDYa	CDYb	DYS442	DYS438		
0.002	0.0017	0.0021	0.00233	0.00432	0.0026	0.006	0.0047	0.0066	0.007	0.0043	0.00035		0.09

DYS531	DYS578	DYF395S1	DYF395S1	DYS590	DYS537	DYS641	DYS472	DYF406S1	DYS511	DYS425	DYS413a	DYS413b	DYS557	DYS594
0.0009	0.00043	0.0004	0.0004	0.00034	0.0009	0.00017	0.000008	0.0009	0.0009	0.000042	0.002	0.002	0.0026	0.00017

DYS436	DYS490	DYS534	DYS450	DYS444	DYS481	DYS520	DYS446	DYS617	DYS568	DYS487	DYS572	DYS640	DYS492	DYS565
0.00004	0.00007	0.00315	0.00011	0.0018	0.004	0.0015	0.0028	0.0005	0.00005	0.0007	0.00092	0.00015	0.00015	0.0019

DYS710	DYS485	DYS632	DYS495	DYS540	DYS714	DYS716	DYS717	DYS505	DYS556	DYS549	DYS589	DYS522	DYS494	DYS533
0.0073	0.0016	0.0001	0.0014	0.0013	0.0045	0.0005	0.0006	0.003	0.0008	0.002	0.00068	0.0016	0.0001	0.00125

DYS636	DYS575	DYS638	DYS462	DYS452	DYS445	Y-GATA-A	DYS463	DYS441	Y-GGAAT-	DYS525	DYS712	DYS593	DYS650	DYS532
0.00023	0.0003	0.00027	0.0003	0.001	0.0005	0.0032	0.0025	0.0021	0.001	0.0017	0.0062	0.00012	0.0037	0.0035

DYS715	DYS504	DYS513	DYS561	DYS552	DYS726	DYS635	DYS587	DYS643	DYS497	DYS510	DYS434	DYS461	DYS435	
0.0032	0.003	0.0015	0.0013	0.0036	0.00017	0.0022	0.0012	0.00257	0.0011	0.003	0.0003	0.0014	0.00011	0.198

Figure 1. A list of the individual mutation rate constants for the 111 markers, arranged in the FTDNA format. The numbers in the last column, 0.02, 0.046, 0.09, and 0.198 show the cumulative mutation rate constants for the 12-, 25-, 37-, and 111-marker haplotypes. The numbers for the intermediate 17- and 67-marker haplotypes are shown in text.

(line 13) in the Calculator. In the quadratic method all said haplotype formats can be employed without any restrictions. The mutation rate constants for all the 111 markers are shown in **Figure 1**.

The calculator handles all seven difficulties named above, some of them directly, other indirectly. Regarding the first item (in the list of difficulties), it is preferred that branches (lineages, subclades) to be identified before the calculator is employed, for example by composing a haplotype tree, see, for example, **Figure 2** and **Figure 3** for haplogroup R1b-M269, **Figure 4** and **Figure 5** for haplogroup I1-M253, **Figure 6** for haplogroup I2-M438,
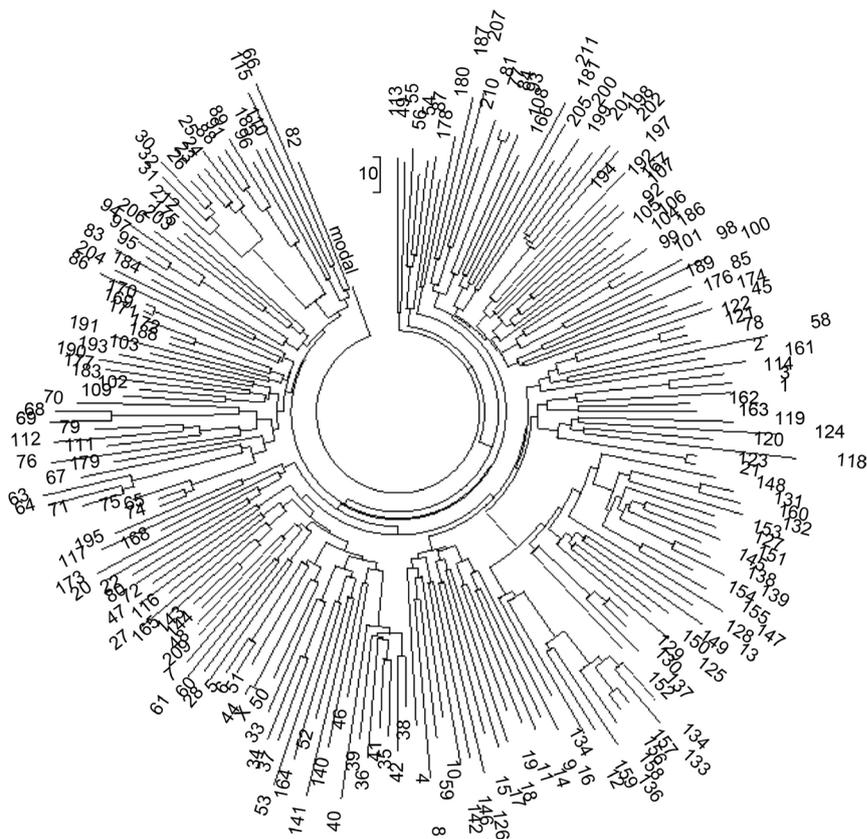
Figure 2. A haplotype tree for R1b-M269 dataset of 204 haplotypes in the 111 marker format. The dataset includes downstream subclades L150, L51, P310, P311, CTS4528, CTS7822, CTS9216. The haplotypes are listed in the hp 35 FTDNA database. The numbers at bars representing haplotype on the tree, as well as on all other trees in this paper, do not bear any particular meaning in this context, and the trees provide here just illustrations with respect to their general appearance.

		KKK111	KKK22	LM6	LM12	LM17	LM25	LM37	LM67	LM111	LM22	
TMRCA		5454	6650	5339	4974	4620	5126	4687	4746	4583	5870	
standard deviation KKK		371	1350	3095	2367	1140	1216	863	678	458	1709	
standard deviation LM				620	532	482	528	477	481	462	694	
		(1+EXP(CYM))	17	2.6	100							
STR	age of haplotype	DYS393	DYS390	DYS19	DYS391	DYS385a	DYS385b	DYS426	DYS388	DYS439	DYS389i
P=1/k(mutation rate constant)		0.00059	0.0022	0.00179	0.0022	0.0028	0.0036	0.00009	0.00022	0.003	0.0008
P for LM		0.00059	0.0022	0.00179	0.0022	0.0028	0.0036	0.00009	0.00022	0.003	0.0008
Variance/P		492	159	73	119	81	192		156	173	227
AverageValueSTR		12	24	14	11	11	14	12	12	12	13
AverageAbsDeviation/p		823	136	52	136	96	159	1089	111	147	227
AverageAbsDeviation		0.49	0.30	0.09	0.30	0.27	0.57	0.10	0.02	0.44	0.18
AbsDeviation		99	61	19	61	55	117	20	5	90	37
		1.07	-0.40	-0.78	-0.58	-0.75	-0.26	4.19	-0.42	-0.34	-0.10
		1.97	0.64	0.29	0.48	0.33	0.77	4.79	0.62	0.70	0.91
haplotype 1		12	24	14	10	11	14	11	12	12	13
haplotype 2		12	24	14	11	11	14	11	12	12	13
haplotype 3		12	25	14	10	11	14	11	12	13	13
haplotype 4		12	24	15	10	11	14	12	12	12	14

Figure 3. A screenshot of the TMRCA Calculator for the series haplotypes of downstream R1b-M269 subclades, with four haplotypes are partially shown out of 204 in the whole series, with the first 11 markers out of the 111 total.
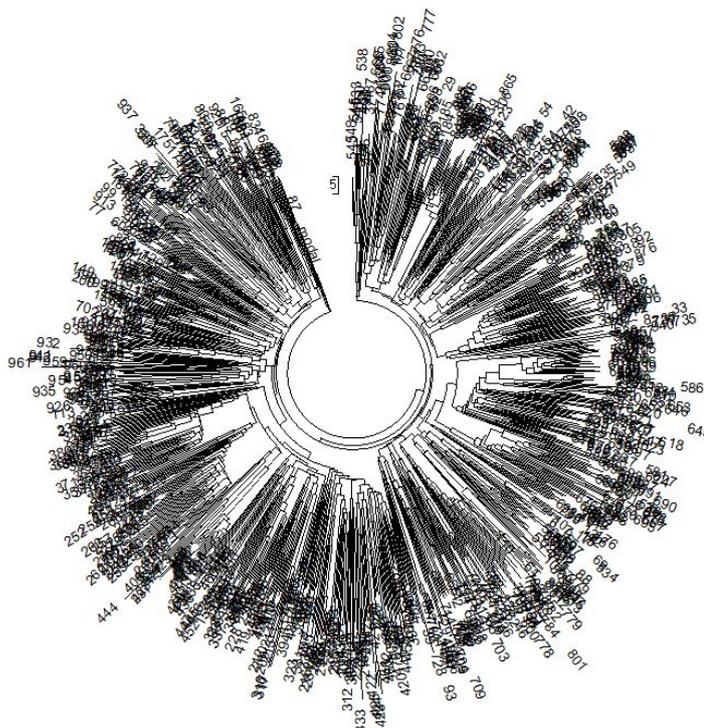
Figure 4. A haplotype tree for I1-M253 dataset of 968 haplotypes in the 111 marker format. The haplotypes are listed in the FTDNA haplogroup I1 database.

		KKK111	KKK22	LM6	LM12	LM17	LM25	LM37	LM67	LM111	LM22			
TMRCA		3549	3136	4167	3752	3367	3593	3469	3618	3686	3088			
standard deviation KKK		234	642	2227	1174	358	637	462	452	372	1238			
standard deviation LM				436	382	341	362	349	363	369	332			
		(1+EXP(CYM))	17	2.6	100									
STR	age of haplotype	DYS393	DYS390	DYS19	DYS391	DYS385a	DYS385b	DYS426	DYS388	DYS439	DYS389i	DSY392	DSY389ii
P=1/k(mutation rate constant)		0.00059	0.0022	0.00179	0.0022	0.0028	0.0036	0.00009	0.00022	0.003	0.0008	0.0004	0.00231
P for LM		0.00059	0.0022	0.00179	0.0022	0.0028	0.0036	0.00009	0.00022	0.003	0.0008	0.0004	0.00231
Variance/P		243	130	153	50	124	94	46		116	151	195	162
AverageValueSTR		13	22	14	10	13	14	11	14	11	12	11	28
AverageAbsDeviation/p		214	182	152	53	165	105	46	667	96	136	145	136
AverageAbsDeviation		0.13	0.40	0.27	0.12	0.46	0.38	0.00	0.15	0.29	0.11	0.06	0.31
AbsDeviation		122	388	263	112	446	365	4	142	278	105	56	303
		0.29	-0.25	-0.14	-0.63	-0.28	-0.42	-0.65	4.38	-0.32	-0.15	0.06	-0.10
		1.34	0.71	0.84	0.27	0.68	0.52	0.25	6.03	0.64	0.83	1.07	0.89
haplotype 964		15	23	15	10	13	14	11	14	11	13	11	29
haplotype 965		13	22	14	10	13	14	11	15	11	12	11	28
haplotype 966		13	23	14	10	13	14	11	14	11	12	11	28
haplotype 967		13	23	14	10	13	14	11	14	11	12	11	29
haplotype 968		13	23	14	10	13	14	11	14	10	12	11	29

Figure 5. A screenshot of the TMRCA Calculator for the series of I1-M253 haplotypes, with five last haplotypes are partially shown out of 968 in the series, with the first 11 markers out of the 111 total.

**Figures 7-9** for haplogroup J1-M267, compared to **Figure 10** and **Figure 11** for haplogroup N1c1-M46, and **Figure 12** and **Figure 13** for haplogroup R1b-U106, and paste haplotypes in the calculator for different branches separately (examples are given below). Another way to detect heterogeneity of a dataset is to employ the linear and the logarithmic method concurrently, and to compare their TMRCAs. If the TMRCAs are different for the two methods, the dataset is non-uniform, and the TMRCA is a "phantom". However, the calculator can point to a non-uniformity of the dataset, if the TMRCAs calculated from haplotypes of a various lengths are "jumpy", and vary for different set of markers. Ideally, it should be the same for all columns in the calculator display.

Regarding the second item, the calculator provides all 111 mutation rate constants for automatic and manual (if needed) calculations. Regarding the third item above, the calculator automatically counts a number of mutations for each panel of haplotypes; it needs just to highlight a sequence of mutations for the given haplotype format and read the cumulative number at the bottom of the computer screen. Regarding the fourth item, the

Figure 6. A haplotype tree for I2-M438 dataset of 244 haplotypes in the 111 marker format**.** The haplotypes are listed in the FTDNA haplogroup I2 database. The tree consists of at least three branches, containing 157 haplotypes (the upper part), 52 haplotypes (at the bottom right), and 35 haplotypes (at the bottom left).
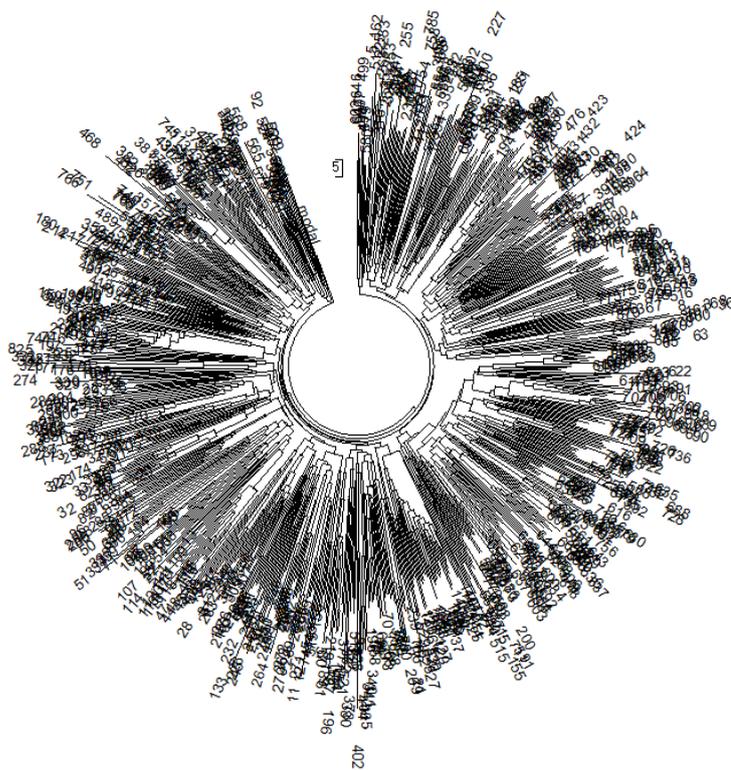

Figure 7. A haplotype tree for J1-M267 dataset of 739 haplotypes in the 111 marker format**.** The haplotypes are listed in the FTDNA haplogroup J1 database. The tree consists of three major branches, containing 74 haplotypes (at the bottom left), 417 haplotypes (right-hand side), and 248 haplotypes (left-hand side).

	KKK111	KKK22	LM6	LM12	LM17	LM25	LM37	LM67	LM111	LM22
TMRCA	2381	1807	2294	2946	2424	2555	2480	2509	2479	1909
standard deviation KKK	199	540	2726	1369	402	686	476	427	279	1115
standard deviation LM			269	311	251	262	252	254	250	241

STR	age of haplotype	DYS393	DYS390	DYS19	DYS391	DYS385a	DYS385b	DYS426	DYS388	DYS439	DYS389i
	(1+EXP(CYM))	17	2.6	100							
P=1/k(mutation rate constant)		0.00059	0.0022	0.00179	0.0022	0.0028	0.0036	0.00009	0.00022	0.003	0.0008
P for LM		0.00059	0.0022	0.00179	0.0022	0.0028	0.0036	0.00009	0.00022	0.003	0.0008
Variance/P		401	22	15	120	74	206	27		71	105
AverageValueSTR		12	23	14	11	13	19	11	17	11	13
AverageAbsDeviation/p		236	22	15	124	61	187	27	698	74	96
AverageAbsDeviation		0.14	0.05	0.03	0.27	0.17	0.67	0.00	0.15	0.22	0.08
AbsDeviation		58	20	11	114	71	280	1	64	92	32
		2.01	-0.64	-0.69	0.05	-0.27	0.65	-0.60	7.31	-0.29	-0.06
		3.55	0.19	0.13	1.06	0.66	1.82	0.24	10.25	0.63	0.93
haplotype 1		12	23	14	11	13	18	11	17	11	13
haplotype 2		11	24	14	13	13	17	11	14	11	13
haplotype 3		12	23	14	11	13	19	11	17	12	13
haplotype 4		12	24	15	11	13	19	11	17	12	13
haplotype 5		13	23	14	11	13	19	11	17	12	13

Figure 8. A screenshot of the TMRCA Calculator for the series of 417 haplotypes of J1-M267 branch located on the right-hand side of the haplotype tree in **Figure 7**, with five first haplotypes are partially shown out of 739 total in the series, with the first 10 markers out of the 111 total.

	KKK111	KKK22	LM6	LM12	LM17	LM25	LM37	LM67	LM111	LM22
TMRCA	4116	4091	3701	4594	3670	4144	3885	4019	3849	5968
standard deviation KKK	339	1318	8250	4077	423	2083	1426	1018	635	2871
standard deviation LM			438	488	383	427	395	407	388	686

STR	age of haplotype	DYS393	DYS390	DYS19	DYS391	DYS385a	DYS385b	DYS426	DYS388	DYS439	DYS389i	DSY392	DSY389ii
	(1+EXP(CYM))	17	2.6	100									
P=1/k(mutation rate constant)		0.00059	0.0022	0.00179	0.0022	0.0028	0.0036	0.00009	0.00022	0.003	0.0008	0.0004	0.00231
P for LM		0.00059	0.0022	0.00179	0.0022	0.0028	0.0036	0.00009	0.00022	0.003	0.0008	0.0004	0.00231
Variance/P		92	92	76	91	128	390	0		142	181	40	285
AverageValueSTR		12	23	14	10	13	18	11	16	11	13	11	30
AverageAbsDeviation/p		96	94	70	97	95	240	0	2053	120	176	40	241
AverageAbsDeviation		0.06	0.14	0.13	0.21	0.27	0.86	0.00	0.45	0.36	0.14	0.02	0.56
AbsDeviation		14	35	31	53	66	214	0	112	89	35	4	138
		-0.42	-0.42	-0.48	-0.42	-0.29	0.67	-0.76	7.09	-0.24	-0.09	-0.61	0.29
		0.45	0.45	0.37	0.44	0.62	1.89	0.00	10.36	0.69	0.88	0.20	1.38
haplotype 1		12	23	14	10	13	15	11	16	11	13	11	29
haplotype 2		12	23	14	10	13	15	11	16	11	13	11	29
haplotype 3		12	23	14	10	13	15	11	16	11	13	11	29
haplotype 4		12	24	14	10	14	18	11	16	12	13	11	29
haplotype 5		13	23	15	10	12	19	11	18	11	14	11	30

Figure 9. A screenshot of the TMRCA Calculator for the series of 248 haplotypes of J1-M267 branch located on the left-hand side of the haplotype tree in **Figure 7**, with five first haplotypes are partially shown out of 739 total in the series, with the first 12 markers out of the 111 total.
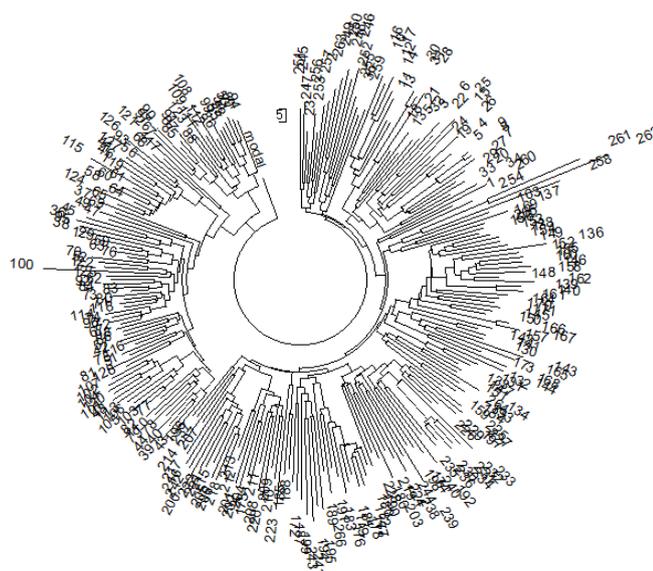
Figure 10. A haplotype tree for N1c1-M46 dataset of 275 haplotypes in the 111 marker format. Three outliers on the right were removed from calculations, since they belong to a different (Siberia-Altayan) subclade. The haplotypes are listed in the FTDNA haplogroup N1c1 database.

	KKK111	KKK22	LM6	LM12	LM17	LM25	LM37	LM67	LM111	LM22			
TMRCA	3086	2546	3951	2924	2735	4127	3461	3287	3233	2546			
standard deviation KKK	271	546	502	743	588	1263	889	581	386	921			
standard deviation LM			458	317	287	424	352	333	326	330			
		(1+EXP(CYM))	17	2.6	100								
STR	age of haplotype	DYS393	DYS390	DYS19	DYS391	DYS385a	DYS385b	DYS426	DYS388	DYS439	DYS389i	DSY392	DSY389ii
P=1/k(mutation rate constant)		0.00059	0.0022	0.00179	0.0022	0.0028	0.0036	0.00009	0.00022	0.003	0.0008	0.0004	0.00231
P for LM		0.00059	0.0022	0.00179	0.0022	0.0028	0.0036	0.00009	0.00022	0.003	0.0008	0.0004	0.00231
Variance/P			174	130	117	90	30	39	0	115	57	310	238	140
AverageValueSTR			14	23	14	11	11	13	11	12	10	14	14	30
AverageAbsDeviation/p			149	186	139	98	27	40	0	83	61	380	211	155
AverageAbsDeviation			0.09	0.41	0.25	0.22	0.08	0.14	0.00	0.02	0.18	0.30	0.08	0.36
AbsDeviation			24	112	68	59	21	39	0	5	50	83	23	98
			0.17	-0.07	-0.14	-0.28	-0.60	-0.55	-0.76	-0.14	-0.45	0.89	0.51	-0.01
			1.22	0.91	0.82	0.63	0.21	0.27	0.00	0.81	0.40	2.17	1.67	0.98
haplotype 1		13	23	12	11	11	13	11	12	10	14	14	30
haplotype 2		13	23	14	10	11	13	11	12	10	14	14	30
haplotype 3		13	23	14	11	11	13	11	12	10	13	14	29
haplotype 4		13	23	14	11	11	13	11	12	10	13	14	29
haplotype 5		13	23	14	11	11	13	11	12	10	13	14	29

Figure 11. A screenshot of the TMRCA Calculator for the series of 275 haplotypes of N1c1-M46 haplotype tree in the 111 marker format, with five first haplotypes are partially shown, with the first 12 markers out of the 111 total.

Figure 12. A haplotype tree for R1b-U106 dataset of 829 haplotypes in the 111 marker format. The haplotypes are listed in the FTDNA haplogroup U106 database.

calculator does not "assume" a length of generation, it employs the conditional generation of 25 years, since the mutation rate constants are adjusted to this generation length, as explained in the Introduction. Regarding the fifth item, in both methods, quadratic and linear, a correction for back mutations is imbedded in the calculator (as it was noticed above, the correction is not required in the quadratic method). Regarding the sixth item, the Calculator counts each multimarker mutation and each zero mutation as a single mutation. Finally, regarding the seventh item, the calculator operates with a number of haplotypes up to 10 thousand, and calculations typically are completed in a few seconds.

If haplotypes are presented in the 12 marker format, the TMRCA will be identical in windows for 12-, 25-, 37-, 67-, and 111-marker haplotypes. In the 17 marker window the TMRCA will be different and, clearly, incorrect, since the 12 marker haplotype misses a number of markers for the 17 marker.

The first two short lines, shown in the downloaded calculator, are for guidance only, they show where to paste

	KKK111	KKK22	LM6	LM12	LM17	LM25	LM37	LM67	LM111	LM22			
TMRCA	4109	4631	4247	3577	3701	4166	3958	3780	3584	4092			
standard deviation KKK	260	784	694	468	421	474	350	1156	711	3413			
standard deviation LM			447	366	375	420	398	379	359	437			
		(1+EXP(CYM))	17	2.6	100								
STR	age of haplotype	DYS393	DYS390	DYS19	DYS391	DYS385a	DYS385b	DYS426	DYS388	DYS439	DYS389i	DSY392	DSY389ii
P=1/k(mutation rate constant)		0.00059	0.0022	0.00179	0.0022	0.0028	0.0036	0.00009	0.00022	0.003	0.0008	0.0004	0.00231
P for LM		0.00059	0.0022	0.00179	0.0022	0.0028	0.0036	0.00009	0.00022	0.003	0.0008	0.0004	0.00231
Variance/P		189	230	47	138	84	91	188	148	149	259	273	229
AverageValueSTR		13	23	14	11	11	14	12	12	12	13	13	29
AverageAbsDeviation/p		153	247	45	157	41	76	161	93	140	225	148	181
AverageAbsDeviation		0.09	0.54	0.08	0.34	0.12	0.27	0.01	0.02	0.42	0.18	0.06	0.42
AbsDeviation		75	450	67	286	96	226	12	17	347	149	49	347
		-0.04	0.10	-0.52	-0.21	-0.40	-0.37	-0.05	-0.18	-0.18	0.19	0.24	0.09
		0.94	1.14	0.23	0.69	0.42	0.45	0.93	0.73	0.74	1.29	1.36	1.14
haplotype 1		13	24	14	11	11	13	12	12	11	13	13	29
haplotype 2		13	24	14	11	11	13	12	12	14	13	13	28
haplotype 3		13	24	14	11	11	14	12	12	12	13	13	29
haplotype 4		13	25	14	10	11	16	12	12	12	13	14	29
haplotype 5		13	25	15	11	11	15	11	12	11	13	13	30

Figure 13. A screenshot of the TMRCA Calculator for the series of 829 haplotypes of R1b-U106 haplotype tree in the 111 marker format, with five first haplotypes are partially shown, with the first 12 markers out of the 111 total.

a haplotype dataset. The second column (column B) in the downloaded calculator should be kept empty for present time haplotypes. Column B is needed when ancestral (base) haplotypes are employed in calculations (for example, when TMRCA of two ancient base haplotypes is sought). In those cases dates (calculated or actual, from excavated haplotypes, should be written into Column B.

2.2. Examples of TMRCA Calculations Using the Calculator

In this section some examples of calculations are provided. It should be noticed that some discrepancies between KKK (quadratic method) and LM (linear method), and between various haplotype formats (in the respective LM columns – D (6 marker haplotypes), E (12 marker), F (17 marker), G (25 marker), H (37 marker), I (67 marker) and J (111 marker haplotypes) do not necessarily reflect shortcomings of the methods or mistakes in the individual or cumulative values of mutation rate constants. Those discrepancies more likely reflect some non-uniformity of the haplotype datasets. Those are never ideal, and could not be ideal. Many mutations, and particularly slow mutations (assembled in the 22 marker panel) are inherited and not random in any given haplotype dataset. There are always some distortions in datasets, however, their contributions are compensated in more extended haplotypes and in more numerous haplotype dataset. As one can see, in some extended datasets a fit between KKK and LM in all the columns is fairly consistent.

2.2.1. R1b-M269, Downstream Subclades

A haplotype tree for 204 haplotypes of downstream subclades of haplogroup R1b-M269 is shown in **Figure 2**. The tree is fairly symmetrical, which shows that those 204 individuals likely descend from one common ancestor in terms of DNA genealogy. "One common ancestor" should not be taken literally, it could have been a more or less tight group of relatives, who shared a common base haplotype, shown in line 10 of the Calculator, 12 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 – 35 15 9 16 12 26 26 19 12 11 13 12 10 9 12 12 10 11 11 30 12 13 24 13 10 10 21 15 19 13 24 17 12 15 24 12 23 18 10 14 17 9 11 11, an offspring survived only from one individual, who lived, as the Calculator shows, 4583 ± 462 years before present (ybp) when calculated in the 111 marker format, apparently the most accurate of all the panels of the Calculator. The 67 marker panel results in the TMRCA of 4746 ± 481 ybp, which differs by only 3.6% from the 111 marker value. It should be added here, that the margin of error is calculated with an assumption that the cumulative mutation rate constant (0.198 and 0.12 mutations per 25 years, respectively) themselves have a margin of error of 10%, which in fact is around 3% - 4%. However, we prefer to set a higher margin of error in order to be on a safe side with TMRCA calculations. Overall, one can see that all ten TMRCAs overlap with each other within the margins of error. Of course, the TMRCA as 4583 ± 462 years is too "accurate" to be mathematically correct, it should be rounded up, as a variant, to 4600 ± 500 years for the 111 marker dataset. **Figure 3** data overall confirm this approximation.

The obtained values of the TRMCA employing the Calculator can be verified using the mutation counter (Abs Deviation line). For all the 111 markers the sum of the individual mutations in each is shown at the bottom of

the screen. It says Count: 111, Sum: 6370. For manual calculations, all 204 haplotypes in the 111 marker format show the apparent (observed) number of conditional generations to the common ancestor to be equal 6370/204/0.198 = 157.7, which should be corrected for back mutations as follows:

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

where: λ_{obs} is an observed average number of mutations per marker in the haplotype dataset, and λ is corrected value (Klyosov, 2009a, 2012). In the given case λ_{obs} equals to 6370/204/111 = 0.281, the formula above is trans- formed to

$$\lambda = \frac{0.281}{2}\left(1 + \exp\left(0.281\right)\right) = \frac{0.281}{2}\left(1 + 1.325\right) = 0.326$$

and the correction equals 0.326/0.281 = 1.16. Therefore, 157.7 (observed) conditional generations become 157.7 × 1.16 = 183 conditional generations, or 4575 ± 461 years to the common ancestor. The calculator gave 4583 ± 462 years, which is practically the same value. The reason for the insignificant difference is because the Calculator does not round up the intermediate numbers. The error margins are calculated using standard statistical approaches (Klyosov, 2009a). In this case the correction is equal to a noticeable figure of 16%, which is accumulated during 4600 years. "Back mutations" in this particular case are those 16% of the mutation alleles which are returned to the initial state, and we cannot see them in the counting by the linear method.

The obtained TMRCA for said 204 individuals to be 4600 ± 600 years which is significantly lower compared to the time of 13,500 ybp when the R1b-M269 subclade was formed as it was calculated employing the SNP values (https://www.yfull.com/tree/R1b/), however, and even significantly lower compared to the formation time for the downstream subclade R1b-M269-L23, which is between 5500 and 7300 ybp for various datasets. It seems that the tree in **Figure 2** (and the respective dataset) shows mainly L23 downstream subclades, and practically none of the parent M269* and L23* subclades.

A more extended dataset of 596 haplotypes of R1b-M269-L23 in the 67 marker format gave the TMRCA of 4661 ± 468 years, which is practically the same as 4583 ± 462 years, shown above, within the margin of error. When rounded, they gave 4700 ± 500 and 4600 ± 500 years, respectively. Again, it seems that most of those 596 haplotypes belong to downstream L23 subclades. In practical terms, subclade M269 arose in Siberia, and L23 arose apparently to the immediate west of the Ural mountains, however, the dataset represents haplotypes of mainly the Caucasus and Middle East origin (Klyosov, 2012).

2.2.2. I1-M253

A haplotype tree for 968 haplotypes of haplogroup I1-M253 in the 111 marker format is shown in **Figure 4**. The tree is symmetrical again, which shows that those 968 individuals likely descend from one common ancestor, in terms of DNA genealogy. The base, or ancestral haplotype, which Calculator shows in line 10, is as follows: 13 22 14 10 13 14 11 14 11 12 11 28 – 15898 11 23 1620 28 12 14 15 16 – 10 10 19 21 14 14 1620 35 37 – 12 10 11 8 15 15 8 11 10 8 99 12 23 25 15 10 12 12 16 8 13 25 20 13 13 11 12 11 11 12 11 – 32 128 17 12 24 27 19 11 12 12 13 11 9 11 11 10 12 12 31 11 13 21 16 11 10 24 15 19 11 24 17 13 15 25 12 22 18 12 14 18 9 12 11.

As **Figure 3** shows, the TMRCA values obtained from the 111- and 67-marker haplotypes are remarkably close to each other, that is 3686 ± 369 and 3618 ± 363 years, which gives only 1.9% difference. It shows a good balance between the mutation rate constants in the respective panels. Furthermore, all other LM panels show very similar figures, namely 3752, 3367, 3593, and 3469 years for the 12-, 17-, 25-, and 37 marker haplotypes (for the margins of error see **Figure 3**). Only the 6-marker haplotypes show a deviation by 13% - 15% to higher figures, which is not surprising, since statistics is limited there due to just a few markers; still, its TMRCA is within the margin of error with all other panels. The KKK111 (quadratic) value for the TMRCA is practically identical with those for the LM method. The KKK22 and LM22 TMRCA values are practically identical (3136 and 3088 years, respectively), and both are 7% - 17% lower than the respective LM values. Overall, the TMRCA for I1-M253 of those 968 individuals is around 3600 years, which reflects a population bottleneck, since the excavated haplotypes of haplogroup I have archaeological dates around 7000 ybp, and haplogroup I1 itself arose about 27,500 ybp. SNP-based data also point at a disconnection of the parent haplogroup I1 and its TMRCA (https://www.yfull.com/tree/I1/).

2.2.3. I2-M438

A haplotype tree of 244 haplotypes of haplogroup I2-M438 in the 111 marker format is shown in **Figure 6**. It represents an example of a non-uniform tree, which consists of at least three major branches. The branches could be split further, however, it is not really informative, as our experience shows. The larger branch of 157 haplotypes, in the upper part, has the base haplotype as follows: 13 24 16 11 13 15 11 13 12 13 1131 – 1781011 11 25 152031 12 13 14 15 – 10 1021 21 16 13 1818 34 36 – 12 10 11 8 15 157 12 10 8 1210 12 22 22 16 10 12 12 147 11 27 21 13 13 10 13 11 11 12 10 – 32 148 15 11 26 27 19 12 11 11 12 11 9 12 11 10 11 12 31 11 12 23 15 11 11 22 15 18 11 25 15 11 15 23 12 22 19 11 14 189 12 11.

The TMRCAs for the 111 and 67 marker haplotypes equal to 7285 ± 734 and 6986 ± 707 years, respectively, which differ by only 4.0% and practically is the same within the margin of error. The 12- and 17-marker haplotypes produce 7542 ± 802 and 6586 ± 686 years, KKK111 (quadratic method) results in 6281 ± 489 years. Overall, the TMRCA is around 7000 ± 700 years. KKK22 and LM22 produce, based on slow markers, a decreased and increased TMRCA, respectively, albeit with a large margin of error, which shows that the 22 markers, per se, do not introduce a systematic error. The problem is in a some non-uniformity of the branch, as it is seen from **Figure 6**.

A smaller branch of 52 haplotypes, at the lower right side, has the base haplotype as follows: 13 23 16 10 12 12 11 14 12 13 1129 – 168911 11 25 152130 11 14 14 15 – 11 1111 21 14 12 1819 33 34 – 12 10 11 8 16 168 12 10 8 117 12 21 21 16 11 12 12 148 12 22 20 13 13 10 13 11 11 12 11 – 30 148 16 11 26 27 18 11 11 10 11 11 9 12 11 10 12 12 30 11 13 22 16 10 12 21 15 19 11 25 16 12 14 26 12 22 18 12 15 159 11 11.

The TMRCAs for the 111 and 67 marker haplotypes equal to 5066 ± 521 and 5098 ± 533 years, respectively, which is practically the same. The KKK111 produces 5547 ± 515 years, which is almost equal to LM17, LM25 and LM37, which in turn are equal to 5649 ± 641, 5782 ± 638, and 5168 ± 549 years, within the margin of error.

Overall TMRCA is around 5100 ± 600 years. KKK22 and LM22 produce, based on slow markers, elevated TMRCA, again with a large margin of error, overlapping with all other TMRCA values for this branch.

The third branch, of 35 haplotypes, at the lower left side, has the base haplotype as follows: 13 23 15 10 12 15 11 15 12 14 1130 – 188911 11 26 141829 11 14 14 15 – 10 1021 21 14 10 1717 33 35 – 12 10 11 8 16 178 11 10 8 1210 12 21 21 17 10 12 12 158 14 26 20 11 14 12 13 10 11 12 11 – 29 158 15 11 26 27 19 11 11 12 12 10 9 13 11 10 11 12 30 11 13 22 16 11 10 23 15 20 10 24 18 12 14 26 12 21 18 12 14 1810 12 11.

The TMRCA for the 111, 67, and 25 marker haplotypes equal to 3542 ± 374, 3517 ± 384, and 3584 ± 437 years, respectively, which is practically the same. Other panels gave either elevated figures (KKK111, KKK22, LM12, LM22) or reduced ones (LM6, LM17, LM37), albeit within margins of error with other panels. Again, there is no any systematic trend in the TMRCAs for any panel, the deviations are due to some heterogeneity of the branches. Overall, the TMRCA is around 3600 ± 400 years.

All three base haplotypes have 108 mutations between them, which gives 108/3/0.198 = 182 → 222 conditional generations, that is 5550 years. The TMRCA for the whole tree in **Figure 6** approximately equals to this 5550 years plus the average TMRCA for the three branches considered above, resulting in approximately 10,800 years. The calculator produced the TMRCA for the whole dataset of 244 haplotypes in the 111 marker format equals to 11,510 ± 1271 years (111KKK) and 9575 ± 961 years (LM111), which are within the margin of error for all the three TMRCAs. It means that despite the significant heterogeneity of the tree (**Figure 6**) the Calculator gave a reasonable overall TMRCA, consistent with a semi-manual calculation.

2.2.4. J1-M267

A haplotype tree of 739 haplotypes of haplogroup J1-M267 in the 111 marker format is shown in **Figure 7**. It represents again an example of a non-uniform tree, which consists of three principal branches. The most ancient branch of 74 haplotypes, in the lower left side, has the base haplotype as follows: 12 24 14 10 14 18 11 15 11 13 1129 – 188911 11 26 142029 14 15 15 16 – 11 1020 22 15 13 1718 33 37 – 12 10 11 8 15 168 11 10 8 119 12 21 22 18 10 12 13 168 12 26 20 15 12 11 14 11 12 12 11 – 35 158 15 12 25 26 19 13 12 12 11 11 9 1211 10 10 1229 11 12 22 14 11 11 22 15 209 23 15 11 15 25 12 21 189 14 179 12 11.

The TMRCAs for the 111, 67 and 37 marker haplotypes equal to 8469 ± 858, 8114 ± 830 and 8274 ± 854 years, respectively, which differ by only 4.0% and are practically the same within the margin of error. KKK111 (quadratic method) results in 9819 ± 1105 years, again within the same margin of error. Overall, the TMRCA is around 8800 ± 700 years. The deviations are clearly in some non-uniformity of the branch, as it is seen from **Figure 7**.

The branch of 417 haplotypes on the right-hand side of the tree is obviously a rather young, since the bars, representing haplotypes, are short, compared to other two branches. The branch represents maily Arabic haplotypes from rather recent common ancestors. It has the base haplotype as follows: 12 23 14 11 13 19 11 17 11 13 1130 – 198911 11 26 142025 12 14 16 17 – 10 1022 22 14 14 1818 32 35 – 11 10 11 8 15 168 11 10 8 119 12 21 22 18 10 12 12 158 12 26 21 14 12 11 13 12 12 12 11 – 34 158 15 11 25 2720 13 12 13 11 13 9 1111 10 11 1129 11 13 22 15 11 10 20 15 2010 24 15 11 15 24 12 21 189 15 189 11 11.

The TMRCA values in all the ten panels of the Calculator are shown in **Figure 8**. For all the LM panels the TMRCA values are the same within the margins of error, particularly for LM111, LM67, LM37, and LM25 marker panels, which are equal to 2479 ± 250, 2509 ± 254, 2480 ± 252 and 2555 ± 262 years, respectively, and all differ by only 5.4%. KKK111 (quadratic method) results in 2381 ± 199 years, again within the same margin of error. For the slow 22 markers, KKK22 and LM22 the TMRCA values are slightly lower, still within margins of error. Overall, the TMRCA is around 2400 ± 400 years.

The branch of 248 haplotypes on the left-hand side is a remarkable one. It consists of Jewish and Arabic haplotypes, 98of them contain the "Cohen Modal Haplotype" 12 23 14 10 16 11 in the 6 marker format (DYS 393, 390, 19, 391, 388, 392). None of the 417 haplotypes on the right-hand side contains the CMH "signature", which, of course, does not belong to Cohens only. Many Arabs have the same "CMH", which makes it rather the "Abraham Modal Haplotype", if to continue the Biblical line. Theremaining 138 haplotypes in the left branch contain a mutated "signature", in a full accord with mutation dynamics. A "half-life time" of the 6-market haplotypes is (ln2)/0.0074 = 94 → 104 conditional generations, that is 2600 years, where 0.0074 is the mutation rate constant for the 6-marker haplotypes, and [ln(N/n)]/k = A is the basic formula for the logarithmic method for the TMRCA calculations (Klyosov, 2009), where N is the total number of haplotypes in the dataset, n is non-mutated/ancestral haplotypes in the dataset, and k is the mutation rate constant for the given haplotype format. If to apply the logarithmic formula to the branch on the left, we obtain the TMRCA of the branch equal to [ln(248/98)]/0.0074 = 125 → 145 conditional generations, that is 3625 ± 515 years. It is equal within the margin of error to 3849 ± 388 years, obtained from the 111 marker haplotypes by the linear method of calculations (**Figure 9**).

The branch representing the "Abraham" branch of the tree in **Figure 7** has the base haplotype as follows: 12 23 14 10 13 18 11 16 11 13 11 30 – 18 8 9 11 11 25 14 20 25 12 14 16 17 – 11 10 22 22 15 14 18 18 32 36 – 12 10 11 8 15 16 8 11 10 8 11 9 12 21 22 18 10 12 12 16 8 12 25 21 14 12 11 14 12 12 12 11 – 34 15 8 15 12 25 27 20 13 12 12 11 12 9 11 11 10 11 11 29 11 13 22 15 11 10 20 15 20 10 23 15 11 15 24 12 21 18 9 15 17 9 11 11.

It is remarkable again, that the TMRCA values are practically the same for all the ten panels in the Calculator, within the margin of error (**Figure 9**), and the overall TMRCA is equal to 4000 ± 300 years. All the LM and KK TMRCA values vary mostly within 2% - 3%, and only LM22 deviates, however, with a very large margin of error, still overlapping with all other TMRCA values.

All 248 haplotypes in the "Abraham" branch have 6651 mutations from the base haplotype, shown above (the figure shown in the calculator display when all 111 numbers in line 13 are highlighted), which gives for manual calculations the TMRCA of 6651/248/0.198 = 135 → 154 conditional generations, or 3850 ± 390 years. This is virtually identical with the TMRCA of 3849 ± 388 years produced by the Calculator.

The base haplotypes for the branches on the left and on the right hand side in **Figure 7** ("Abraham" branch and the young Arabic branch, respectively) differ by 17 mutations, however, those are "observed" mutations after rounding the respective alleles. Often alleles show, for example, values of 14.45 and 14.65, and they are rounded to 14 and 15, respectively, however, the difference between them equals to only 0.2. If to take it into account, the mutational difference between the two base haplotypes equals to 14. This gives a time-difference between the both haplotypes equal to 14/0.198 = 71 → 77 conditional generations, or approximately 1925 years, and a common ancestor of the "Abraham" and the "Young Arabic" branch lived (1925 + 2400 + 4000)/2 = 4160 ± 300 years ago, that is the "Abraham" itself (the TMRCA of 4000 ± 300 years). The logarithmic method (see above) gave the same TMRCA value, 3625 ± 515 years to the common ancestor, within the margin of error. In other words, both the "Abraham" and "Young Arabic" branches descended from the common ancestor of the Jews and the Arabs, who lived approximately 4000 years ago.

2.2.5. N1c1-M46

A haplotype tree of 275 haplotypes of haplogroup N1c1-M46 in the 111 marker format is shown in **Figure 10**. It is barely passable in terms of uniformity, since its left-hand side branch is noticeably younger compared to the

rest of the tree. If to assume that the tree is uniform, and employ the Calculator, it shows the base haplotype (in line 10) as follows: 14 23 14 11 11 13 11 12 10 14 1430 – 179911 12 25 141929 14 14 15 15 – 11 1118 20 14 15 1719 36 36 – 13 10 11 8 15 1788 10 8 1110 12 21 22 14 10 12 12 177 13 20 21 16 12 11 10 11 11 12 11 – 39 158 15 12 2327 19 13 14 11 12 13 9 1112 10 10 1231 12 12 21 18 119 23 15 2112 22 13 13 14 26 12 21 1811 13 168 12 11.

The TMRCA values in all the ten panels of the Calculator, shown in **Figure 11**, are practically the same, within the margin of error. The LM values for the 111, 67, and the 37 marker panels are 3233 ± 326, 3287 ± 333, and 3461 ± 352 years. The KKK111 TMRCA is 3086 ± 271 years. The KKK22 and LM22 TMRCA values, that is for slow markers, are both reduced in this case (but still are equal to all other within the margin of error), 2546 ± 546 and 2546 ± 921 years. Overall, TMRCA is around 3300 ± 300 years.

2.2.6. R1b-U106

A haplotype tree of 829 haplotypes of haplogroup R1b-U106 in the 111 marker format is shown in **Figure 12**. It is remarkably symmetrical, as much as a haplotype tree can be. The Calculator shows the base haplotype (in line 10) as follows: 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 38 – 12 12 11 9 15 16 8 10 10 8 10 10 11 23 23 16 10 12 12 15 8 12 22 20 13 12 11 13 11 11 13 12 – 35 15 9 16 12 26 26 19 12 11 13 12 10 9 12 12 10 11 11 30 12 13 24 13 10 10 21 15 19 13 24 17 12 15 24 12 23 18 10 14 17 9 12 11.

The TMRCA values in all the ten panels of the Calculator, shown in **Figure 13**, are practically the same, within the margin of error. The LM values for the 111, 67, and the 37 marker panels are 3584 ± 359, 3780 ± 379, and 3958 ± 398 years. The KKK111 and LM22 TMRCA values are 4109 ± 260 and 4092 ± 437 years, respectively. The KKK22 TMRCA value, that is for slow markers, is elevated to 4631 ± 784, but still overlaps with all other TMRCA values within the margin of error. It shows again that fluctuations of KKK and LM values for the 22 marker haplotypes are random, and do not reveal any systematic (one-sided) errors. Overall, TMRCA is around 3700 ± 400 years.

The base haplotypes for R1b-M269-… (see above) and R1b-U106 are surprisingly similar, there are only 8 mutations between them. This translates to 8/0.198 = 40 → 42 conditional generations, or 1050 years between them. This date can be interpreted differently depending on whether the R1b-M269… subclades are parent ones with respect to R1b-M269-L23-U106, or they are "parallel", such as R1b-M269-L23-Z2103.

2.2.7. A Few More Brief Examples of European Subclades R1b-L21, R1a-Z283, R1a-M458

Hundreds of examples could have been listed here, illustrating usage of the Calculator for the last several years, however, we restrict ourselves with a few more figures for some important subclades which are often considered in the literature. For the 3466 haplotypes of R1b-L21 subclade, the TMRCAs for 111, 67, 37, 25 and 17 marker haplotypes are equal to 3810 ± 381, 3841 ± 384, 3576 ± 358, 3571 ± 358, and 3679 ± 369 years. Here, across this paper, we give dates with excessive number of digits, unrealistic for practical purposes, just to show the Calculator outputs. For any practical goals, the figures should be rounded.

For 113 haplotypes of R1a-Z283 subclade, the TMRCAs for 67, 37, 17 and 12 marker haplotypes are equal to 4503 ± 461, 4898 ± 505, 4549 ± 492, and 4529 ± 513 years. For 24 haplotypes of the same haplogroupin the 111 marker format the KKK111 equals to 4281 ± 553 years.

For 754 haplotypes of R1a-M458 subclade, the TMRCAs for 67, 37, 25, 17 and 12 marker haplotypes are equal to 3668 ± 368, 3799 ± 382, 3866 ± 391, 3308 ± 336, and 3833 ± 393 years.

2.2.8. The Calculator and Some Examples of Do Cumented Genealogy

We will give here only two examples which illustrate an appropriate fit of the Calculator data to the documented genealogy, representing two quite different cases. In one, a series, albeit small, of 111 marker haplotypes was available. In another, a set of assorted haplotypes in different formats was provided. Other cases of documented genealogy along with several haplotypes are typically positioned between these two extreme cases. Larger sets of haplotypes coupled with documented genealogy data practically always are non-uniform ones, such as those of the Donald Clan of haplogroup R1a (**Figure 14**). According to the official (but not necessarily documented) genealogy, the common ancestor of the haplogroup R1a branch of the Clan, John Lord of the Isles, died in 1386, that is 630 years ago. However, the tree suggests that among the Clan members are many individuals who descended from more ancient common ancestors, making the haplotype dataset a mix. The Calculator shows a
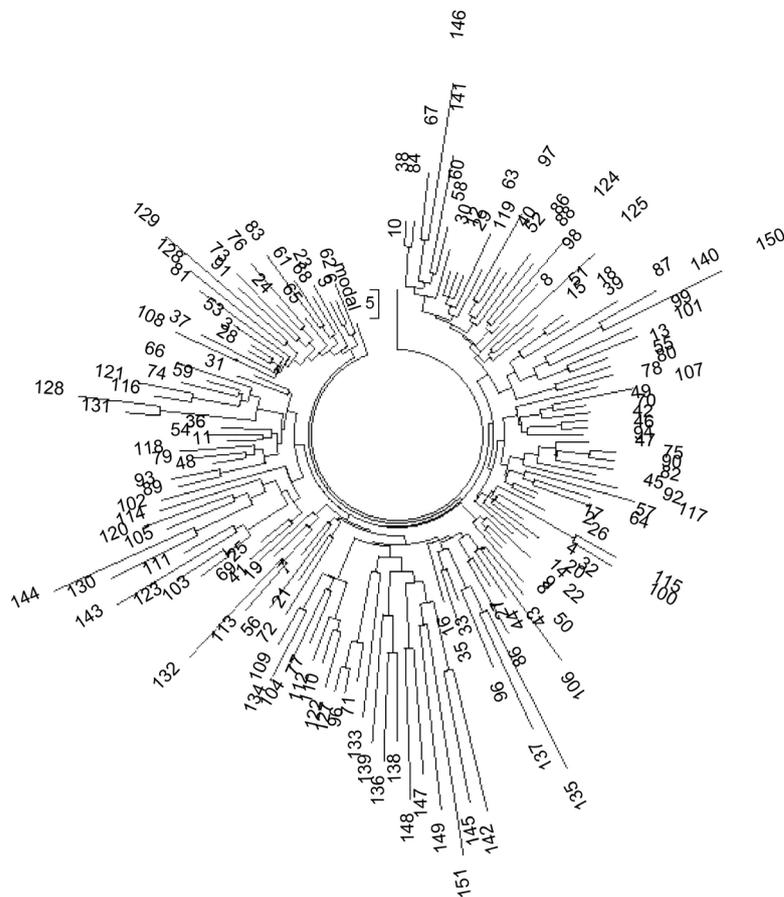
Figure 14. A haplotype tree of the Donald Clan, haplogroup R1a, containing 151 haplotypes in the 67 marker format. There are at least five branches on the tree. The haplotypes are listed in the Donald Clan database.

"phantom" common ancestor of the dataset of 151 haplotypes with the TMRCA of 1031 ± 110, 1034 ± 113, and 1054 ± 128 years for the 67, 37, and 17 marker haplotypes, respectively. The KKK111 gave the TMRCA of 828 ± 116 years. The discrepancy obviously reflects a non-uniformity of the haplotype dataset.

Let us consider five 111 marker haplotypes of Ashkenazi Jews (the Horowitz rabbinical family) of R1a-Z93-Z94-YP264 subclade. Their known documented genealogy places their common ancestor to 1507-1572, that is 442 - 507 years ago. [https://sites.google.com/site/levitedna/y-dna-analysis/snp-analysis---klyosov-s-comment]. The Calculator gives 436 ± 114, 466 ± 148, and 454 ± 167 years ago by the 111, 67, and 37 marker panels.

Another example concerns Capt. Thomas Osborne, b 1580 in England, who came to Virginia in 1619 (http://freepages.genealogy.rootsweb.ancestry.com/~tlosborne/AusburnSurnameProject/fdnatip.htm), that is 397 years ago. A set of 10 assorted haplotypes in the 12, 25, and 37 marker format was provided in the above link, in which the first 12 markers were identical in all the ten haplotypes, eight 25 marker haplotypes contained four mutations, and four 37 marker haplotypes contained two mutations. Only the quadratic method could have been employed in this case, and the KKK111 panel showed 379 ± 270 years to a common ancestor.

2.2.9. Ancient and Very Ancient Common Ancestors
R1a-R1b

Figure 15 shows two base haplotypes of R1a-Z280 and R1b-M269 (Klyosov & Rozhanskii, 2012) in the 22 slow marker format. The Calculator in this case, as for other 22 marker haplotypes was rearranged to make it more compact. Since it was determined earlier (Klyosov, 2012; Klyosov & Rozhanskii, 2012b) that the TMRCA for the respective subclades equaled 4900 and 7000 years, respectively, these figures were added to Column B of the Calculator. The overall TMRCA is equal to 27,992 years, which, of course, is 28,000 years after the figure was rounded, and the common ancestor is obviously the haplogroup R1. The SNP-based calculations for time of

formation of haplogroup R1 is equal to 28,200 ± 2300 years (https://www.yfull.com/tree/R1/).

R1a-A0

Figure 16 shows the base haplotypes of R1a-Z280 and a contemporary haplotype of A0-V166, listed in the FTDNA Haplogroup A Project, in the 22 slow marker format. Manual calculations by the linear method are close to impossible there, since a number of mutations between the two haplotypes equals to 18, and close to the number of markers in the haplotypes, which makes the system unstable. The Calculator produces the TMRCA equal to 187,482 years, which can be rounded to 187,000 and the common ancestor is obviously the haplogroup A0 itself. The SNP-based calculations for time of formation of haplogroup A0 is not listed in (https://www.yfull.com/tree/A0-T/), however, it can be estimated as 161,300 ± 8600 years, based on some information provided there. If so, a difference between them is about 16%.

A0-A00

Let us consider the most ancient *Homo sapiens* Y Chromosomal haplogroup known to date, which is haplogroup A00. An approach based on counting SNP-mutations, provided the TMRCA values between 208,000 years (Elhaik et al., 2014) and 235,900 years (https://www.yfull.com/tree/A00/); the earlier value of 338,000 years (Mendez et al., 2013) was dismissed in the literature, since the calculations employed too low mutation rates taken from autosomal data. **Figure 17** shows contemporary haplotypes of haplogroups A0 and A00, listed in the FTDNA Haplogroup A Project. Manual calculations by the linear method are practically impossible there, since a number of mutations between the two haplotypes equals to 27, and exceeds the number of markers in the haplotypes, which makes the system unstable. The Calculator, using the 22 market slow panel, produced the TMRCA for a common ancestor of A00 and A0-V166 of 217,436 years ago, which is, of course, 217,000 years ago after rounding.

R1b-Chimpanzee

It is assumed in the literature that a common ancestor of chimpanzee and modern humans lived about 4 - 5 million years ago. Those assumptions are based on some anthropological data, and direct estimates are absent. We have conducted search for chimpanzee Y chromosome markers in the European Nucleotide Archive (ENA) database, employing Whole Genome Shotgun (WGS), and in the National Center for Biotechnology Information (NCBI) GenBank and ENA, as described in (Klyosov et al., 2012). We have succeeded in retrieving alleles of 16 chimpanzee Y chromosome markers, as indicated in **Figure 18**. The Calculator produced the TMRCA for a common ancestor of chimpanzee and modern humans, exemplified with a haplotype R1b-P312, of 4,290,000 years ago.

	TMRCA 27992																							
standard deviation	9642	3	100																					
STR	age of haplotype	DYS426	DYS388	DYS392	DYS455	DYS454	DYS438	DYS531	DYS578	DYS395S1a	DYS395S1b	DYS590	DYS641	DYS472	DYS425	DYS594	DYS436	DYS490	DYS450	DYS617	DYS568	DYS640	DYS492	
P=1/k(mutation rate constant)		0.00009	0.00022	0.0004	0.0001	0.0003	0.00035	0.0009	0.00043	0.0004	0.004	0.00034	0.00017	0.000008	0.000042	0.00017	0.00004	0.00007	0.00011	0.0005	0.00005	0.00015	0.00015	0.00539
Variance/P		5556	0	5000	0	0	1429	0	1163	5000	1250	0	0	0	0	0	0	0	0	0	0	0	0	
AverageValueSTR		12	12	12	11	11	12	11	9	16	17	8	10	8	12	10	12	12	8	12	11	11	12	
		2.45	-0.62	2.15	-0.62	-0.62	0.17	-0.62	0.02	2.15	0.07	-0.62	-0.62	-0.62	-0.62	-0.62	-0.62	-0.62	-0.62	-0.62	-0.62	-0.62	-0.62	
		4.96	0.00	4.47	0.00	0.00	1.28	0.00	1.04	4.47	1.12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
R1a-Z280	4900	12	12	11	11	11	11	11	8	17	17	8	10	8	12	10	12	12	8	12	11	11	12	
R1b-M269	7000	11	12	13	11	11	12	11	9	15	16	8	10	8	12	10	12	12	8	12	11	11	12	

Figure 15. A screenshot of the TMRCA Calculator for the 22 marker base haplotypes of R1a-Z280 and R1b-M269 (Klyosov, 2012).

	TMRCA 187482																							
standard deviation	109316	3	100																					
STR	age of haplotype	DYS426	DYS388	DYS392	DYS455	DYS454	DYS438	DYS531	DYS578	DYS395S1a	DYS395S1b	DYS590	DYS641	DYS472	DYS425	DYS594	DYS436	DYS490	DYS450	DYS617	DYS568	DYS640	DYS492	
P=1/k(mutation rate constant)		0.00009	0.00022	0.0004	0.0001	0.0003	0.00035	0.0009	0.00043	0.0004	0.004	0.00034	0.00017	8E-06	4.2E-05	0.00017	0.00004	0.00007	0.00011	0.0005	0.00005	0.00015	0.00015	0
Variance/P		0	2273	1250	5000	0	1429	0	0	5000	5000	0	2941	0	11905	0	0	28571	0	0	90000	0	0	
AverageValueSTR		12	13	11	11	11	11	11	8	16	16	8	10	8	6	10	11	13	8	12	10	11	12	
		-0.41	-0.34	-0.37	-0.24	-0.41	-0.36	-0.41	-0.41	-0.24	-0.24	-0.41	-0.31	-0.41	-0.01	0.41	3.37	0.55	-0.41	-0.41	2.61	-0.41	-0.41	
		0.00	0.19	0.10	0.41	0.00	0.12	0.00	0.00	0.41	0.41	0.00	0.24	0.00	0.97	0.00	9.16	2.33	0.00	0.00	7.33	0.00	0.00	
R1a-Z280	4900	12	12	11	11	11	11	11	8	17	17	8	10	8	12	10	12	12	8	12	11	11	12	
A0-V166		12	13	10	10	11	10	11	8	15	15	8	9	8	0	10	9	14	8	12	8	11	12	

Figure 16. A screenshot of the TMRCA Calculator for the 22 marker base haplotype of R1a-Z280 and contemporary A0-V166 haplotype.

	TMRCA 217436																							
standard deviation	84643	3	100																					
STR	age of haplotype	DYS426	DYS388	DYS392	DYS455	DYS454	DYS438	DYS531	DYS578	DYS395S1a	DYS395S1b	DYS590	DYS641	DYS472	DYS425	DYS594	DYS436	DYS490	DYS450	DYS617	DYS568	DYS640	DYS492	
P=1/k(mutation rate constant)		0.00009	0.00022	0.0004	0.0001	0.0003	0.00035	0.0009	0.00043	0.0004	0.004	0.00034	0.00017	8E-06	4.2E-05	0.00017	0.00004	0.00007	0.00011	0.0005	0.00005	0.00015	0.00015	0.00539
Variance/P		5556	9091	5000	0	0	51429	556	0	1250	1250	0	2941	0	11905	11765	50000	28571	0	0	0	0	3333	
AverageValueSTR		13	12	11	10	11	13	11	8	15	15	8	9	8	5	11	10	13	8	12	10	11	12	
		-0.28	-0.18	-0.30	-0.44	-0.44	1.01	-0.42	-0.44	-0.40	-0.40	-0.44	-0.35	-0.44	-0.10	-0.11	0.97	0.36	-0.44	-0.44	4.05	-0.44	-0.34	
		0.36	0.58	0.32	0.00	0.00	3.30	0.04	0.00	0.08	0.08	0.00	0.19	0.00	0.76	0.76	3.21	1.83	0.00	0.00	10.27	0.00	0.21	
(A00,Mbo)		13	11	12	10	11	16	10	8	14	14	8	8	8	9	12	11	12	8	12	12	11	11	
(A0-V166)		12	13	10	10	11	10	11	8	15	15	8	9	8	0	10	9	14	8	12	8	11	12	

Figure 17. A screenshot of the TMRCA calculator for the 22 marker base haplotypes of A00 and A0-V166 (Klyosov et al., 2012).

The database Y Search contains a different haplotype of chimpanzee, under ID 6RCUU (http://www.ysearch.org/lastname_view.asp?uid=&letter=&lastname=chimpanzee&viewuid=6RCUU&p=0) listed there by Thomas Krahn. The haplotype differs, albeit significantly, from the base R1b-P312 haplotype only in a few markers, and in the rest it is too close to R1b-P312. It resulted in an unrealistically recent TMRCA for chimpanzee and humans, that was about 143,000 years (**Figure 19**). Apparently, the main reason of this absurd results is that Krahn determined, unfortunately, mainly fast markers for the chimpanzee haplotype, which are not applicable for data analysis for such a distant common ancestor. Indeed, fast markers for ancient primates do not make sense for any meaningful calculations.

2.2.10. Alternative Sets of Mutation Rate Constants for Individual Markers

Since 2006 a few sets of individual mutation rate constants for 67 and 111 markers have appeared. Among them are Chandler table (2006) for the first 37 markers, then upgraded to 67 markers, Heinila table (Heinila, 2012) for a set of 111 markers (Genealogy-DNA Digest, vol. 9, Issue 232), and an unnamed set of mutation rates known as "estimated for 3565 haplotypes" (Anonymous, 2014). Besides, there are extended father-son studies for intended 111 markers by Ballantyne et al. (2010) and Burgarella et al. (2011), which have many omissions in their series, and, as it will be shown below, the Ballantyne table is practically not applicable for the TMRCA calculations. Since the Burgarella et al. data are similar in kind with the latter, it is not considered here.

The Chandler's table of mutation rates, despite its history for 10 years, was barely used for TMRCA calculations in the literature. Partly it was neglected because of so-called "Zhivotovsky mutation rates", or (a synonym) "population mutation rates" was published in 2004 (Zhivotovsky et al., 2004), and soon became the only accepted mutation rates for reviewers of academic publications. Besides, it did not reach the scientific community because 1) the scientific community was not ready for individual mutation rate constants, and 2) the mutation rates were set per "generation" while chronology is history is not measured in "generations", it is measured in years; however, Chandler did not provide a factor which would allow to translate "generations" to years. The same was essentially applicable to the Heinila and the "3565 haplotypes" tables, which were not published in the scientific literature, and were known only in the net. It is not very productive to provide a loose criticism of the tables, instead, we will give here some specific examples what those tables result in.

The Ballantyne table presents one big confusion. Unfortunately, those who employ their table, do not realize that their data are not applicable for actual calculations. First, the table has too many omissions. For about 1700 father-son pairs (tested for intended 111 markers), 24 markers were not tested at all, in 17 additional markers there were no any mutation, in 15 additional markets there was only one mutation (per 1700 father-son pairs), which altogether makes 56 markers out of 111 (that is 50% of all) being non-functional in terms of the mutation rate constants. On top of it, in 11 markers there were only two mutations, which do not provide any meaningful statistics. Overall, two-thirds of those 111 markers are practically non-usable. Some estimated mutation rates are obviously erroneous, due to, probably, poor statistics, such as mutation rate for DYS393 reported to be faster than that for DYS390 (0.00211 and 0.00152, respectively, in mutations per generation), while anyone who works with mutations in Y chromosome knows that reverse is true (0.00059 and 0.00220, respectively, in mutations per 25 years, see **Figure 1**). Last, but not the least, data obtained in the mutation rate table, were related to "one generation", without any explanation what it might be in terms of real time/chronology. One generation can be almost anything between, say, at least 20 years to 35 years, since among the father-son pairs in

TMRCA	4290010																							
STR	age of haplotype	DYS426	DYS388	DYS392	DYS455	DYS454	DYS438	DYS531	DYS578	DYS395S1a	DYS395S1b	DYS590	DYS641	DYS472	DYS425	DYS594	DYS436	DYS490	DYS450	DYS617	DYS568	DYS640	DYS492	
P=1/k(mutation rate constant)		0.00009	0.00022	0.0004	0.0001	0.0003	0.00035	0.0009	0.00043	0.0004	0.004	0.00034	0.00017	0.000008	0.000042	0.00017	0.00004	0.00007	0.00011	0.0005	0.00005	0.00015	0.00015	0.00539
Variance/P		88889	20455		11250	245000		70000				0			562500	47619	105882	800000	178571	64000	490000	30000	30000	
AverageValueSTR		10	14	12	8	9		9				10			7	11	7	8	10	8	8	10	11	
Chimp		8	15	10	4	5		9				10			5	10	4	4	7	4	4	8	9	
Human	4500	12	12	13	11	12		9				10			8	12	10	12	12	12	11	11	12	

Figure 18. A screenshot of the TMRCA calculator for the 16 marker haplotype of chimpanzee (Klyosov et al., 2012) and the respective 16 marker base haplotype of R1b-P312 subclade (Klyosov, 2012).

TMRCA	142950																							
STR	age of haplotype	DYS426	DYS388	DYS393	DYS389I	DYS437	DYS578	DYS537	DYS578	DYS511	DYS441	DYS485	DYS641	DYS533	DYS71	GATAA10	GATAH4	DYS444	DYS520	DYS446	DYS461	DYS640	DYS492	
P=1/k(mutation rate constant)		0.00009	0.00022	0.00059	0.0008	0.00064	0.00043	0.0009	0.00043	0.0009	0.00021	0.0016	0.00017	0.00125	0.0006	0.0032	0.0017	0.0018	0.0015	0.0028	0.0014	0.00015	0.00015	0.02342
Variance/P		22222	0	7627	625	781	18605	556	18605	5000	3810	313	0	3600	0	10000	1176	1111	1333	11429	357	3333	13333	
AverageValueSTR		11	12	12	13	15	7	10	7	12	11	15	10	11	19	9	10	11	19	17	13	12	11	
Chimp		10	12	10	12	14	5	9	5	13	9	14	10	9	19	5	9	10	18	21	13	12	10	
Human	4500	12	12	13	13	15	9	10	9	10	13	15	10	12	19	13	11	12	20	13	12	11	12	

Figure 19. A screenshot of the TMRCA calculator for a non-standard 22 marker haplotype of chimpanzee (Krahn, ID 6RCUU in Y search) and the respective base haplotype of R1b-P312 subclade (Klyosov, 2012).

Ballantyne et al. (2010) were 70 years old fathers. As a result, the Ballantyne and Burgarella tables of mutation rates cannot be employed for meaningful TMRCA calculations.

Let us show it by employing several specific examples, taking a conditional generation for 25 years (if someone wants to take it for 30 years or any other timespan, the data can be easily recalculated, by adjusting the mutation rate constants accordingly).

Example 1. The TMRCA for five 111 marker haplotypes of the Horowitz rabbinical family (see above), whose known documented genealogy places their common ancestor to 1507-1572 AD, that is 442 - 507 years ago. A comparison of the TMRCAs for said four mutation rate tables is shown in **Table 1**.

The Ballantyne mutation rate table misses 20 markers out of the 111 markers, hence, it misses several mutations with them. The differences with the Calculator TMRCAs can be explained if a "generation" in the Ballantyne table would be taken as 35 to 39 years, rather than a "conditional generation" of 25 years employed for a calibration of mutation rate constants in our study, but how people who calculate the TMRCAs would know it?

The Chandler mutation rate table (data for only 67 markers are available) gives the TMRCAs, which present only half of the timing of the documented genealogy, and of the Calculator data. This was expected, since the Chandler cumulative mutation rates for the 12, 25, 37, and 67 marker haplotypes (0.0224, 0.0695, 0.182, and 0.224 per generation, respectively) were progressively inflated compared to the Calculator values (0.020, 0.046, 0.009, and 0.120 mutations per 25 years).

The Heinila table gives the TMRCAs, which are again much lower compared to the documented genealogy data and to the Calculator data. This again was expected, since the Heinila cumulative mutation rates for the 12, 25, 37, 67, and 111 marker haplotypes (0.0243, 0.0605, 0.132, 0.173, and 0.291 per generation, respectively) were progressively inflated compared to the Calculator values (0.020, 0.046, 0.009, 0.120, and 0.198 mutations per 25 years).

The anonymous "estimated with 3565 haplotypes" set gives the TMRCAs, which are again much lower compared to the documented genealogy data and to the Calculator data. This again was expected, since the "3565 haplotypes" cumulative mutation rates for the 12, 25, 37, 67, and 111 marker haplotypes (0.021, 0.054, 0.122, 0.158, and 0.255 per generation, respectively) were progressively inflated compared to the Calculator values (0.020, 0.046, 0.009, 0.120, and 0.198 mutations per 25 years).

Example 2. The TMRCAs for 829 haplotypes of haplogroup R1b-U106 in the 111 marker format are shown in **Table 2**. As always, the Ballantyne mutation rate table misses 20 markers out of the 111. The differences with the Calculator TMRCAs can be explained if a "generation" in the Ballantyne table would be taken as 30 to 40 years, rather than a "conditional generation" of 25 years employed for a calibration of mutation rate constants in our study.

As always, the Chandler mutation rate table (data for only 67 markers are available) gives the TMRCAs, which present only half of the Calculator TMRCA data. The reason for it is explained in the preceding section.

The Heinila table gives the TMRCAs, which are again much lower compared to the Calculator TMRCA.

Table 1. The TMRCA for five 111 marker haplotypes of the Horowitz rabbinical family, whose known documented genealogy places their common ancestor to 1507-1572 AD, that is 442 - 507 years before present. Margins of error are omitted here, since they are the same in all the columns, percent-wise, because a number of mutations in the haplotype dataset is the same in all the lines in the Table (except the Ballantyne set of data, since a few mutation are missing there due to lack of the respective markers). Due to a short series of haplotypes and only a few mutations in them, the margins of error are rather high, and equal to ±26%, ±32%, and ±37%, in the 111, 67, and 37 marker haplotypes. In the Ballantyne set of data the margins of error equal 29%, 35%, and 42%, respectively.

	111 marker haplotype	67 marker haplotype	37 marker haplotype
Documented genealogy	***************442 - 507***************		
Kilin-Klyosov calculator	436	466	454
Chandler	n.a.	249	225
Ballantyne	317	393	313
Heinila	297	323	309
Anonymous "3565 haplo"	338	354	335

The anonymous "estimated with 3565 haplotypes" set gives the TMRCAs, which are again much lower compared to the Calculator TMRCA.

Example 3. The TMRCAs for 3466 haplotypes of haplogroup R1b-L21 in the 111 marker format are shown in **Table 3**. As always, the Ballantyne mutation rate table misses 20 markers out of the 111. The differences with the Calculator TMRCAs can be explained if a "generation" in the Ballantyne table would be taken as 36 to 41 years.

As always, the Chandler mutation rate table (data for only 67 markers are available) gives the TMRCAs, which present only half of the Calculator TMRCA data. The reason for it is explained in the preceding section.

The Heinila table gives the TMRCAs, which are again is significantly (33% lower) compared to the Calculator TMRCA.

The anonymous "estimated with 3565 haplotypes" set gives the TMRCAs, which are again significantly lower (22% - 26% lower) compared to the Calculator TMRCA.

Example 4. The TMRCAs for 968 haplotypes of haplogroup I1-M253 in the 111 marker format are shown in **Table 4**. As always, the Ballantyne mutation rate table misses 20 markers out of the 111. The differences with the Calculator TMRCAs can be explained if a "generation" in the Ballantyne table would be taken as 34 to 42 years.

As always, the Chandler mutation rate table (data for only 67 markers are available) gives the TMRCAs, which present only half of the Calculator TMRCA data. The reason for it is explained in the preceding sections.

Table 2. The TMRCA for 829 haplotypes of haplogroup R1b-U106 in the 111 marker format. Margins of error for all five lines equal to ±10%.

	111 marker haplotype	67 marker haplotype	37 marker haplotype
Kilin-Klyosov calculator	3584	3780	3958
Chandler	n.a.	2022	1956
Ballantyne	2230	2406	2944
Heinila	2440	2622	2691
Anonymous "3565 haplo"	2780	2868	2919

Table 3. The TMRCA for 3466 haplotypes of haplogroup R1b-L21 in the 111 marker format. Margins of error for all five lines equal to ±10%.

	111 marker haplotype	67 marker haplotype	37 marker haplotype
Kilin-Klyosov calculator	3810	3841	3576
Chandler	n.a.	2055	1768
Ballantyne	2298	2338	2462
Heinila	2594	2665	2431
Anonymous "3565 haplo"	2955	2914	2637

Table 4. The TMRCA for 968 haplotypes of haplogroup I1-M253 in the 111 marker format. Margins of error for all five lines equal to ±10%.

	111 marker haplotype	67 marker haplotype	37 marker haplotype
Kilin-Klyosov calculator	3686	3618	3469
Chandler	n.a.	1936	1715
Ballantyne	2201	2264	2582
Heinila	2510	2511	2358
Anonymous "3565 haplo"	2680	2746	2558

The Heinila table gives the TMRCAs, which are again is significantly (23% - 32% lower) compared to the Calculator TMRCA.

The anonymous "estimated with 3565 haplotypes" set gives the TMRCAs, which are again significantly lower (22% - 26% lower) compared to the Calculator TMRCA.

Example 5. The TMRCA for several pairs of haplotypes descended from ancient common ancestors are shown in **Table 5**. As always, the Ballantyne mutation rate table misses 20 markers out of the 111. The differences with the Calculator TMRCAs can be explained if a "generation" in the Ballantyne table would be taken as 34 to 42 years.

As always, the Chandler mutation rate table (data for only 67 markers are available) gives the TMRCAs, which present only half of the Calculator TMRCA data. The reason for it is explained in the preceding sections.

The Heinila table gives the TMRCAs, which are again is significantly (23% - 32% lower) compared to the Calculator TMRCA.

The anonymous "estimated with 3565 haplotypes" set gives the TMRCAs, which are again significantly lower (22% - 26% lower) compared to the Calculator TMRCA.

3. Discussion and Conclusions

This paper presents a multifunctional automatic TMRCA Calculator for haplotypes of any number of markers within 111 markers in the FTDNA nomenclature. A number of haplotype datasets from a few assorted (in terms of their markers) haplotypes to thousands of 111 marker haplotypes were presented, within a time span of their TMRCAs from a few hundred years to four million years. The Calculator showed an appropriate reproducibility and cross-verification between the datasets and with respect to documented genealogy and SNP-based and anthropological data. In the course of examining the Calculator, with hundreds of datasets over several years, the conclusion was that the capacity and reliability of the Calculator is unmatched. It operates with multi marker alleles, with zero markers, it detects "foreign" haplotypes which do not belong to the dataset (spotting excessive dispersion of alleles) and switches the respective marker off, and this feature of the Calculator can be regulated on different levels of sensitivity.

An attention was paid to other mutation rate tables, exemplified with tables by Chandler, Ballantyne et al., Heinila, and an anonymous author whose table is mentioned not once in the literature. It was shown that those tables commonly results in significant underestimates of the TMRCAs. The most deviated TMRCAs have resulted by using the Ballantyne et al. table of mutation rates, particularly when a fraction of "slow" markers is rather high. The reason is simple—Ballantyne et al. employed the father-son estimates, and typically could not detect even a single mutation between "slow markers" in ~1700 father-son pairs, that is below $1/1700 = 0.0006$ mutation/generation. There are at least 40 of such markers among the 111 marker panel. Surprisingly, Balantyne et al. assigned "material" mutation rates to each one of them, and almost all of them were significantly overestimated. As a result, the TMRCA for the R1a – R1b haplogroups (which by definition is R1 haplogroup, which arose $28,200 \pm 2300$ years ago, as determined using SNP data (https://www.yfull.com/tree/R1/), and for which the Calculator gave 28,000 years, Ballantyne et al. mutation rates gave 9000 years, which is totally unrealistic figure. The trend continues across all datasets, considered in this work: the TMRCA for A0 (187,000 years by

Table 5. The TMRCA for R1a – R1b, R1a – A0, A0 – A00 in the 22 marker format, and human (R1a) and chimpanzee in the 16 marker format. Details are listed in **Figures 15-19**. The TMRCAs are rounded for a nearest thousand years. A common ancestor of R1a – R1b lived $28,200 \pm 2300$ years before the present (ybp), as determined employing SNP data (https://www.yfull.com/tree/R1/). Haplogroup A0 is not dated, however, the parallel haplogroup A1 formed $161,300 \pm 8500$ ybp (https://www.yfull.com/tree/A0-T/). Haplogroup A00 arose $235,900 \pm 7700$ ybp (https://www.yfull.com/tree/A00/).

	R1a – R1b	R1a – A0	A0 – A00	Human-chimp
Kilin-Klyosov calculator	28,000	187,000	217,000	4,290,000
Chandler	34,000	82,000	121,000	1,930,000
Ballantyne	9000	21,000	41,000	604,000
Heinila	26,000	116,000	130,000	1,808,000
Anonymous "3565 haplotypes"	34,000	61,000	163,000	2,355,000

the Calculator) Ballantyne et al. data gave 21,000 years, for A00 (217,000 years by the Calculator, and between 209,000 and 235,000 by the genomic data), Ballantyne et al. gave 41,000 years, for chimpanzee (4.29 million years by the Calculator), Ballantyne et al. data gave 604,000 years, which is absurd, of course.

The other three tables of mutation rates turned out to be suffered from overestimates of mutation rates from slow markers in particular. Generally, it is clear why. Slow markers are sensitive to inherited mutations, which stick to a dataset in multiple quantities, which are in fact not random, but inherited. Let us consider, as an example, a dataset of 100 haplotypes, in which DYS426 (a slow marker, which mutates on average once in 1/0.00009 = 11,111 generations) mutated—randomly—just once (statistically, it is quite possible, particularly if the TMRCA for the dataset equals 111 conditional generations, that is about 2800 years). Then the dataset contains just one mutated DYS426, if it does not includes members of the same lineage. If, however, the dataset includes 10 members of this particular lineage, each with the same mutation in DYS426, then the dataset contains 10 DYS426 mutated haplotypes, and the mutation rate for DYS426 would be taken as 10 times higher. To avoid such mistakes, haplotype trees should be considered for each analyzed data, in which such repeated allele, particularly in slow markers, would form a separate branch. Apparently, such an approach was not considered in Chandler, Heinila, and other mutation rate tables, and many slow markers produced overestimated values. The Heinila table as well as of the anonymous table showed better results compared to the Chandler table, however, they also deviated quite significantly for some datasets.

Haplotype trees were composed using PHYLIP, the Phylogeny Inference Package program (Felsenstein, 2004); for multiple examples of usage and calculations see Klyosov, 2009a, 2009b, 2009c, 2012; Klyosov and Rozhanskii, 2012a, 2012b, and references therein).

The authors express hope that a long debate over which mutation rates to employ, how to count mutations, and how to calculate the TMRCA in various cases, including complicated ones, is eventually over.

Acknowledgements

References

Anonymous (2014). Mutation Rates. http://www.johnbrobb.com/Content/DNA/MarkerPanelsCompared.pdf

Ballantyne, K. N., Goedbloed, M., Fang, R., Schaap, O., Lao, O., Wollstein, A. et al. (2010). Mutability of Y-Chromosomal Microsatellites: Rates, Characteristic, Molecular Bases, and Forensic Implications. *American Journal of Human Genetics, 7,* 341-353. http://dx.doi.org/10.1016/j.ajhg.2010.08.006

Burgarella, C., & Navascues, M. (2011). Mutation Rate Estimates for 110 Y-Chromosome STRs Combining Population and Father-Son Pair Data. *European Journal of Human Genetics, 19,* 70-75. http://dx.doi.org/10.1038/ejhg.2010.154

Chandler, J. F. (2006). Estimating Per-Locus Mutation Rates. *Journal of Genetic Genealogy, 2,* 27-33.

Elhaik, E., Tatarinova, T. V., Klyosov, A. A., & Graur, D. (2014). The "Extremelyancient" Chromosome That Isn't: A Forensic Bioinformatic Investigation of Albert Perry's X-Degenerate Portion of the Y-Chromosome. *European Journal of Human Genetics, 22,* 1111-1116. http://dx.doi.org/10.1038/ejhg.2013.303

Felsenstein, J. (2004). *PHYLIP (Phylogeny Inference Package).* Version 3.6., Seattle: Department of Genome Sciences, University of Washington.

Goldstein, D. B., Linares, A. R., Cavalli-Sforza, L. L., & Feldman, M. W. (1995). Genetic Absolute Dating Based on Microsatellites and the Origin of Modernhumans. *Proceedings of the National Academy of Sciences of the United States of America, 92,* 6723-6727. http://dx.doi.org/10.1073/pnas.92.15.6723

Heinila, M. (2012). http://dna.cfsna.net/HAP/Mutation_Rates.htm

Klyosov, A. A. (2009a). DNA Genealogy, Mutation Rates, and Some Historical Evidences Written in Y-Chromosome. I. Basic Principles and the Method. *Journal of Genetic Genealogy, 5,* 186-216.

Klyosov, A. A. (2009b). DNA Genealogy, Mutation Rates, and Some Historical Evidences Written in Y-Chromosome. II. Walking the Map. *Journal of Genetic Genealogy, 5,* 217-256.

Klyosov, A. A. (2009c). A Comment on the Paper: Extended Y Chromosome Haplotypes Resolve Multiple and Unique Lineages of the Jewish Priesthood. *Human Genetics, 126,* 719-724. http://dx.doi.org/10.1007/s00439-009-0739-1

Klyosov, A. A. (2012). Ancient History of the Arbins, Bearers of Haplogroup R1b, from Central Asia to Europe, 16,000 to 1500 Years before Present. *Advances in Anthropology, 2,* 87-105. http://dx.doi.org/10.4236/aa.2012.22010

Klyosov, A. A., & Rozhanskii, I. L. (2012a). Haplogroup R1a as the Proto Indo-Europeans and the Legendary Aryans as Witnessed by the DNA of Their Current Descendants. *Advances in Anthropology, 2,* 1-13. http://dx.doi.org/10.4236/aa.2012.21001

Klyosov, A. A., & Rozhanskii, I. L. (2012b). Re-Examining the "Out of Africa" Theory and the Origin of Europeoids (Caucasoids) in Light of DNA Genealogy. *Advances in Anthropology, 2,* 80-86. http://dx.doi.org/10.4236/aa.2012.22009

Klyosov, A. A., Rozhanskii, I. L., & Ryanbchenko, L. E. (2012). Re-Examining the Out-of-Africa Theory and the Origin of Europeoids (Caucasoids). Part 2. SNPs, Haplogroups and Haplotypes in the Y Charomosome of Chimpanzee and Humans. *Advances in Anthropology, 2,* 198-213. http://dx.doi.org/10.4236/aa.2012.24022

Mendez, F. L., Krahn, T., Schrack, B., Krahn, A.-M., Veeramah, K. R., Woerner, A. E. et al. (2013). An African American Paternal Lineage Adds an Extremely Ancient Root to the Human Y Chromosome Phylogenetic Tree. *American Journal of Human Genetics, 92,* 454-459. http://dx.doi.org/10.1016/j.ajhg.2013.02.002

Poznik, G. D., Xue, Y., Mendez, F. L., Willems, T. F., Massaia, A., Wilson Sayres, M. A. et al. (2016). Punctuated Bursts in Human Male Demography Inferred from 1,244 Worldwide Y-Chromosome Sequences. *Nature Genetics, 48,* 593-599. http://dx.doi.org/10.1038/ng.3559

Rozhanskii, I. L., & Klyosov, A. A. (2011). Mutation Rate Constants in DNA Genealogy (Y Chromosome). *Advances in Anthropology, 1,* 26-34. http://dx.doi.org/10.4236/aa.2011.12005

Zhivotovsky, L. A., Underhill, P. A., Cinnioglu, C., Kayser, M., Morar, B., Kivisild, T. et al. (2004). The Effective Mutation Rateat Y Chromosome Short Tandem Repeats, with Application Tohuman Population-Divergence Time. *American Journal of Human Genetics, 74,* 50-61. http://dx.doi.org/10.1086/380911