A comment on the Paper:

Extended Y chromosome haplotypes resolve multiple and unique lineages of the Jewish Pristhood

By M.F. Hammer, D.M. Behar, T.M. Karafet, F.L. Mendez, B. Hallmark, T. Erez, L.A. Zhivotovsky, S. Rosset, K. Skorecki, Hum. Genet. , Published online 08 August 2009

Journal:	*Human Genetics*
Manuscript ID:	draft
Manuscript Type:	Letter to the Editor
Date Submitted by the Author:	
Complete List of Authors:	Klyosov, Anatole
Keywords:	

Human Genetics

Letters to the Editors

A comment on the Paper:

Extended Y chromosome haplotypes resolve multiple and unique lineages of the Jewish Pristhood

By M.F. Hammer, D.M. Behar, T.M. Karafet, F.L. Mendez, B. Hallmark, T. Erez, L.A. Zhivotovsky, S. Rosset, K. Skorecki, Hum. Genet. , Published online 08 August 2009

Anatole A. Klyosov

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

The cited paper is remarkable with respect to haplogroups and haplotypes typing and the large number of individuals tested, including Cohanim, which is the prime target of the study. However, it is open to critique regarding calculations of time spans to common ancestors of presented series of haplotypes.

First, the authors have employed an inadequate methodolody for the calculations, namely the infamous "population mutation rates" of 0.00069 mutation/marker/generation. It should not have been used in the first place. Much more adequate procedures have been developed lately. Second, the authors have not defined criteria when the "population" and when the "pedigree", or the "genealogical" mutation rates, should be used. As a result, they have applied the "population" mutation rate to "genealogical" haplotype series, which increased their "TMRCA", that is a time span to common ancestors, by about 300%. Third, they have not analyzed genealogical lineages with recent common ancestors (such as 500-800 years bp), therefore, they missed valuable information regarding history of Cohanim in this millennium.

Allow me to explain.

Cohanim J1e*-P58*

Here is the haplotypes tree (it is explained in the last section of this Comment) for Cohanim, haplogroup J1e*, composed from the Hammer et al (2009) data, 22-marker haplotypes, Table S3.

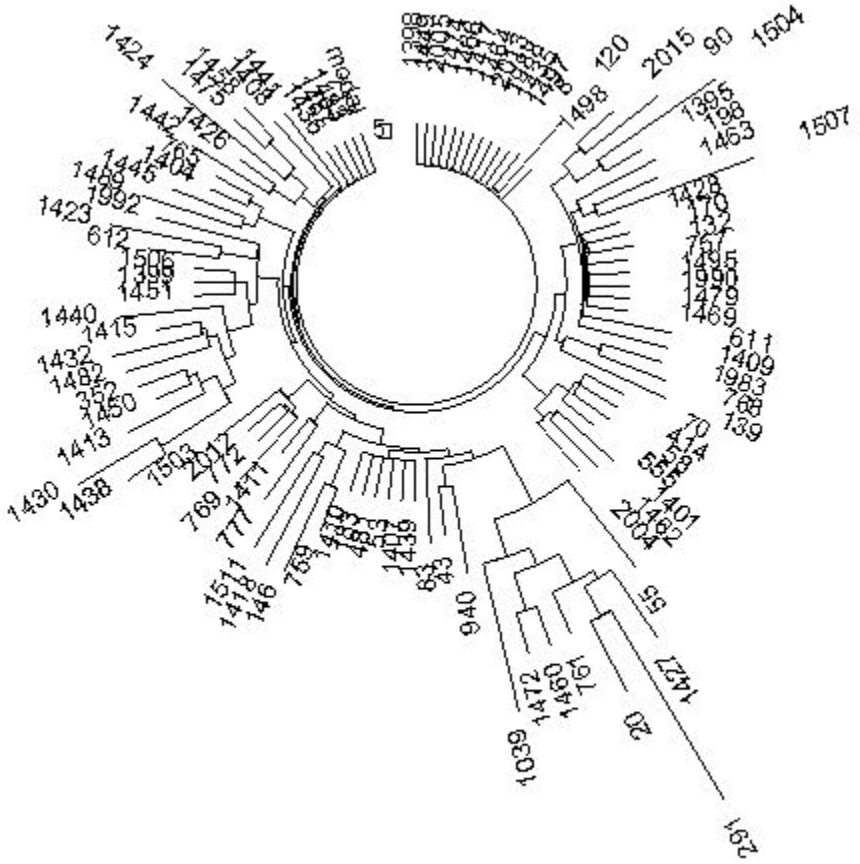

Figure 1. 22-marker haplotype tree for Cohanim, haplogroup J1e*-P58*, for 98 haplotypes (Hammer et al, 2009). Haplotypes numbering corresponds to the numbering in the paper, Table S3

The tree contains 16 base (ancestral) haplotypes (see the last section for definitions) on the top (in the entire order DYS 393, 390, 19, 391, 385a, 385b, 426, 388, 439, 389-1, 393, 389-2, 458, 459a, 459b, 455, 454, 447, 437, 448, 449 and 438):

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

All 98 haplotypes contain 191 mutations from the above base haplotype, which gives – employing the linear method (see the last section of this paper) – 191/98/0.047 = 41 generations (without a corrections for back mutations [Klyosov, 2009]), or 43 generations with the correction, to a common ancestor, that is 1075±130 years bp, the 10-th century AD plus-minus a century.

The authors of the cited paper have obtained 3200±1100 years to a common ancestor for the Cohanim, using the same haplotypes, albeit with a set of 17-marker haplotypes, and 3000±1500 years to a common ancestor using a set of 9-marker haplotypes, and claimed that is it nicely fit to the Biblical description for the origin of the Jewish higher priest. It is not. See below for more details.

Since all 98 haplotypes contain 16 base (identical to each other) 22-marker haplotypes (see above), it gives – employing the logarithmic method (see the last section of this paper) – ln(98/16)/0.047 = 39 generations (without a correction), or 41 generation (with the correction), that is 1025 years bp to a common ancestor for Cohanim J1e* haplotypes.

Since the logarithmic and the "linear" methods give the same dating to the common ancestor, it means that there was indeed just one common ancestor for the whole series of 98 of 22-marker haplotypes (see the last section), who lived about 1075±130 years bp.

The authors of the cited paper have removed some markers for their analysis, namely five markers, DYS 385a, 385b, 459a, 459b and 458, reducing the series from 22 to 17-marker haplotypes. Let us consider if there would be any noticeable change. 98 of 17-marker haplotypes contain 140 mutations from the above base haplotype (less of the five markers), which gives 1175±160 years (bp) for the common ancestor. It is practically the same timespan within the margin of error.

As it was noticed above, the authors of the cited paper have obtained 3200±1100 years to a common ancestor for the Cohanim, using the same set of 17-marker haplotypes, and claimed that is nicely fits to the Biblical description for the origin of the Jewish higher priest. It is not, with 1075±130 and 1175±160 years (bp).

Why such a large difference? The answer is rather obvious. The ratio of 3200 and 1075 years is 3.0. The ratio of the "population" mutation rate (0.00069) to the actual, "genealogical" one (0.00214) is 3.1. The authors of the cited paper have applied the "population" mutation rate to the "genealogical" series of haplotypes. Actually, that "population" mutation rate should have never been applied in the first place. It is ill defined, it is not governed by any criteria when it should be applied and when it should not.

In Table S4 of the cited paper the authors listed 99 of 12-marker Cohanim J1e* (J-P58*) haplotypes. All of them contain 98 mutations from the above base haplotype (in the first 12 markers), which gives 1175±170 years to a common ancestor. It is the same figure obtained above, within the margin of error.

Finally, the authors reduced the 22-marker haplotypes to 9-marker one (the 12-marker haplotype in the FTDNA format less of DYS 385a, 385b and 426), and, using the same "population" mutation rate, have obtained 3000±1500 years to a common ancestor of Cohanim J1e*. In fact, this whole series contain 67 mutations, which gives 67/98/0.018 = 38 generations (40 generations with a correction for back mutations). That is 1000±160 years to a common ancestor. This is again practically the same figure as given above for 22-marker (1075±130 ybp), 17-marker (1175±160 лет), and 12-marker (1175±170 ybp) haplotypes, and is far away from the authors' 3000±1500 ybp for Cohanim J1e*.

In fact, the ancestral haplotype of Cohanim J1e* (in the format previously described)

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

was identified in (Klyosov, 2008a; 2009b) in the form of a 67-marker haplotype (the respective 22 markers are shown below for a comparison)

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

(so-called "the recent extended CMH") of a common ancestor who lived 1,050±190 years bp, around the 10th century. One can see, it is practically the same ancestral haplotype and the same time span to a common ancestor.

Recent DNA-genealogical branches of Cohanim J1e*

The tree (Fig. 1) reveals three principal branches (on top and the left-hand side; on the right-hand side; and at the bottom), with their ancestral haplotypes, respectively :

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

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

12 23 14 10 13 15 11 16 **11** 13 11 30 – 17 8 9 11 11 26 14 21 26 10

each having only one mutation (in bold) from the upper base haplotype, common for the whole tree and for the top-left-hand side. Common ancestors for these all branches lived approximately in the range of 625 to 875 years bp. Just one example – the 27-haplotype branch on the right-hand side has 8 base haplotypes as shown above, with the whole branch having 32 mutations from it. Therefore, a common ancestor of the branch lived ln(27/8)/0.047 = 26 generations bp, that is 650 years bp. The linear method gives 32/27/0.047 = 25 generations bp, that is 625±130 years bp. The fit shows that the branch has one common ancestor indeed, and that he lived around the 14th century AD. Just one mutation per 22-marker haplotypes indicates that all the common ancestors lived within the current millennium, with THEIR common ancestor at the bottom of this time period.

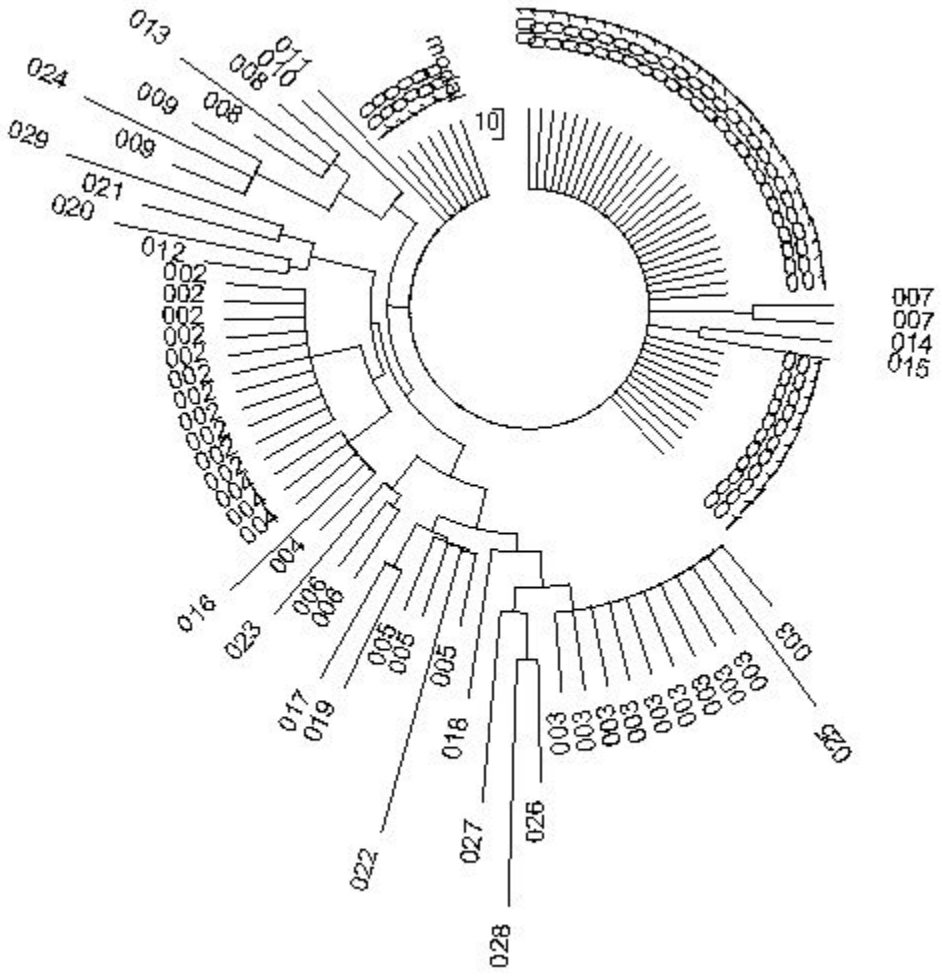

Figure 2. 12-marker haplotype tree for Cohanim, haplogroup J1e*-P58*, for 99 haplotypes (Hammer et al, 2009). Haplotypes numbering follows the order of haplotypes in the paper, Table S4

That the "young" branches are really young ones can be seen from Fig. 2, which shows a haplotype tree for 12-marker haplotypes of the Cohanim of haplogroup J1e*, listed in the cited paper.

Since these haplotypes are lacking ten markers after their first twelve ones, compared to those in Fig. 1, the tree is slightly rearranged in Fig. 2, following the computer program. However, it shows the main tree branches in the haplotype series. Since the branches lost many mutations (in markers 13 through 22), "ages" of their common ancestors are reduced for some branches. For example, the right-hand side branch now contains 48 haplotypes, 43 of which are identical to each other (base haplotypes), and the other contains only seven mutations. This gives $\ln(48/43)/0.022 = 5.0$ generations, that is 125±50 years to a common ancestor, and $7/48/0.022 = 6.6$ generations, that is 165±65 years to a common ancestor. Since the logarithmic and linear methods again give practically identical results, it points to a single (in terms of DNA genealogy) common ancestor for the whole branch. It also shows that 12-marker haplotypes contain much poorer information regarding a time span to a common ancestor compared to that of 22-marker haplotypes. Hence, it is surprising why the authors were reluctant to consider the 22-marker haplotypes of Cohanim, persistently reducing a number of markers in the haplotypes to 17, 12, and even 9.

Cohanim J2a-M410*.

The 22-marker haplotype tree is shown in Fig. 3

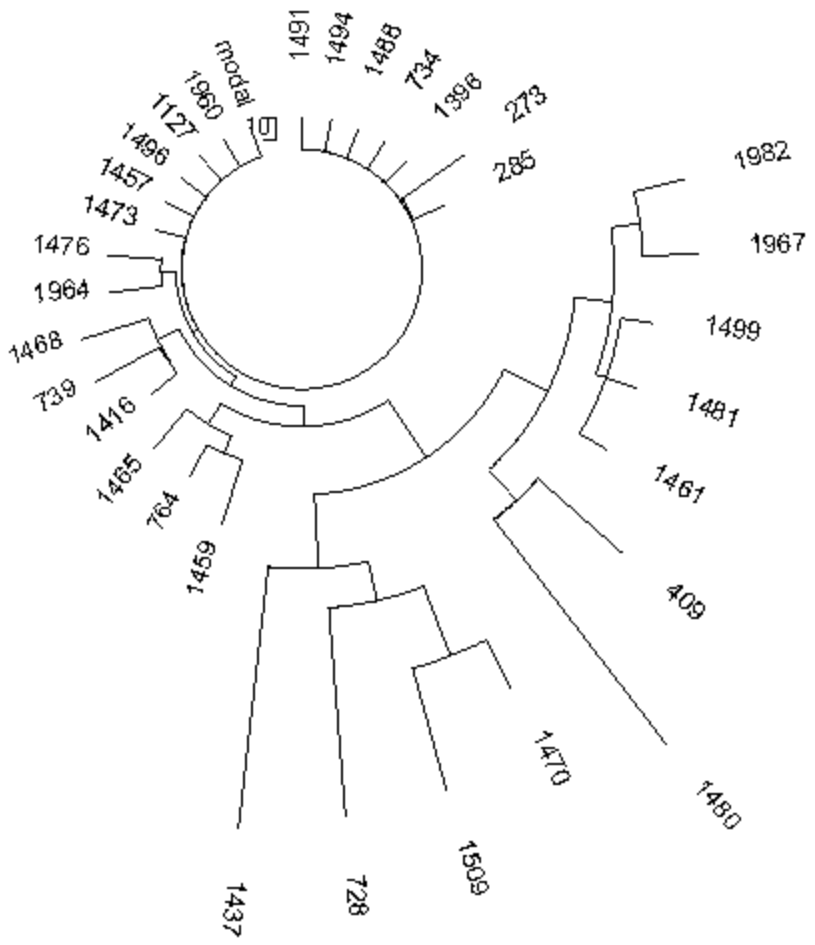

Figure 3. 22-marker haplotype tree for Cohanim, haplogroup J2a-M410, for 31 haplotypes (Hammer et al, 2009). Haplotypes numbering corresponds to the numbering in the paper, Table S3

One can see that the tree consists of two quite different branches, the older one in turn contains two sub-branches. It is clear that to treat the tree as a whole, without subdivision into branches, as it is done in the cited paper, would be incorrect and lead to a phantom common ancestor (though the authors in the paper actually indicated this may be a problem as not all SNPs down stream from M410 were considered). However, the authors did exactly that, they took the whole series of 31 haplotypes, applied the "population" mutation rate and obtained 5900±2000 years to a common ancestor using 17-marker haplotypes, and 4900±1900 years to a common ancestor using 9-marker ones. Again, they have eliminated markers DYS385a,b and DYS459a,b, as well as DYS449, and reduced the 22-marker haplotypes to 17-marker ones. It is obvious that the younger series of haplotypes and the respective recent common ancestor were lost from considerations in the cited paper.

Let us consider the younger branch, containing 20 of 22-marker haplotypes, having 14 mutations from the base haplotype (in the format described above):

12 23 15 10 14 17 11 16 12 14 11 30 – 15 8 9 11 11 25 15 21 31 9

It contains 11 base haplotypes, hence, ln(20/11)/0.047 = 13 generations to a common ancestor. The linear method gives 14/20/0.047 = 15±4 generations to a common ancestor, which is practically the same value obtained by the logarithmic method, within margin of error. Therefore, a common ancestor for the "recent" Cohanim branch lived 375±110 years bp, around the 17th century AD.

The older branch split to two sub-branches. The first one contains five haplotypes of descendants of Cohanim of the former Russian Empire, namely Russia, Ukraine, Belarus, Lithuania and Poland, with only three mutations per 110 markers from the base haplotype:

12 23 15 10 14 17 11 **15** 12 **13** 11 29 – 15 8 9 11 11 **24** 15 21 **32** 9

which gives 325±190 years to their common ancestor, who lived around the 17th century. The alleles in bold are different from ones of the above base haplotype with a common ancestor of 375±110 ybp. This places their common ancestor to about 1500 ybp, around the 6th century AD.

It is exactly the ancestral haplotype of Jewish J2 population identified in a 37-marker format (Klyosov, 2008a) (the respected alleles are shown here for a comparison):

12 23 15 10 14 17 11 15 12 13 11 29 – 15 8 9 11 11 24 15 21 32 9

with a common ancestor who lived 850±260 years bp (12 mutations in 148 markers).

The older sub-branch also splits into two sub-branches, one includes descendants from Iran, Iraq, Argentina and North Africa, with a common ancestor of 3650±830 years bp (24 mutations in four 22-marker haplotypes). The last two branches (Cohanim from the Russian Empire, on the one hand, and Iraq, Iran, Argentina and Africa, on the other) differ by 12 mutations per 22 markers, that places their common ancestor to about 6200 ybp.

If, following the authors of the cited paper, to reduce the amount of markers from 22 to 17, then the "young" branch contains 9 mutations, and a common

ancestor lived 9/20/17/0.00184 = 14 generations ago, that is 350±120 ybp. This is practically the same as 375±110 ybp, determined using 22-marker haplotypes. The Russian Empire Cohanim (see above), whose common ancestor lived 325±190 ybp (22-marker haplotypes), now fits the value of 325±230 ybp (17-marker haplotypes, two mutations in 110 markers), that is just the same value. It is not clear why the authors of the cited paper reduced the haplotypes from 22-market to 17-marker, 12-marker and 9-marker haplotypes. 22-marker ones worked perfectly.

The same was for the ancient branch. 24 mutations in four 22-marker haplotypes (3650±830 ybp) reduced to 16 mutations in four 17-marker haplotypes (3575±960 ybp), which is just the same, albeit with a larger margin of error.

For 9-marker haplotypes the "younger" branch of 20 haplotypes contains 9 mutations (625±220 ybp) which for the 22-marker haplotypes it contains 14 mutations (375±110 ybp). These values are similar and fit within margin of error, and at any rate point at a recent common ancestor of Cohanim J2a.

Finally, for ancient Cohanim J2a branches, 22- and 9-marker haplotypes give similar time spans to a common ancestor: in the first case 24 mutations in four haplotypes results in 3650±830 years to a common ancestor, in the second case the number of mutations reduced to 7, and gave 2425±1200 years to a common ancestor. One can see that the two values are within the margin of error, however, 9-marker haplotypes give much larger error margin.

Cohanim J2e-M12

Fifteen of 22-marker haplotypes of Cohanim J2e-M12 are listed in Table S3 in the cited paper (the authors have reported in Table 1 that they have considered 16 haplotypes). According to the cited paper, a common ancestor of the haplotype lived 12100±4400 years bp as determined with 22-marker haplotypes. The haplotype tree is shown in Fig. 4.

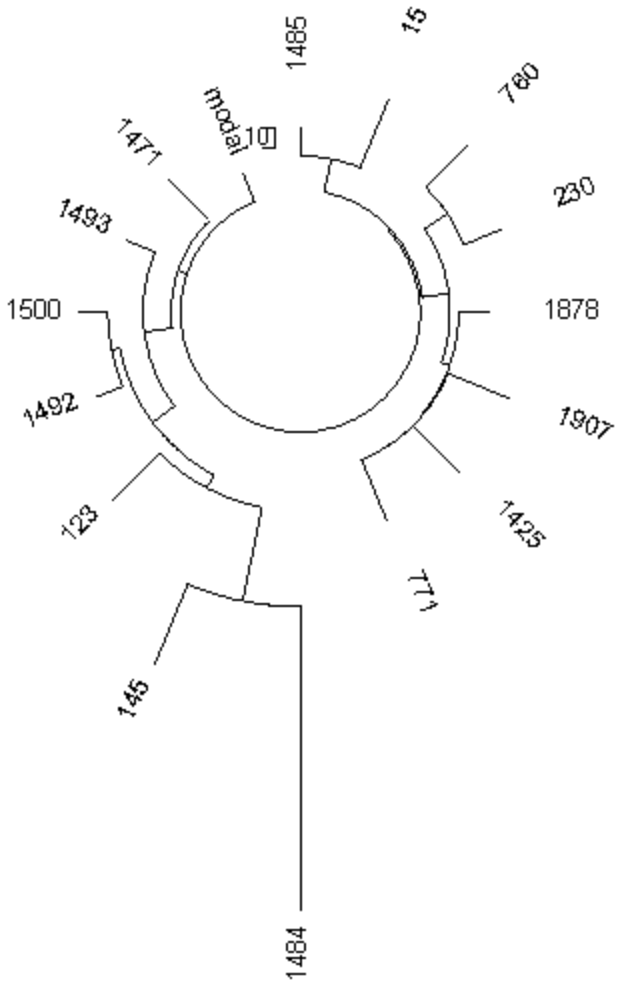

Figure 4. 22-marker haplotype tree for Cohanim, haplogroup J2b-M12, for 15 haplotypes (Hammer et al, 2009). Haplotypes numbering corresponds to the numbering in the paper, Table S3

The tree reveals two branches. The 8-haplotype branch on the right-hand side contains 15 mutations from the following base haplotype (in the format, explained above)

13 24 15 10 15 17 11 15 12 12 11 29 – 19 8 9 8 11 27 16 19 29 9

It gives 15/8/0.047 = 40 generations (not corrected for back mutations), or 42 generations (corrected), that is 1050±290 years to a common ancestor.

The 7-haplotype branch on the left-hand side contains 37 mutations from the base haplotype

12 24 15 10 15 17 11 15 12 12 11 **28** – **18** 8 9 8 11 27 16 19 29 9

It gives 3150±600 years from a common ancestor of the branch. The two base haplotypes are separated by three mutations, which places THEIR common ancestor to about 2950 years bp, which is the common ancestor of the left-hand branch itself. In other words, the right-hand side branch is a daughter branch of the left-hand side one, and a common ancestor of the whole tree of haplotypes of Cohanim J2b-M12 lived 3150±600 years bp.

The authors of the cited paper, in which the time to a common ancestor was determined as 12100±4400 ybp, were in error. They have treated the "father" and the "daughter" branches of the haplotype series as a just a set of haplotypes with a common ancestor outside of the branches, and, besides of it, employed an erroneous "population" mutation rate, as described above. Then, they have considered 17-marker haplotypes, omitting DYS 385a,b, DYS459a,b, and DYS449. These 15 of 17-marker haplotypes in fact contained 9 and 29 mutations in the right-hand side and the left-hand side branch, respectively, which results in 900±310 and 3725±800 years to their common ancestors, respectively. These figures are essentially the same with those obtained with the 22-marker haplotypes, within the margins of error.

Finally, if to reduce haplotypes to 9-marker ones, which would greatly sacrifice accuracy of calculations, the respective time spans to their common ancestors would be 1100±460 and 1500±600 years to a common ancestor, respectively, for the right-hand and left-hand side branches. It just could not possibly be 5500±1900 ybp, obtained in the cited paper with the 9-marker haplotypes. The main reason of the errors, again, is wrong methodology employed in the cited paper for calculations of time spans to common ancestors.

Conclusions

Cohanim J1e*-P58*

A common ancestor of all 99 Cohanim lived 1075±130 years bp, and this timing is reproducible for 9-, 12-, 17-, 22- and 67-marker haplotypes. A much higher values of 3190±1090 and 3000±1500 years bp were obtained in the cited paper (Hammer et al, 2009) by using incorrect methods and incorrect mutation rates. A common ancestor of all the 99 J1e* Cohanim lived around the 10th century AD. There are three main lineages derived from the common ancestor, with their common ancestors who lived approximately between 625 and 875 years bp.

An emphasis of the cited paper on the conclusion that "an extended CMH on the J1e*-P58* background that … is remarkably absent in non-Jews" and having "the

estimated divergence time of this lineage… 3,190±1,090 years" is incorrect regarding the divergence time. It is much more understandable why the lineage originated only 1,075±130 years ago is "remarkably absent in non-Jews".

Cohanim J2a-M410*

Common ancestors of 31 Cohanim lived 325±190 and 375±110 years bp (recent different DNA-lineages), and their common ancestor lived around 1500 ybp. A common ancestor of another branch of Cohanim of this haplogroup lived 3560±830 years bp, and yet another common ancestor lived about 6200 years bp.

Cohanim J2b-M12

Common ancestors of 15 Cohanim lived 3150±600 and 1050±290 years bp, with the second lineage being descendant from the first common ancestor.

Analysis of mutations and their rates

Haplotype trees were constructed using PHYLIP, the Phylogeny Inference Package program (Felsenstein, 2005), as it was explained in detail in (Klyosov, 2008b; 2009a). 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. The farther the haplotypes lies from the wheel, the more mutations they carry compared to the base haplotype and 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". Hence, "base" haplotypes.

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

$B = Ae^{kt,}$

that is

$\ln(B/A) = kt$

where:

B = 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 that 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 only "linear" method is employed. In this method 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, 2008c; 2009a), and a time span to a common ancestor is calculated, either using the Table (Klyosov, 2008c; 2009a), or applying the respective mutation rates. In other words, it is described by the following equation

$$n/N/\mu = t$$

where n is a number of mutations in all N haplotypes in the given series of haplotype, μ is an average mutation rate per haplotype per generation, and t is a number of generations to a common ancestor, corrected for back mutations.

Average mutation rates employed in this paper, are:

For 22-marker haplotypes – 0.047 mutations per haplotype, 0.00214 mutations per marker; this value was obtained by summing up all 22 mutations rates for each marker, according to (Chandler, 2006).

For 17-marker haplotypes – 0.032 mutations per haplotype, 0.00188 mutations per marker; this value was obtained by summing up all 17 mutation rates for each marker, according to (Chandler, 2006).

For 12-marker haplotypes – 0.022 mutations per haplotype, 0.00183 mutations per marker; this value was obtained by summing up all 12 mutation rates for each marker, according to (Chandler, 2006).

For 9-marker haplotypes – 0.018 mutations per haplotype, 0.00200 mutations per marker; this value was obtained by summing up all 9 mutation rates for each marker, according to (Chandler, 2006).

Error margin for time spans to common ancestors are calculated as described in (Klyosov, 2009a), and correspond to the 95% confidence interval.

Corrections for back (reverse) mutations can be done either numerically (Klyosov, 2009a), or using the conversion table (Klyosov, 2008c; 2009a)

References

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

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

Hammer MF, Behar DM, Karafet TM, Mendez FL, Hallmark B, Erez T, Zhivotovsky LA, Rosset S, Skorecki K. (2009) Extended Y chromosome haplotypes resolve multiple and unique lineages of the Jewish Pristhood. Hum. Genet., published online 08 August 2009

Klyosov, AA. (2008a) Origin of the Jews via DNA genealogy. Proceedings of the Russian Academy of DNA Genealogy (ISSN 1942-7484), 1, No. 1, 54-232. http://www.lulu.com/content/2677603

Klyosov AA. (2008b) Basic rules of DNA genealogy (Y chromosome), mutation rates and their calibration. Proceedings of the Russian Academy of DNA Genealogy (ISSN 1942-7484), 1, No. 1, 3-53. http://www.lulu.com/content/2677603

Klyosov, AA. (2008c) A manual for calculations of timespans to common ancestors based on Y-chromosome haplotypes, and a conversion table for reverse mutations. Proceedings of the Russian Academy of DNA Genealogy (ISSN 1942-7484), 1, No. 5, 812-835. http://www.lulu.com/content/4850521

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

Klyosov, AA. (2009b) DNA Genealogy, mutation rates, and some historical evidences written in Y-chromosome. II. Walking the map. J. Genetic Genealogy, in the press.