ARTICLES

_____

DNA Genealogy, Mutation Rates, and Some Historical Evidences Written in Y-Chromosome

Anatole A. Klyosov[1]

Origin of peoples in a context of DNA genealogy is an assignment of them to a particular tribe (all members of which belong to a certain haplogroup) or its branch (a lineage), initiated in a genealogical sense by a common ancestor and an estimation of a time span between the common ancestor and its current descendants. We have developed an approach to kinetics of haplotype mutations in non-recombinant segments of the Y-chromosome and illustrated it with a number of haplotype series related to various populations (timespans to common ancestors shown as years before the present): Basques R1b1b2 (4,050), near thirty of European countries R1a1 (between 4,100 and 4,825), India R1a1 (3,675), the Arabian Peninsula R1a1 (3,750), South India Chenchu R1a1 (3,200 and 350, two lineages), bearers of so-called Cohen Modal Haplotype (CMH) J1 (4,000 and 1,050, two lineages), and CMH J2 (1,400), Arabian CMH J1 (9,000), Bulgarian and Croatian Gypsies H1 (575 and 1,125, respectively), Polynesians C2 (800), South

African Lemba (625), the oldest Balkan R1a1 population (11,600), and Native Americans Q (16,000). Some of these findings are supported by independent estimates.

_____

[1]MIR International, Inc., 36 Walsh Road, Newton, Massachusetts 02459, USA

aklyosov@comcast.net

Introduction

Origin of peoples in a context of DNA genealogy is an assignment of each of them to a particular tribe or its branch (lineage) initiated in a genealogical sense by a particular ancestor who had a base ("ancestral") haplotype. This also includes an estimation of a time span between the common ancestor and its current descendants. If information obtained this way can be presented in a historical context and supported, even arguably, by other independent archeological, linguistic, historical, ethnographic, anthropological and other related considerations, this can be called a success. The first step is to understand our roots in the historical and genealogical context[1-5].

Principles of DNA genealogy have been developed over the last decade and volumes can be written on each of them. The main principles are summarized briefly below[1-18]:

First- Fragments of DNA (haplotypes) considered in this study have nothing to do with genes. Technically, some of them can be associated with gene fragments. However, those arguable associations are irrelevant in context of this study.

Second- Copying of the Y chromosome from father to son results in mutations of two kinds, single nucleotide polymorphisms (SNP) which are certain inserts and deletions in Y chromosomes and mutations in short tandem repeats (STR), which make them shorter or longer by certain blocks of nucleotides. A DNA Y-chromosome segment (DYS) containing an STR is called a locus, or a marker. A combination of certain markers is called a haplotype.

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

Fourth- Haplotypes can be practically of any length. Typically, the shortest haplotype considered in DNA genealogy is a 6-marker haplotype. It is the most common in peer-review publications on DNA genealogy (along with 7-, 9-, 10- and lately 17-marker haplotypes). For example, a typical 6-marker haplotype found in haplogroups J1 and J2 is 14-16-23-10-11-12*.* This is also called the "Cohen Modal Haplotype" (CMH). It shows a number of tandem repeats as alleles in markers of DYS, numbered as DYS 19, 388, 390, 391, 392, and 393. When this or any other haplotype is transmitted from father to son, a mutation in it can occur with a probability of about 1/100 (one mutation per hundred births) for a 6-marker haplotype.

Twelve-marker haplotypes are also often considered in DNA genealogy; however, they are rather seldom presented in professional publications. For example, a common 12-marker is the "Atlantic Modal Haplotype" (in haplogroup R1b1b2 and its subclades):

13-24-14-11-11-14-12-12-12-13-13-2

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

In a similar manner, 17-, 25-, 37-, 43- and 67-marker haplotypes are considered in DNA genealogy. If to operate with an excessive precision, one mutation occurs on average once in: 2,840 years in 6-marker haplotypes, 1,140 years in 12-marker haplotype, 540 years in 25-marker haplotypes, 280 years in 37-marker haplotypes, and 170 years in 67-marker haplotypes. This gives a general idea of a time scale of operations in DNA genealogy.

However, these considerations generally apply only to series of haplotypes, not to pairs of individuals, as supported below.

Fifth- One cannot calculate time to a common ancestor based upon just a pair of haplotypes, particularly short haplotypes. Mutations are ruled by statistics and can be analyzed quantitatively, often semi-quantitatively, using numerous haplotypes. The smaller the number of haplotypes in a set, the less reliable the result. With a pair of haplotypes, one or two mutations can appear in a recent generation. This may mistakenly lead to an addition or subtraction of a thousand years compared with just one generation. In a large set of haplotypes recent mutations are balanced by statistics across all mutations, both recent and old.

Sixth- An average number of STR mutations per haplotype can serve to calculate the time span lapse from the common ancestor for all haplotypes in the set, assuming they all derived from the same common ancestor. That ancestor had a so-called base, or ancestor haplotype. However, very often haplotypes in a set are derived not from one common ancestor, but represent a mix from different ancestors, as theorized below.

Since descendants retain the base haplotype, which is relayed along the lineage from father to son, and mutations in haplotypes occur on average once in centuries or even millennia, then even after 5000 years, descendants retain 23%

of the base, ancestral, unchanged 6-marker haplotype. In 12-marker haplotypes, 23% of them will still serve as base ones after 72 generations or 1,800 years.

Seventh- Haplotypes are often bound to certain territories. In ancient times, people commonly migrated by tribes. A tribe was a group of people typically related to each other. Their males shared the same or similar haplotypes. Sometimes a tribe population was reduced to a few, or even to just one individual, passing though a bottleneck. If the tribe survived, the remaining individual or multiple tribes' people having certain mutations in their haplotypes passed their mutations to the offspring. Many members left the tribe voluntarily or by force as prisoners, escapees, through journeys, or military expeditions. Survivors continued and perhaps initiated a new tribe in a new territory. As a result, a world DNA genealogy map is rather spotty, with each spot demonstrating its own prevailing haplotype, sometimes a mutated haplotype, which deviated from the initial, base, ancestral haplotype. The most popular haplotype in a territory is called a modal haplotype. It often, but not necessarily, represents the ancestor's haplotype.

SNP mutations are practically permanent. Once they appear, they remain. Theoretically, some other mutations can happen at the same spot, in the same nucleotide, changing the first one. However, with millions of nucleotides such an

event is very unlikely. There are more than three million chromosomal SNP's[19], and DNA genealogists have employed a few hundreds of them.

Eighth- People can be assigned to their original tribes of their ancestors not only based on their haplotypes, but based on their SNP's, which in turn lead to their haplogroups and sub-haplogroups, so-called clades. Examples include haplogroups A and B (African, the oldest ones), haplogroups C (mongoloid, as well as a significant part of Native Americans, descendants of mongoloids), haplogroups J (Middle Eastern) with J1 (mainly Jewish and Arabic), and J2 (predominantly Mediterranean including many Jews). Others include haplogroups N (represented in many Siberian peoples, Northern Europeans, as well as some Native Americans who have originated from Siberians) and haplogroup R1b and its subgroups are observed primarily, but not exclusively, in Western Europe, Asia, and Africa. Haplogroup R1a1 dominates in Eastern Europe and Asia, with a minute percentage in the Atlantics, close to 50% (and higher) in Russia and 16% of the population in India. Haplogroup R1a1 also marks some areas in Central Asia particularly in Kirgizstan and Tadzhikistan.

In other words, each male has a SNP from a certain set, recognized by DNA genealogy, which assigns him to a certain ancient tribe. SNP's, typically range

from 40 to 80 thousand years for "old" SNP's to 10 to15 thousand years for "young" ones.

Ninth- The chronological unit employed in DNA genealogy is commonly a generation. A generation in the context of DNA genealogy is an event that occurs four times per century. A "common" generation cannot be defined precisely in years and floats in its duration in real life. Hence, 25 years defines a generation and represents the time span used for the calibration of mutation rates in this study.

Tenth- It is unnecessary to have hundreds or thousands of different haplotypes in order to determine an ancestral (base) haplotype for a large population and calculate a time span from its common ancestor to present time. Alleles in haplotypes do not have random values. Rather, they are typically restricted in rather narrow ranges. Then, after thousands of years descendants of common ancestors for whole populations of the same haplogroup have typically mixed up across their territory. In Europe, for example, one can hardly find an enclave in which people have stayed put in isolation for thousands of years. Even if this was the case, they hardly made a noticeable input with their haplotypes into databases. Last but not least, wherever bearers of haplotypes are hiding, their mutations are "ticking" with the same frequency as the mutations of anyone else.

9

For example, an ancestral (base) haplotype of the Basques of haplogroup R1b1b2, deduced from only 17 of their 25-marker haplotypes follows (in the FTDNA notation):

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

It is very close to a deduced haplotype of a common ancestor of 184 individuals, who belong to haplogroup R1b1b2, subclade U152:

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

Two different alleles (in bold) are equal (as an average in the 17 Basques) to 14.53 and 18.35, respectively, while in subclade U152 they are 14.86 and 18.91, respectively. These two differences in 0.89 mutations total in 25-marker haplotypes point at only approximately ten generations between them. Furthermore, YSearch database contains 12,090 of 25-marker R1b haplotypes (including subclades), which when combined have a minimum of mutational difference with the following base haplotype:

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

It is exactly the same base haplotype as shown above for R1b1b2-U152, and practically the same for that for the Basques of haplogroup R1b1b2.

It is noteworthy that a common ancestor of the Basques 17 haplotypes lived 4,050 years bp, of 184 haplotypes of U152 subclade lived 4,375 years bp, and of all 12,090 haplotypes of R1b lived 4,500 years bp (0.254, 0.269 and 0.275 mutations per marker in 25-marker haplotypes).

Hence, 17 haplotypes lead to approximately the same results in terms of ancestral haplotypes and times when they lived, as more than 12 thousand haplotypes do. The power of DNA genealogy is not in large numbers, though they are always welcomed, but in randomness of haplotype selections.

Eleventh- Unlike languages, religion, cultural traditions, anthropological features, which are often assimilated over centuries and millennia by other languages, cultures, or peoples, haplotypes and haplogroups cannot be assimilated. They can be physically exterminated, but this does not happen often for whole populations. This non-assimilation makes haplogroups and haplotypes practically priceless for archaeologists, linguists, and historians. They not only

stubbornly transcend other assimilations across millennia, but also provide means for calculations of when, and sometimes where, their common ancestors lived.

Twelfth- When possible, the following two calculations should be used to determine estimates of time span to a common ancestor within a reasonable margin of error: (1) a number of mutations in a set of haplotype with respect to the base haplotype, and (2) a fraction of the base haplotypes in the whole set, despite a number of mutations in the set.

If the two numbers are principally different (often by 2-3 times), the set of haplotypes should be dissected by constructing a tree of haplotypes and identifying separate branches of the tree, for example. The same two calculations should be applied.

An explanation of this principle is provided in the Supplementary Information. It also contains Tables 1 and 2, showing average mutation rates per haplotype and per marker for haplotypes of various lengths, and corrections for reverse mutations[20-23]. Haplotype trees are constructed as described in ref.[24]

Basques, haplogroup R1b1b2

Basque DNA Project[25] lists 76 haplotypes which belong to five major

haplogroups (E, G, I, J and R) and their downstream haplogroups and subclades.

More than half of them belong to haplogroup R1b1b2 (M269), which is

considered to be of Western European origin, though it is more conjecture than

proven fact.

Only 17 of those R1b1 Basque haplotypes were presented in the 25-marker

format (Supplementary Information), and the respective haplotype tree is given in

Fig. 1.

One can see that the tree stems from a single mutation coming from a presumably

common ancestor of all 17 individuals in the haplotype set. The base (ancestral)

haplotype can be identified as follows:

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

All 17 haplotypes have 108 mutations from the above base (ancestral) haplotype,

which gives 0.254 mutations per marker on average. Using Tables 1 and 2 of the

Supplementary Information, one can calculate a time span to a common ancestor

of the Basques presented in the haplotype set, which is equal to 162 generations,

or 4,050 years. A consideration of a series of 44 of 12-marker haplotypes of the

Basques gives 0.259 mutations per marker on average and 4,150 years to a common ancestor. That is practically the same time span despite two different series of haplotypes employed for calculations (17 of 25-marker haplotypes and 44 of 12-marker haplotypes).

The Basque ancestral haplotype is also identical to the so-called Atlantic Modal Haplotype in its 12-marker format[26,27]:

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

The data shown in this section suggests that of the R1b1b2 Basques who participated in the Basque Project, their common ancestor was apparently derived from the R1b1b2-M269 subclade (4,450 ybp), and/or R1b1b2-U152 subclade (4,375 ybp), and lived about 4,050 ybp (all the three figures were obtained using 25-marker haplotypes). Since all the tree figures might be considered as being within an error margin, they may represent the same common ancestor. The principal conclusion is that the male Basques living today have rather recent roots of around four thousand years, contrary to legend that proposes they lived approximately 30 thousand years ago. Despite the ancient language, it is very likely that the present day Basques represent a rather recent population, in terms of genealogy. It is very likely that their ancestors had not encountered

14

Neanderthals in Europe, had not been associated with the Aurignacian culture

(34,000-23,000 years bp), nor did they make sophisticated cave paintings in

South of France, Spain, and Portugal. Those people were likely bearers of

haplogroup I2. However, this theory remains to be proven.

This section (along with the Supplementary Information) has essentially shown

how to make calculations and interpret data extracted from a number of mutations

and a number of base haplotypes in haplotype sets. The following section will

follow with principal illustrations and conclusions, without repeating the

methodology.

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

The "mapping" of the enormous territory outlined in this heading reveals that it is

all marked with practically the same ancestral haplotype, which is about $4,500 \pm$

400 years "old". In other words, a common ancestor of the Eurasian R1a1

haplotype lived between 4,100 and 4,900 years ago. An exception is presented

only in the Balkans (Serbia, Kosovo, Macedonia, Bosnia), where a common

ancestor is significantly more ancient, about 11,500 years bp. This will be explored below in this section.

The entire map of base (ancestral) haplotypes and their mutations, as well as "ages" of common ancestors of R1a1 haplotypes in Europe, Asia, and the Middle East show that approximately six thousand years ago bearers of R1a1 haplogroup started to migrate from the Balkans in all directions, spreading their haplotypes. Their closeness should not be surprising, since a recent excavation of 4,600 year-old R1a1 haplotypes[28] revealed their almost exact closeness to present-day R1a1 haplotypes, as it is shown below.

England and Ireland R1a1 haplotypes

As a result of that spread across Europe, R1a1 haplotypes sampled on the British Isles point at their common ancestor who lived 4,250 ± 150 years ago. His base (ancestral) haplotype follows for England:

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

and for Ireland:

13-25-15-**11**-11-14-12-12-10-13-11-30-15-9-10-11-11-**23**-14-**20**-32-12-15-15-16

Their 25-marker haplotypes totaling 50 for England and 40 for Ireland (refer to Supplementary Information) have on average 0.270 and 0.260 mutations per marker, which translates to 4,350 years and 4,150 years to their common ancestors, respectively.

An apparent difference in two alleles between the British and Irish ancestral haplotypes is in fact too insignificant, since the respective alleles are equal to 10.48 and 10.73, and 23.98 and 23.55, respectively. Hence, their ancestral haplotypes are practically the same, within one mutational difference.

"English" and "Irish" R1a1 haplotype trees are shown in Fig. 2 and Fig. 1S (Supplementary Information), respectively.

A branch at the bottom of the tree in Fig. 2, containing seven haplotypes, consists of haplotypes with DYS388=10 and is analyzed separately below.

A DYS388=10 subfamily of North-Western European R1a1 haplotypes

About 20% of both English and Irish haplotypes have a mutated allele in eighth position (DYS388=12➔10), with a common ancestor of that population who lived 3,600 years ago. This mutation is observed only in northern and western Europe, mainly in England, Ireland, Norway, and to a lesser degree in Sweden, Denmark, Netherlands and Germany. In areas further east and south that mutation is rarely observed.

A common ancestor of both DYS388=12 and DYS388=10 populations lived about 5,625 years before the present in another location, presumably in the Balkans, more than a thousand years before bearers of that mutation arrived to northern and western Europe. This mutation was passed over the generations up to the present time. Refer to the Supplementary Information for calculations.

Scotland R1a1 haplotypes

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

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

and the "Scottish" R1a1 common ancestor lived 4,300 years before present (0.265 mutations per marker in 31 of 25-marker haplotypes).

Germany R1a1 haplotypes

A 67-haplotype sample in Germany revealed the following ancestral haplotype:

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

There is an apparent mutation in the third allele from the left (in bold) compared with the Isles ancestral haplotypes, however, it equals to 15.48 for England, 15.33 for Ireland, 15.26 for Scotland, and 15.84 for Germany. All 67 25-marker haplotypes contain 488 mutations, that is 0.291 mutations per marker on average. It corresponds to 4,700 years from a common ancestor in the German territory.

These results are supported by the very recent data on excavation of R1a (SNP SRY10831.2) and presumably R1a1 haplotypes near Eulau, Germany[28]. The 4,600 old haplotypes follow with dating performed using strontium isotope analysis:

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

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

Norway and Sweden R1a1 haplotypes

Norwegian and Swedish ancestral R1a1 haplotypes are principally the same, and both look exactly as the German 25-marker ancestral haplotype. Their 21-and 20-haplotype sets contain 0.265 and 0.262 mutations per marker, respectively, which gives 4,300 and 4,250 years to their common ancestors. This, of course, is the same time span within an error margin.

The Swedish haplotype tree (not shown) reveals a separate, "younger" branch, which contains only 0.165 mutations per marker, resulting in 2,475 years from the common ancestor, who, in theory, lived not long before the beginning of AD.

Poland, Czech, and Slovak R1a1 haplotypes

These ancestral haplotypes are very similar to each other, having only insignificant deviation in DYS439 (9th marker from the left), which is 10.43 in the Polish base haplotype of 44 haplotypes total.

13-25-16-10-11-14-12-12-**10**-13-11-30-16-9-10-11-11-23-14-20-32-12-15-15-16

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

13-25-16-10-11-14-12-12-**11-**13-11-30-16-9-10-11-11-23-14-20-32-12-15-15-16

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

These haplotypes have 0.282 and 0.259 mutations per marker on average in their 25-marker haplotypes, which results in 4,550 and 4,100 years from their common ancestors, respectively.

The European R1a1 haplogroup (small haplotype sets across Europe)

Many countries are represented with just a few haplotypes in data bases. I have collected 36 of 25-marker R1a1 haplotypes provided into YSearch database by residents of Denmark, Netherlands, Switzerland, Iceland, Belgium, France, Italy,

Lithuania, Romania, Albania, Montenegro, Slovenia, Croatia, Spain, Greece, Bulgaria and Moldavia. (refer to Fig. 2S in Supplementary Information).

The tree does not show any noticeable anomalies and points at just one common ancestor for all 36 individuals, who had the following base haplotype:

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

The ancestral haplotype match is exactly the same as those in Germany, Russia (see below), and has quite insignificant deviations from all other ancestral haplotypes considered above, within fractions of mutational differences. All the 36 individuals have 248 mutations in their 25-marker haplotypes, which gives 0.276 mutations per marker on average and corresponds to 4,450 years to a common ancestor. This is quite a common value for European R1a1 population.

Russia and Ukraine R1a1 haplotypes

The haplotype tree containing 44 of 25-marker haplotypes collected over 10 time zones from the Western Ukraine to the Pacific Ocean and from the northern tundra to Central Asia (Tadzhikistan and Kirgizstan) is shown in Fig. 3.

The ancestral (base) haplotype for the haplotype tree is

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

It is exactly the same ancestral haplotype as that in Germany, and it has only one insignificant deviation from the ancestral haplotype in England, which has the third allele (DYS19) 15.48, which in Russia/Ukraine it is 15.89. There are similar insignificant deviations with the Polish and Czechoslovakian base haplotypes, at 13th and 18th markers (DYS458 and DYS447, respectively), within 0.2-0.5 mutations. This points at a chronological difference between their common ancestors of only several generations.

All 44 haplotypes contain 326 mutations from the base haplotype, that is 0.296 mutations per marker or 4,825 years from a common ancestor.

R1a1 haplotypes of individuals who considered themselves of Ukrainian and Russian origin, present a practically random mix in the haplotype tree. The haplotype tree contains two local Central Asian haplotypes (a Tadzhik and a Kyrgyz, haplotypes 133 and 127, respectively), as well as a local of a Caucasian Mountains Karachaev tribe (haplotype 166). They did not show any unusual

23

deviations from other R1a1 haplotypes. Apparently, they are derived from the same common ancestor as are all other individuals of the set.

India R1a1 haplotypes

The YSearch database contains 15 R1a1 haplotypes from India. Their ancestral haplotype follows:

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

The only one apparent deviation in DYS389-2 (shown in bold) is related to an average alleles equal to 30.67 in Indian haplotypes, and 29.84 in Russian ones. In another accepted format, these alleles are equal to 17.40 and 16.77, respectively that is within one mutational unit. If rounded, they both would be equal to 17.

All 15 Indian R1a1 haplotypes contain 88 mutations, that is 0.235 mutations per marker. It is significantly less than 0.296 mutations per marker in the Russian haplotypes and corresponds to 3,675 years from a common ancestor. This is almost 1,150 years "younger" compared to the Russian one.

Archaeological studies have been conducted since the 1990's in the South Ural's

Arkaim settlement and have revealed that the settlement was abandoned 3,600

years ago. The population apparently moved to Northern India. That population

belonged to Andronovo archaeological culture. Excavations of some sites of

Andronovo culture showed that eight inhabitants out of nine shared R1a1

haplogroup and haplotypes[29] as follows, dating between 5,500 and 1,800 years

bp:

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

In this example, alleles which have not been deciphered are replaced with letters.

One can see that the ancient R1a1 haplotype closely resembles the Russian (as

well as the other R1a1) ancestral haplotypes.

This provides rather strong evidence that the R1a1 tribe migrated from Europe to

the East between 5,000 and 3,600 years bp. The pattern of this migration is

exhibited as follows: 1) the descendants who live today share a common ancestor

of 4,825 years prior, 2) the Andronovo archaeological complex of cultures in

North Kazakhstan and South and Western Siberia dates 4,300 to 3,500 years bp,

3) the reach to South Ural some 4,000 years bp, is where they built Arkaim,

Sintashta (contemporary names) and the so-called "a country of towns" on South

Ural around 3,800 ybp, 4) by 3,600 ybp they abandoned the area and moved to India under the name of Aryans. There they have "established" a common ancestor of R1a1 Indians 3,675 years bp, which chronologically corresponds to the events on the South Ural.

R1a1 haplotypes, the Arabian peninsula

Sixteen R1a1 10-marker haplotypes from Qatar and United Arab Emirates have been recently published[30]. They split into two branches, with base haplotypes

13-25-15-11-11-14-X-Y-10-13-11-**30**

13-25-16-11-11-14-X-Y-10-13-11-**31**

which differ by only one mutation, marked in bold. The first haplotype is the base one for seven haplotypes with 13 mutations in them, on average 0.186 mutations per marker, which gives 2,300 years to a common ancestor in the 3rd century BC. The second haplotype is the base one for nine haplotypes with 26 mutations in them, an average 0.289 mutations per marker or 3,750 years to a common ancestor. Since a common ancestor of R1a1 haplotypes in Armenia and Anatolia

lived 4,525 and 4,050 years bp, respectively[31], it corresponds well with 3,750 years bp in the Arabian peninsula.

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

A series of 67 haplotypes of haplogroup R1a1 from the Balkans was published[32-34]. They were presented in a 9-marker format only (Fig. 4), which is a common case in research publications by population geneticists. One can see a remarkable branch on the lower left side of the tree which stands out as an "extended and fluffy" one. These are typically features of a very old branch compared with others on the same tree. Also, a common feature of old haplotype trees is that they are typically "heterogeneous" ones and consist of a number of branches.

An analysis of the main branches of the tree in Fig. 4 is given in Supplementary Information, where thirteen haplotypes of the most extended and fluffy branch are listed. Those haplotypes collected in Serbia, Kosovo, Bosnia, and Macedonia[32-34] produced the following base haplotype:

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

27

Some deviations from typical ancestral (base) East-European haplotypes are shown in bold, however, the third allele (DYS19) is the same as that in the Atlantic and Scandinavian R1a1 base haplotypes. A mutational analysis results in 11,425 years from a common ancestor, obtained using the "linear method" with correction for reverse mutations, 11,650 years using the quadratic method and an averaged base haplotype, and 11,650 years using the all-permutation quadratic method.

The obtained data suggest that the first bearers of R1a1 haplogroup lived in the Balkans (Serbia, Kosovo, Bosnia, Macedonia) about 11,600 years bp. It is unknown whether R1a1 appeared in the Balkans presumably from R1 or R1a or arrived from a yet unknown location. It was found[27] that haplogroup R1b appeared from R1 about 16,000 years bp apparently in Asia.

The data shown above suggests that only about 6,000-5,000 years bp bearers of R1a1 began to mobilize and migrate to the west toward the Atlantics, to the north toward the Baltic Sea and Scandinavia, to the east to the Russian plains and steppes, to the south to Asia Minor, the Middle East, and far south to the Arabian Sea. All of those local R1a1 haplotypes point at their common ancestors who lived from 4,800 to 4,000 years bp. On their way through the Russian plains and steppes the R1a1 tribe presumably sat up the Kurgan archaeological culture,

apparently domesticated the horse, advanced to Central Asia and left the "Aryan

population" which dated to 4,500 years bp. They then moved to the Ural

mountains about 4,000 years bp and migrated to India as the Aryans circa 3,600-

3,500 years bp. Presently, 16% of the male Indian population, or approximately

100 million people, bear R1a1 haplogroup's SNP mutation, with their common

ancestor of 3,675 years bp. The current Indian R1a1 haplotypes are practically

indistinguishable from Russian, Ukrainian, and Central Asian R1a1 haplotypes,

as well as from many West and Central European R1a1 haplotypes. They

correspond closely to the Indo-European language family.

The next section points at some trails of R1a1 in India, an "Aryan trail".

The Chenchu R1a1 haplotypes

The Chenchus, an australoid tribal group from southern India, bear R1a1

haplogroup in 11 of 41 individuals tested (or 27%).[35] It is tempting to associate

this with the Aryan influx into India which allegedly occurred some 3,600-3,500

years bp. However, questionable calculations of time spans to a common ancestor

of R1a1 in India[35-37] using methods of population genetics rather than those of

DNA genealogy have precluded an objective and balanced discussion of the

events and their consequences.

29

Eleven R1a1 haplotypes of the Chenchus[35] do not provide good statistics; however, they can allow a reasonable estimate of a time span to a common ancestor for these 11 individuals. Logically, if these haplotypes are more or less identical, with just a few mutations in them, a common ancestor lived within a thousand of years bp. Conversely, if these haplotypes are all mutated, and there is no base (ancestral) haplotype among them, a common ancestor lived thousands years bp. Even two base (identical) haplotypes among 11 would tentatively give $\ln(11/2)/0.0088 = 194$ generations, which, corrected to back mutations, would result in 240 generations, or 6,000 years to a common ancestor. If eleven of the 6-marker haplotypes are all mutated, it means that a common ancestor lived earlier than 6 thousand years bp. Hence, even with such a poor set of haplotypes one can obtain useful and meaningful information.

Eleven Chenchu haplotypes have as many as seven identical (base) 6-marker haplotypes (in the format of DYS 19-388-290-391-392-393, commonly employed in scientific publications):

16-12-24-11-11-13

They are practically the same as those common East European ancestral haplotypes considered above, if presented in the same 6-marker format:

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

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

Data in Supplementary Information demonstrates that the Chenchu R1a1 haplotypes represent two lineages, one 3,200 years old and the other only 350 years old, starting from the 17th century AD. Considering that the Aryan (R1a1) wave to northern India took place about 3,600-3,500 years bp, it is quite plausible to refer the appearance of R1a1 in the Chenchu by 3,200 ybp to the Aryans.

The origins of the influx of Chenchu R1a1 haplotypes in the 17th century are found in this passage excerpted from ref.[35]: "Chenchus were first described as shy hunter-gatherers by the Mohammedan army in 1694".

Native Americans haplotypes of haplogroup Q

Let us consider much more distant time periods to further examine and justify the timing methods of DNA genealogy developed in this study. One hundred

seventeen haplotypes of Native American Indians, haplogroup Q-M3, were

published[38], and a haplotype tree, shown in Fig. 5, was developed based upon

their data.

The tree contains 31 identical (base) haplotypes and 273 mutations from that

"base" haplotype. It is obvious that the haplotypes in the tree descended from

different common ancestors, since 31 base haplotypes out of 117 total would give

$\ln(117/31)/0.0088 = 151$ generations to a common ancestor, though 273 mutations

in all 117 haplotypes would give 265 generations (without corrections for back

mutations).

Indeed, the haplotype tree points to at least six different common ancestors for

the 117 individuals. Four of them turned out to be quite recent ancestors, who

lived within the last thousand years (the analysis is provided in Supplementary

Information). The "oldest" brunch turned out to be one in the upper right area,

containing 11 haplotypes. Their common ancestor lived 12,125 years bp.

Furthermore, a common ancestor for the whole haplotype tree lived 16,300 years

bp (Supplementary Information).

Hence, a common ancestor of several groups of individuals among Native

Americans of haplogroup Q-M3 and having largely varied haplotypes, lived

approximately 16 thousand years bp. This dating is in line with many independent

data of archaeological, climatological, and genome study origins. Some

researchers refer the peopling of the Americas to the end of the last glacial

maximum, approximately 23,000 to 19,000 years ago and suggest a strong

population expansion started approximately 18,000 and finished 15,000 years

bp.[39] Others refer to archaeology data of Paleo-Indian people between 11,000 to

approximately 18-22,000 years bp[40-43]. At any rate, the time span of 16,000 years

ago corresponds well with those estimates.

The Cohen Modal Haplotype of haplogroups J1 and J2 (the Jewish and Arabic haplotypes)

-- CMH, Haplogroup J1

The "Cohen Modal Haplotype" (CMH) was introduced[13] ten years ago to

designate the following 6-marker haplotype:

14-16-23-10-11-12

Further research showed that this haplotype presents in both J1 and J2 haplogroups. In haplogroup J1 it splits into two principal lineages[20], with base (ancestral) haplotypes:

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

and

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

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

In haplogroup J2 the "Cohen Modal Haplotype", if to follow its 6-marker notation, has the following 12-marker base haplotype:[20]

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

It differs from both J1 CMHs by three and one mutations.

It is still unknown to what extent each of the three 12-marker haplotypes can be attributable to actual cohens (cohanim) or higher priests. An answer may be obtained in principle only, taking into account complex history of cohens in the Jewish society. However, DNA genealogy can answer when a common ancestor

of each of the family of haplotypes (shown as base haplotypes) lived, and how different those haplotypes expressed when their extended formats are considered, such as 25-, 37- and 67-marker haplotypes (analysis and haplotype trees are presented in Supplementary Information).

In short, when only haplotypes from individuals of the Jewish descent are analyzed, the J1 "Cohen Modal Haplotype" exists as two different lineages, a "recent CMH", with a common ancestor 1,400 years bp, and another "older CMH", with a common ancestor 2,400 years bp. A common ancestor of these two lineage lived 4,225 years bp (refer to Supplementary Information).

In order to detail this information, as many as 85 of 37-marker "Cohen" haplotypes were collected from both Jewish and non-Jewish descendants, and 33 haplotypes among them contained as many as 67 markers. Fig. 6 shows a 37-marker haplotype tree, in which all haplotypes belong to the CMH series, that is all of them have the 14-16-23-10-11-12 string in their haplotypes (shown here in the 6-marker format). The left-hand side represents the "recent CMH", the branch at the lower right represents the "older CMH" haplotypes.

The "older CMH" 22-haplotype branch contains 126 mutations in 25-marker format and 243 mutations in 37-marker format, which gives 0.229 and 0.299

mutations per marker, respectively. Thirty seven-marker trees include a number of "fast" markers, hence, the higher figure. This results in 3,575 and 3,525 years from a common ancestor, respectively, on average $3,550 \pm 25$ years. However, such high precision is coincidental. A realistic error margin should be within 200 to 300 years.

The "recent CMH" branch contains 0.069 mutations and 0.110 mutations per marker in 25- and 37-marker haplotypes, which corresponds to 975 and 1175 years to a common ancestor, respectively, on average $1,075 \pm 100$ years bp.

Maximum high resolution CMH haplotypes are shown on the 67-haplotype tree (Fig 3S, Supplementary Information). The tree splits into two quite distinct branches: a recent one on the right and an older one on the left. Again, these are two principal "Cohen Modal Haplotype" branches, each one with its own common ancestor, who lived 3,000 years apart. A common ancestor of the "older CMH", calculated from the two branches in Fig. 3S, however using 25- and 37-marker haplotypes for calculations, lived 4,150 and 3,875 years bp, on average $4,000 \pm 200$ years bp. A common ancestor of the "recent CMH", calculated from 25- and 37-marker haplotypes, lived 950 and 1,150 years to a common ancestor, respectively, on average $1,050 \pm 150$ years bp, in the 9th to the 11th century. This coincides with the Khazarian times; however, it would be a stretch to claim so.

36

According to data provided in databases, two-thirds of bearers of the "older CMH" (10 individuals of the 16) in the respective 67-marker branch denied Jewish heritage. They are descendants of people who lived in Italy, Cuba, Lebanon, Puerto-Rico, Spain, England, and France (Basque). That may explain why the "older CMH" haplotype differs in three alleles in the first 25 markers from the Jewish oCMH, shown earlier. On the contrary, 16 out of 17 haplotypes in the recent CMH branch claimed their Jewish heritage, and several claimed themselves to be descendants of Cohens. Their base haplotype is identical with the Jewish rCMH shown earlier.

To verify this concept, three haplotypes of inhabitants of the Arabian Peninsula with typical Arabic names, having the "CMH" haplotypes (see Supplementary Information) were added to a set of haplotypes shown in Fig. 6. A source of these haplotypes is described in the next Section. All three Arabic haplotypes joined the lower branch on the right which is predominantly non-Jewish (Fig. 6). All 25 haplotypes in the branch contained 162 mutations, which gives 4,125 years to a common ancestor for a collection of 88 Jewish and non-Jewish haplotypes, all of them belonging to the "CMH" family of haplotypes. This time is close to that of the legendary Biblical split into the Jewish and the Arabic lineages, whether it is applicable or not to the results of this study.

The following Section demonstrates that the "CMH" in fact appeared as long as 9,000 years bp or earlier on the Arabian Peninsula. The above time span of about 4,100 and 1,000 years bp was generated as a result of drifts of haplotype bearers from the Arabian Peninsula to the Middle East and further to the north. We can neither prove nor disprove as yet that the "recent CMH" appeared in the Khazar Kaganat between 9th and 11th centuries. At any rate, the bearer of the base "recent CMH" became a common ancestor to perhaps millions of present-day bearers of this lineage.

-- *The Arabic "CMH", the Arabian Peninsula*

Obviously, the name "Cohen Modal Haplotype" was a misleading one. Though, by the end of the 1990's it had certainly attracted attention to DNA genealogy. Even as a "modal" haplotype it is not exclusively associated with a Jewish population. A haplotype tree of Arabs from the Arabic Peninsula is shown in Supplementary Information (Fig. 4S). The tree was composed from 19 of 37-marker haplotypes of haplogroup J1 listed in the "Arabian Peninsula YDNA Project".[44]

There are two branches of one tree, one having the "Cohen Modal Haplotype" (in its 6-marker format) as an ancestral (base) haplotype:

12-23-14-10-12-18-11-16-11-13-11-30-18-8-9-11-11-25-14-20-26-12-13-15-16-

10-10-19-22-15-14-18-18-33-35-12-10

Seventy three mutations in the 6-haplotype branch, or 0.532 mutations per marker on average, gives 9,000 years to a common ancestor.

The second, a 7-haplotype branch (Fig. 4S), is certainly a younger one, with an average number of 0.154 mutations per marker, which gives 2,300 years to a common ancestor. It turns out to be a direct descendant of the older common ancestor (Supplementary Information).

It seems that the "Cohen Modal Haplotype" was likely an ancestral haplotype some 9,000 years bp for both historical inhabitants of the Arabian Peninsula and the Jews, along with the Arabs of the Middle East. About 4,000 years ago the establishing Jewish population carried this "modal haplotype" although it is a slightly drifted structure, which was coined as the "older CMH". By the 7th century AD, the "recent CMH" split from the "older CMH" and became the

ancestral haplotype for a separate Jewish lineage within haplogroup J1. If to consider only "CMH" haplotypes within this population, a common ancestor can be identified who lived about 1,000 years bp.

-- CMH, Haplogroup J2

The "Cohen Modal Haplotype" of haplogroup J2 represents a rather compact group of haplotypes with a recent ancestor who lived 1,450 years bp (see Supplementary Information).

The compact group of ten haplotypes sitting rather close to the trunk of the tree (see Fig. 5S), which indicates their "younger age", represents the J2-CMH. Their 37-marker base haplotype is as follows:

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

All 10 haplotypes contain 25 and 44 mutations in the 25- and 37-marker format, respectively. This gives 1,450 and 1,300 years to a common ancestor or 1,375 $\pm$ 100 years bp on average who lived between the 6th and 8th centuries.

The J2-CMH base haplotype differs from J1 "recent" and "older" base (ancestral) haplotypes by 29 and 25 mutations, respectively, in their 37-marker haplotypes. This corresponds to 11,800 and 9,600 years of mutational difference, respectively. Clearly, J1- and J2-CMH represent quite distant lineages. After all, they belong to two different haplogroups.

The South African Lemba haplotypes

Lemba is a South African tribe whose people live in Limpopo Province, Zimbabwe, Malawi, and Mozambique. A list of 136 Lemba haplotypes was published[14], and the authors alluded that some Lemba belong to the CMH Jewish lineage. We will demonstrate that it is very unlikely.

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

15-12-21-10-11-13

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

Another 23 Lemba who were tested had the following base haplotype (a haplogroup was not identified, as well as of any haplogroups for the published haplotypes[14]):

14-12-23-10-15-14

All 23 haplotypes had only 16 mutations, which gives 2,150 years to a common ancestor for these individuals.

There were a few scattered Lemba haplotypes, apparently from different unidentified haplogroups, and finally there were 57 haplotypes of apparently haplogroup J, which in turn split into three different branches (Fig. 7). Three base haplotypes, one for each branch, are shown below:

14-16-24-10-13-12
14-15-24-10-11-12
14-16-23-10-11-12

The first one represents 16 identical haplotypes (the upper right area in Fig. 7), which obviously came from a very recent common ancestor. As one can see from

the haplotype tree, none of these haplotypes is mutated. Its common ancestor should have lived no more than a few centuries ago.

The second one, being a base haplotype for a 26-haplotype branch on the left-hand side in Fig. 7, is a rather common haplotype in the Arabic world, and belongs likely to haplogroups J and J2. The branch contains 21 mutations, which gives 2,550 years to a common ancestor, who most likely lived in the middle of the first millennia BC. It is clearly not the "Cohen Modal Haplotype" and differs from the last by two mutations.

The third base haplotype, which is the CMH in its 6-marker format, supports a branch of 15 haplotypes on the lower right-hand side. Twelve of those CMH haplotypes are identical to each other and form a flat branch. There are no mutations in them, and they must have come from a very recent ancestor of only a few centuries ago. From a fraction of the base haplotype, their common ancestor lived only ln (15/12)/0.0088 = 25 generations ago, or about 625 years bp in the 14th century.

The three mutated haplotypes in this series are quite different from the CMH, and apparently do not belong to the same group of haplotypes. All of them have two or four mutations from the CMH:

14-23-14-10-11-12

14-23-14-10-11-12

16-24-14-10-11-12

Unfortunately, more extended haplotypes are not available. It is very likely that they are rather typical mutated Arabic haplotypes. Besides, it is not known to which haplogroup they belong, J1 or J2.

Obviously, to call the Lemba haplotypes the "Cohen haplotype" is a stretch. They could have been Jewish and originated just a few centuries ago, or they could have been Arabic. Hence, the so-called "Cohen Modal Haplotype" in the "Black Jews of Southern Africa" has nothing to do with an ancient history of either the Lemba or the Jewish people. It is a rather recent acquirement.

A conclusion

This paper and its Supplemental Information present a rather consistent concept of dating both recent and distant common ancestors based upon appearance of haplotype trees, verify a haplotype set in terms of a singular common ancestor for the set or a multiplicity of them applying the "logarithmic method", and calculate a time span to a common ancestor corrected for reverse mutations. Obviously,

44

time spans to common ancestors refer to those ancestors whose descendants survived and present their haplotypes and haplogroups for testing in the present. Naturally, their tribes themselves could have appeared earlier, since it is very likely that in many cases offspring and/or descendants did not survive.

Amazing as it is, all of it is written in the DNA of each of us, the survivors. It is a scribble on the cuff of the DNA. Though there is a deeper meaning to these scribbles. If we look at them for a single individual, without comparisons to others, they do not say much. They represent just a string of numbers. However, when compared with those in other people, these scribbles start to tell a story. These collective stories are about origins of mankind, appearances of tribes, their migrations, about our ancestors, and their contributions to current populations. This study advances quantitative descriptions in the field of DNA genealogy.

References

1. Nei, M. Genetic support for the out-of Africa theory of human evolution. *Proc. Natl. Acad. Sci. USA*, **92**, 6720-6722 (1995)

2. Karafet, T.M., Zegura, S.L., Posukh, O., Osipova, L., Bergen, A., et al. Ancestral Asian source(s) of New World Y-chromosome founder haplotypes. *Am. J. Hum. Genet*. **64**, 817-831 (1999).

3. Underhill, P.A., Shen, P., Lin, A.A., Jin, L., Passarino, G., et al. Y chromosome sequence variation and the history of human populations. *Nature genetics* **26**, 358-361 (2000).

4. Semino, O., Passarino, G., Oefner, P.J., Lin, A.A., Arbuzova, S., et al. The genetic legacy of Paleolithic *Homo sapiens sapiens* in extant Europeans: a Y chromosome perspective. *Science* **290,** 1155-1159 (2000)

5. Weale, M.E., Yepiskoposyan, L., Jager, R.F., Hovhannisyan, N., Khudoyan, A., et al. Armenian Y chromosome haplotypes reveal strong regional structure within a single ethno-national group. *Hum. Genet*. **109**, 659-674 (2001)

6. Goldstein, D.B., Linares, A.R., Cavalli-Sforza, L.L. and Feldman, M.W. Genetic absolute dating based on microsatellites and the origin of modern humans. *Proc. Natl. Acad. Sci*. US, **92**, 6723-6727 (1995).

7. Goldstein, D.B., Linares, A.R., Cavalli-Sforza, L.L. and Feldman, M.W. An evaluation of genetic distances for use with microsatellite loci. *Genetics* **139**, 463-471 (1995).

8. Zhivotovsky, L.A. and Feldman, M.W. Microsatellite variability and genetic distances. *Proc. Natl. Acad. Sci. US*, **92**, 11549-11552 (1995).

9. Jobling, M.A., and Tyler-Smith, C. Fathers and sons: the Y chromosome and human evolution. *TIG*, **11**, 449-456 (1995).

10. Takezaki, N. and Nei, M. Genetic distances and reconstruction of phylogenic trees from microsalellite DNA. *Genetics* **144**, 389-399 (1996).

11. Heyer, E., Puymirat, J., Dieltjes, P., Bakker, E., and de Knijff, P. Estimating Y chromosome specific microsatellite mutation frequencies using deep rooting pedigrees. *Human Mol. Genetics* **6**, 799-803 (1997)

12. Skorecki, K., Selig, S., Blazer, S., Bradman, R., Bradman, N., et al. Y chromosomes of Jewish Priests. *Nature* **385,** 32 (1997)

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

14. Thomas, M.G., Parfitt, T., Weiss, D.A., Skorecki, K., Wilson, J.F., et al. Y Chromosomes traveling South: the Cohen Modal Haplotype and the origin of the Lemba – the "Black Jews of Southern Africa". *Am. J. Hum. Genet*. **66**, 674-686 (2000).

15. Nebel, A., Filon, D., Weiss, D.A., Weale, M., Faerman, M., et al. High-resolution Y chromosome haplotypes of Israeli and Palestinian Arabs reveal geographic substructure and substantial overlap with haplotypes of Jews. *Hum. Genet*. **107**, 630-641 (2000).

16. Kayser, M., Roewer, L., Hedman, M., Henke, L., Hemke, J., et al. Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs. *Am. J. Hum. Genet*. **66**, 1580-1588 (2000).

17. Hammer, M.F., Redd, A.J., Wood, E.T., Bonner, M.R., Jarjabazi, H., et al. Jewish and Middle Eastern non-Jewish populations share a common pool of Y-chromosome biallelic haplotypes. *Proc. Natl. Acad. Sci. US*. **97**, 6769-6774 (2000).

18. Nebel, A., Filon, D., Brinkmann, B., Majumder, P.P., Faerman, M. and Oppenheim, A. The Y chromosome pool of Jews as part of the genetic landscape of the Middle East. *Am. J. Hum. Genet*. **69**, 1095-1112 (2001).

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

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

21. Klyosov, A.A. Basic rules of DNA genealogy (in Russian). *Proc. Russian Academy of DNA Genealogy* (ISSN 1942-7484), **1** (2), 252-348 (2008).

22. Klyosov, A.A. Calculations of time spans to common ancestors for haplotypes of Y chromosome (in Russian). *Proc. Russian Academy of DNA Genealogy* (ISSN 1942-7484), **1** (5), 812-835 (2008).

23. Adamov, D.S. and Klyosov, A.A. Theoretical and practical estimates of reverse mutations in haplotypes of Y chromosome (in Russian). *Proc. Russian Academy of DNA Genealogy* (ISSN 1942-7484), **1** (4), 631-645 (2008).

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

25. Basque DNA Project http://www.familytreedna.com/public/BasqueDNA

26. Adamov, D.S. and Klyosov, A.A. Evaluation of an "age" of populations from Y chromosome using methods of average squared differences (in Russian).

Proc. Russian Academy of DNA Genealogy (ISSN 1942-7484), **1** (5), 855-905 (2008).

27. Klyosov, A.A. The features of the "West European" R1b haplogroup (in Russian). *Proc. Russian Academy of DNA Genealogy* (ISSN 1942-7484), **1** (4), 568-629 (2008).

28. Haak, W., Brandt, G., de Jong, H.N., Meyer, C., Ganslmeier, R., et al. Ancient DNA, stroncium isotopes, and osteological analyses shed light on social and kinship organization of the later Stone Age. *Proc. Natl. Acad. Sci. US,* on-line before print, 10.1073/pnas.0807592105, November 17, 2008

29. Bouakaze, C., Keyser, C., Amory, S. and Crubezy, E. First successful assay of Y-SNP typing by SNaPshot minisequencing on ancient DNA. *Int. J. Legal Med*. **121**, 493-499 (2007)

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

31. Klyosov, A.A. Where is the origin of the Slavs and Indo-Europeans? DNA genealogy gives an answer (in Russian). *Proc. Russian Academy of DNA Genealogy* (ISSN 1942-7484), **1** (3), 400-477 (2008).

32. Barac, L., Pericic, M., Klaric, I.M., Janicijevic, B., Parik, J., et al. Y chromosome STRs in Croatians. *Forensic Sci. Internat.* **138**, 127-133 (2003)

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

34. Barac, L., Pericic, M., Klaric, I.M., Rootsi, S., Janicijevic, B., et al. Y chromosomal heritage of Croatian population and its island isolates. *Europ. J. Human Genetics* **11**, 535-542 (2003)

35. Kivisild, T., Rootsi, S., Metspalu, M., Mastana, S., Kaldma, K., et al. The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. *Am. J. Hum. Genet.* **72**, 313-332 (2003)

36. Sengupta, S., Zhivotovsky, L.A., King, R., Mehdi, S.Q., Edmonds, et al. Polarity and temporality of high-resolution Y-chromosome distributions in

India identify both indigenous and exogenous expansions and reveal minor genetic influence of Central Asian Pastoralis. *Amer. J. Human Genet*. **78**, 202-221 (2006)

37. Sahoo, S., Singh, A., Himabindu, G., Banerjee, J., Sitalaximi, T., et al. A prehistory of Indian Y chromosomes: evaluating demic diffusion scenarios. *Proc. Natl. Acad. Sci. US*, **103**, 843-848 (2006)

38. Bortolini, M.-C., Salzano, F.M., Thomas, M.G., Stuart, S., Nasanen, S.P.K., et al. Y-chromosome evidence for differing ancient demographic histories in the Americas. *Am. J. Hum. Genet*. **73**, 524-539 (2003)

39. Fagundes, N.J.R., Kanitz, R., Eckert, R., Valls, A.C.S., et al. Mitochondrial population genomics supports a single pre-clovis origin with a coastal route for the peopling of the Americas. *American Journal of Human Genetics* **82**, 583–592 (2008).

40. Haynes, G. The Early Settlement of North America: *The Clovis Era*. Cambridge University Press, 52 (2002)

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

42. Bradley, B., and Stanford, D. The North Atlantic ice-edge corridor: a possible Palaeolithic route to the New World. *World Archaeology* **36**, 459 – 478 (2004).

43. Seielstad, M., Yuldasheva, N., Singh, N., Underhill, P., Oefner, P., et al. A novel Y-chromosome variant puts an upper limit on the timing of first entry into the Americas. *Am. J. Hum. Genet*. **73**, 700-705 (2003)

44. Arabian Peninsula YDNA Project
http://www.familytreedna.com/public/ArabianPeninsulaDNAProject

Supplementary Information is linked to the online version of the paper at www.nature.com/nature

Figure Legends

Figure 1. The 25-marker haplotype tree for Basque R1b1 (mainly R1b1b2) haplotypes. The 17-haplotype tree was composed according to data of the Basque DNA Project[25].

Figure 2. The 25-marker haplotype tree for England, haplogroup R1a1. The 57-haplotype tree was composed from data of YSearch database. A seven-haplotype branch at the bottom (between 035 and 043) is a family of haplotypes with DYS388=10 (all other mostly have DYS388=12, in one case =14, haplotype 031).

Figure 3. The 25-marker haplotype tree for Russia and Ukraine, haplogroup R1a1. The 44-haplotype tree was composed from data of YSearch database and private sources.

Figure 4. The 9-marker haplotype tree for the Balkans, haplogroup R1a1.

The 67-haplotype tree was composed from data of Barac et al[32-34]

Figure 5. The 6-marker haplotype tree for the Native Americans, haplogroup Q-M3. The 117-haplotype tree was composed from data of Bortolini et al[38]

Figure 6. The 37-marker haplotype tree for the "Cohen Modal Haplotypes", haplogroup J1. The 85 haplotype tree was composed of haplotypes collected in YSearch database[20] and private "Cohen Haplotype" projects, and provided by Dr. Alberto Aburto.

Figure 7. The Lemba 6-marker haplotype tree, apparently of haplogroup J. The 57 haplotype tree was composed of data published in [14].