ISSN 0006-2979, Biochemistry (Moscow), 2011, Vol. 76, No. 5, pp. 517-533. © Pleiades Publishing, Ltd., 2011.
Published in Russian in Biokhimiya, 2011, Vol. 76, No. 5, pp. 636-653.

REVIEW

Dedicated to the 75th Anniversary of Biokhimiya/Biochemistry (Moscow)

Biological Chemistry as a Foundation of DNA Genealogy: the Emergence of "Molecular History"

A. A. Klyosov

MIR International, Inc., 36 Wells Avenue, Newton, Massachusetts 02459, USA; E-mail: aklyosov@comcast.net

Received November 17, 2010
Revision received December 17, 2010

Abstract—This paper presents the basis of DNA genealogy, a new field of science, which is currently emerging as an unusual blend of biochemistry, history, linguistics, and chemical kinetics. The methodology of the new approach is comprised of chemical (biological) kinetics applied to a pattern of mutations in non-recombinant fragments of DNA (Y chromosome and mtDNA, the latter not being considered in this overview). The goal of the analysis is to translate DNA mutation patterns into time spans to the most recent common ancestors of a given population or tribe and to the dating of ancient migration routes. To illustrate this approach, time spans to the common ancestors are calculated for ethnic Russians, that is Eastern Slavs (R1a1 tribe), Western Slavs (I1 and I2 tribes), and Northern (or Uralic) Slavs (N1c tribe), which were found to live around 4600 years before present (R1a1), 3650 ybp (I1), 3000 and 10,500 ybp (I2, two principal DNA lineages), and 3525 ybp (N1c) (confidence intervals are given in the main text). The data were compared with the respective dates for the nearest common ancestor of the R1a1 "Indo-European" population in India, who lived 4050 years before present, whose descendants represent the majority of the upper castes in India today (up to 72%). Furthermore, it was found that the haplotypes of ethnic Russians of the R1a1 haplogroup (up to 62% of the population in the Russian Federation) and those of the R1a1 Indians (more than 100 million today) are practically identical to each other, up to 67-marker haplotypes. This essentially solves a 200-year-old mystery of who were the Aryans who arrived in India around 3500 years before the present. Haplotypes and time spans to the ancient common ancestors were also compared for the ethnic Russians of haplogroups I1 and I2, on one hand, and the respective I1 and I2 populations in Eastern and Western Europe and Scandinavia, on the other. It is suggested that the approach described in this overview lays the foundation for "molecular history", in which the principal tool is high-technology analysis of DNA molecules of both our contemporaries and excavated ancient DNA samples, along with their biological kinetics.

DOI: 10.1134/S0006297911050026

Key words: DNA, DNA genealogy, haplogroups, haplotypes, mutations, Y chromosome

This paper describes a new scientific discipline that is progressively accumulating new experimental material. Dozens or even hundreds of new "experimental points" are flowing daily into databases, and those data points provide feedback for correcting the methodology of the new science, which in turn brings about a refining of the methods of calculation using the data.

The name of this science is "DNA genealogy". Its experimental data are essentially patterns of mutations in the non-recombinant part of the male Y chromosome and female mitochondrial DNA. This paper includes only Y chromosomal data due to limits of space. This Y chromosomal data will be related to both individuals and their groups, i.e. populations. The methodology of the new science aims to translate a dynamic pattern of mutations into chronological data, that is time spans to common ancestors of (ancient) populations, tribes, clans, and sometimes whole peoples. In fact, the calculations provide us with times when those tribes and clans lived. Careful consideration of these calculated times along with geographical location of populations under consideration, identified from either historical sources or excavated data, or both, describes ancient migrations, ancient genealogies of certain DNA lineages, routes of the spread of language, and traces of lost connections between certain individuals and populations.

Mutations considered in DNA genealogy are either single (or involving just a few nucleotides) replacements

of nucleotides in DNA, such as adenine by cytosine, or cytosine by thymine, or insertions or deletions of nucleotides, collectively called SNPs (single nucleotide polymorphisms), or some more complex "tandem" mutations. The latter, called STRs (short tandem repeats), occur when a mistake of the copying enzyme (or rather molecular copying machinery in the cell nucleus) brings about an extra copying of an extended block of nucleotides being inserted into the DNA, or, on the contrary, an elimination of an extended block of nucleotides. SNP mutations selected for the goals of DNA genealogy are very stable and occur only once (rarely twice) during the history of mankind. Therefore, they actually serve as markers of human tribes, which are the principal DNA genealogical lineages of the highest hierarchical level and the following downstream levels, the so-called subclades. STR mutations are more frequent and occur in certain and well identified loci in the Y chromosome once every few hundred or few thousand generations. A set of those loci is called a haplotype.

Haplotypes for DNA genealogy are selected in such a way so that they contain both slow and fast-mutating loci, or STR markers, and contain as many of them as possible, albeit practical for routine experimental work (Y chromosome testing). In earlier studies, that is in the end of the 1990s and beginning of the 2000s, 6-marker haplotypes were commonly used and the resulting data published in academic papers. Since then academic studies have shifted routinely to 7, 8, 9, 10, and lately 17 and 19-marker haplotypes, but occasionally 21 and even 39-marker haplotypes. Commercial Y chromosome testing, which provides a great resource for DNA genealogy, has progressed from the initial 12 and then 25-marker haplotypes to 37 and 67-marker ones. In 67-marker haplotypes, for example, one mutation occurs on average each eight generations, which allows for a rather detailed description of family DNA genealogies for the last several centuries.

Since the mutations in Y chromosome loci picked for DNA genealogy occur pretty much randomly, as it has been shown by various methods, their time-wise behavior can be described in terms of chemical (or biological) kinetics. In this context chemical and biological kinetics mean the same thing, namely, a description of time course of accumulation of mutations in haplotypes or disappearance of initial (non-mutated, or ancestral, or base) haplotypes, employing certain (and experimentally determined) mutation rate constants. The mutation of each marker (a certain Y chromosome locus, identified and numbered) is described with its specific mutation rate constant, as well as mutation of each haplotype being a set of markers is described with its average mutation rate constant, which is a sum of mutation rate constants of its constituent markers [1, 2]. These mutation rate constants can be determined from either massive father–son pair studies, or consideration of numerous haplotype datasets, or from known and extended family genealogies, by comparing a pattern of mutations in family members and their position in a genealogical tree. Obviously, the more ancient a common ancestor of a population, the more mutations have accumulated in haplotypes of his descendants, the more generations (and years) separate him from his current descendants. That number of years can be calculated by the methods of DNA genealogy.

Since mutations happen "between" generations, on transitions from father to son, the calculations are technically done by generations as principal units, and then by translation of a number of generations into a number of years separating a common ancestor from his present-day descendants (a time span to a (most recent) common ancestor – TCA). Obviously, the length of a generation is a "floating" figure and depends on many factors including the culture of a society, historical epoch, and such circumstances as wars, famine, epidemics, natural cataclysms, etc. It would be impossible to clearly define a length of a generation in years for all times and all peoples. Therefore, DNA genealogy employs an apparent generation, equal to 25 years, which essentially is a mathematical figure, not a "real" generation length. In other words, if a common ancestor of a given genealogical lineage lived, say, 650 years ago, he lived, by definition, 26 generations ago.

Let us give a few specific examples to illustrate which kind of data are considered in DNA genealogy, the essence of calculations, and what kind of data (results) are typically obtained from those considerations.

It was experimentally established that ethnic Russians include three major tribes (in terms of DNA genealogy) and a number of minor tribes. For the purpose of this kind of work, ethnic Russians are typically defined as those who consider Russian as their native tongue within at least three generations and whose ancestors lived for at least three generations in the central part of Russia (which is designated as the 12 official regions of Russia, currently the Russian Federation, for centuries). In the study [3] 545 ethnic Russians, unrelated in at least three generations, were tested for their SNP (haplogroups) and STR (haplotypes). The three major tribes include 84% of the population of ethnic Russians, and, if tentatively to use a linguistic classification, they are Eastern Slav, Western Slav, and Finno-Ugric populations. The latter can be appropriately called Northern Slavs. In terms of DNA genealogy those tribes differ by SNP mutations as follows: all "Eastern Slavs" have SNP M198 (rs2020857), which is the mutation of cytosine to thymine in a certain position in the Y chromosome. All "Western Slavs" have SNP M170 (rs2032597), which is the mutation of adenine to cytosine in another position. All "Northern Slavs" have SNP M46 (Tat, rs34442126), which is the mutation of cytosine to thymine in yet another position of the Y chromosome. The haplogroup to which all the "Eastern Slavs" belong (but which is not

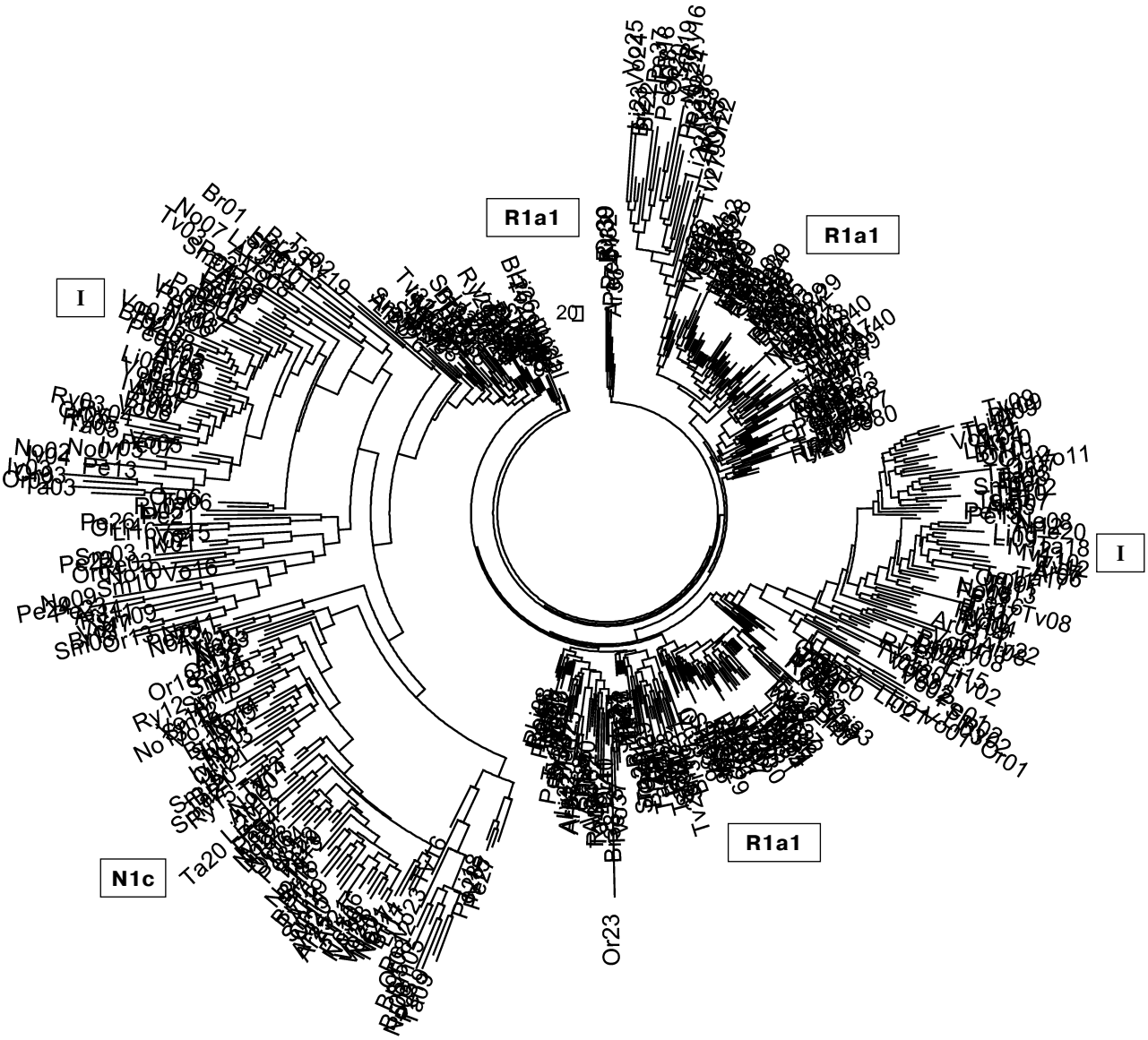

Fig. 1. Tree of 545 ethnic Russian 17-marker haplotypes collected in 12 administrative regions of the Russian Federation (Arkhangelsk, Bryansk, Vologda, Ivanovo, Lipetsk, Penza, Novgorod, Orel, Smolensk, Ryazan, Tambov, Tver) composed from haplotypes listed in [3]. The most numerous branches are marked as haplogroups to which the branches belong – R1a1 (47% on average, up to 62% in the South), N1c (14% on average, more numerous in the North from the Arkhangelsk through Pskov regions), I (22% on average, more numerous near the Baltic Sea and in the West of Russia).

restricted to them) is defined as R1a1, that to which the "Western Slavs" belong to is called I, and that which includes the "Northern Slavs" is designated as N1c. Those tribes split from their common ancestors tens of thousands of years ago, and for millions of individuals they came together in the Russian Federation.

When did these tribes have a common ancestor? Where did these common ancestors live? When did they come to territories that eventually became Russia? Those are the questions of DNA genealogy.

Paper [3] lists 17-marker haplotypes of 545 ethnic Russians selected for DNA testing according to the criteria described above. We used these data to compose the haplotype tree shown in Fig. 1. The computer program use to create the tree (see [1, 2] and references therein) sorts the haplotypes according to their mutations and arranges them into branches in accord with the most likely order of the mutations. The most mutated haplotypes are placed by the program at the most distance from the "trunk" of the tree; the least mutated are positioned close to the "trunk"; the original, ancestral, base haplotypes, identified by the program, are placed directly in the "trunk".

The tree in Fig. 1 is composed of 17-marker haplotypes, which generally possess little high resolution

capacity compared to that for 25, 37, or 67-marker haplotypes. It can be seen, nevertheless, that the main haplogroups are more or less resolved, and the other, minor haplogroups are combined into smaller branches of haplogroups R1b (5%), J2 (3%), E (3%), G (2%), K (2%), F (1%), and C (0.4%). Percent values here are related to fractions of the mentioned haplogroups out of the total 545 ethnic Russians in the Russian Federation. These fractions are generally reproducible from study to study.

Such complex multi-haplogroup haplotype trees typically serve for illustration purposes only. For their quantitative analysis separate and distinct branches are considered, particularly when they are confirmed by their SNP (haplogroup) data. Figure 2 shows only haplotypes that belong to haplogroup R1a1, the major one among ethnic Russians. It can be seen that the tree in Fig. 2 is a relatively symmetrical one compared to a quite heterogeneous multi-haplogroup tree in Fig. 1. A detailed consideration of the tree shown in Fig. 2 has been done in [4], which showed that it contains nine separate branches. They could have corresponded to nine ancient Slavic tribes, each one of them represented by an extended family. This would have indeed created separate branches of the haplotype tree. In the same work [4] time spans to common ancestors of all the nine branches were calculated. However, since this is only an illustration here, we will

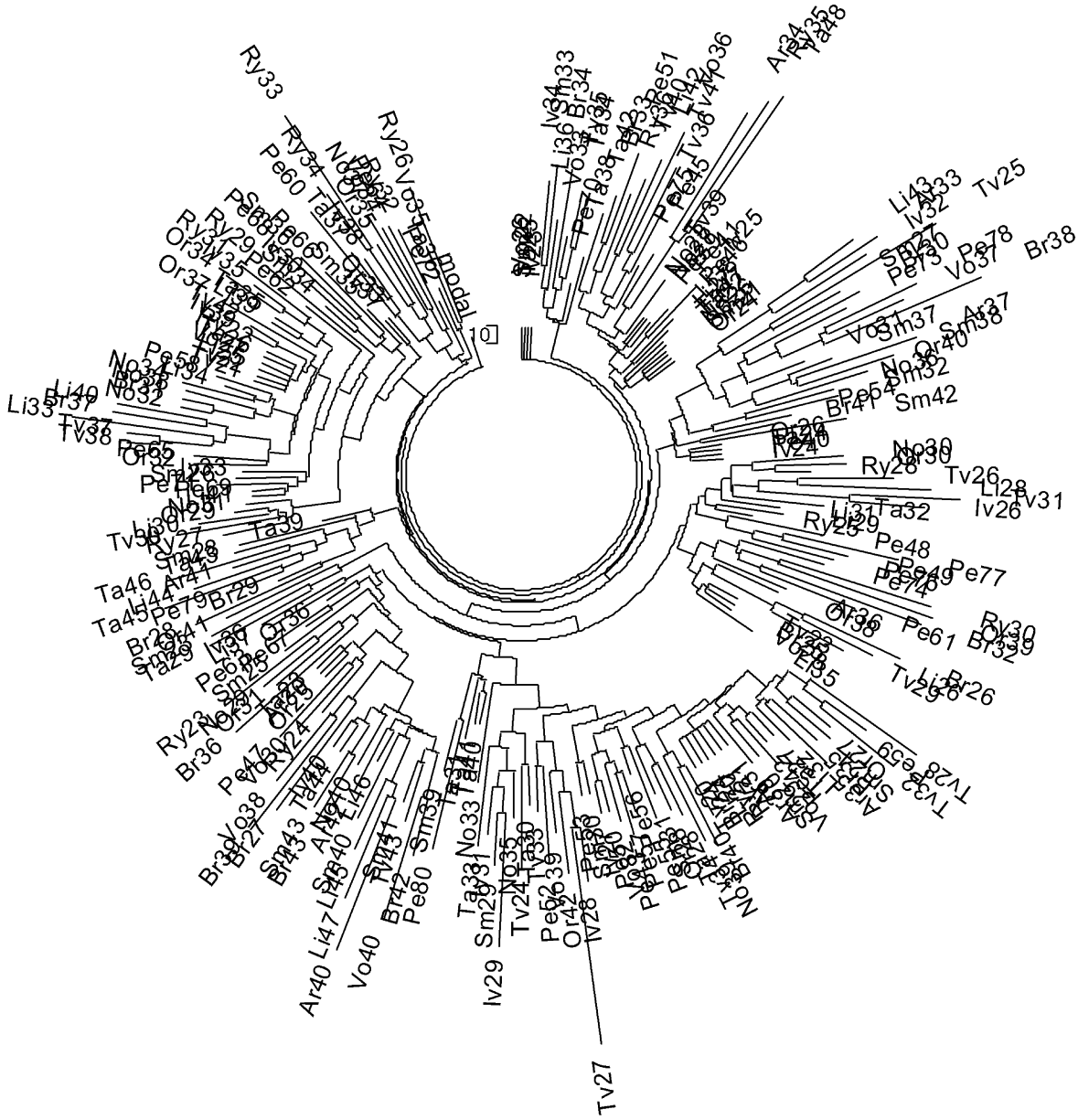

Fig. 2. The 17-marker haplotype tree for 255 haplotypes of haplogroup R1a1 in the 12 regions of the Russian Federation as listed in Fig. 1. The haplotypes were taken from [3].

not go into detail but just give the final calculations. All 255 haplotypes coalesce in terms of their mutations to one single haplotype, from which they spread as circles on the surface of water. In other words, there is one "central" haplotype, minimized on mutations, which was most likely the ancestral one. Figuratively speaking it will be "minimized on mutations", that is on chemical transformations. The minimization of all 255 haplotypes in this dataset in their mutations leads to an ancestral, or base haplotype, which in the given 17-marker format can be presented as follows: 16 11 14 13 30 25 11 11 13 14 11 10 20 16 15 23 11.

"Base" haplotype in this particular case is a synonym for "ancestral" haplotype; however, it is not an actual ancestral haplotype but a deduced one. It can be equal to the ancestral haplotype; however, it can also be its approximation. The above series of numbers describes the tandem repeats and their mutations. In the first marker, or the locus DYS19, where DYS is the DNA Y segment and 19 is the number of the locus according to the classification, a certain sequence of nucleotides repeats 16 times. In the second marker another sequence of nucleotides repeats 11 times, in the third marker yet another sequence repeats 14 times, and so on.

It turned out that all 255 haplotypes, that is all 4335 markers (255 × 17) contain 1320 mutations with respect to the base haplotype of haplogroup R1a1 of the ethnic Russians. This is equal on average to 0.304 ± 0.017 mutations per marker with 95% confidence.

Since an average mutation rate in the 17-marker haplotypes is equal to 0.002 mutation per marker per conditional generation of 25 years [1], as described above (if needed, the above mutation rate constant can be re-calibrated per any other conditional generation, such as 0.0024 mutation per marker for 30 years per generation, or 0.0028 mutation per marker for 35 years per generation, etc.), then a common ancestor of the 255 ethnic Russians of haplogroup R1a1 ("Eastern Slavs") lived 152 ± 16 generations ago (without a correction for back mutations) or 179 ± 19 generations (with the correction) [1], that is 4475 ± 460 years before present (the middle of the 3rd millennium BC).

The protocol for calculations, mutation rate constants for various haplotype formats, calibration procedures, tables of corrections for back mutations, calculation of confidence intervals and margins of errors, etc. are described in [1]. Now, let us verify the value of 4475 ± 460 years to a common ancestor obtained with the 17-marker haplotype format by considering 67-marker haplotypes of 148 individuals from the "post-Soviet" region, all of them having the R1a1 haplogroup. These were all the R1a1 haplotypes we could find in a commercial database for the "post-Soviet" region, published up to the middle of November, 2010. Haplotypes of such length could not be found in "academic" papers, except those by the author of this article. There are two main reasons why 67-marker haplotypes are not described in "academic" papers: they are too expensive to obtain, and it is not known how to analyze them quantitatively. Those 148 haplotypes were determined on requests of their bearers, who personally paid for the testing, and placed the haplotypes into the public database YSearch. The respective haplotype tree is shown in Fig. 3.

Some of these haplotypes belong to bearers of nationalities other than Russian. However, despite nationalities of their bearers all the haplotypes belong to the R1a1 haplogroup, which, as it was calculated above, arose some 4500 years ago among those who are now ethnic Russians. Obviously, in the course of those millennia the R1a1 tribe has expanded and included people of many different present-day nationalities. Indeed, the haplotypes of people of those nationalities listed in the legend to Fig. 3 are rather evenly mixed in the tree and do not form distinct branches. Let us see how the "age" of a common ancestor of the tree in Fig. 3 would correspond to the "age" of a common ancestor of the 255 ethnic Russians, that is "Eastern Slavs" (Fig. 2).

The base 67-marker haplotype of the series of 148 haplotypes can be written as follows: 13 25 16 11 11 14 12 12 10 13 11 30 15 9 10 11 11 24 14 20 32 12 15 15 16 11 11 19 23 16 16 18 19 34 39 13 11 11 8 17 17 8 12 10 8 11 10 12 22 22 15 10 12 12 13 8 14 23 21 12 12 11 13 11 11 12 13.

This is exactly the same base haplotypes for the 16 markers (shown below in bold) from the 17 determined above (the 17th marker is absent in the 67-marker format): **13 25 16 11 11 14** 12 12 **10 13 11 30 15** 9 10 11 11 24 **14 20** 32 12 15 15 16 11 **11** 19 23 **16** 16 18 19 34 39 13 **11** 11 8 17 17 8 12 10 8 11 10 12 22 22 15 10 12 12 13 8 14 23 21 12 12 11 13 11 11 12 13.

Let us take a look at the first 25 markers, the mutation rate constant for which is equal to 0.00183 mutations per marker per conditional generation of 25 years [1]. One can see that it is "slower" compared to that for the 17-marker format (0.002 mutation per marker per "generation"). There are 1037 mutations in the first 25 markers from the above base haplotype, which gives an average 0.280 ± 0.017 mutation per marker and 153 ± 16 generations without a correction for back mutations, or 180 ± 19 generations with the correction, which is 4500 ± 470 years to a common ancestor. For the 17-marker haplotypes it was 4475 ± 460 years. Such a practically absolute fit is a coincidence, of course; however, those coincidences within 95% confidence interval are rather the rule than an exception.

If we go beyond the 25-marker haplotypes and calculate mutations in all of the 148 haplotypes in the first 37 markers, we find 2023 mutations. The mutation rate constant for the 37-marker haplotypes equals 0.00243 per marker per "generation" [1]. This gives us 0.369 ± 0.016 mutation per marker on average, or 152 ± 16 "generations" without the correction for back mutations, or 179 ± 18 "generations" with the correction, which is

4475 ± 460 years to a common ancestor. This is again almost an absolute fit with the above values for 17 and 25-marker haplotypes. Finally, in all the 67 markers the 148 haplotypes contain 2748 mutations, which gives 2748/148/0.12 = 155 ± 16 "generations" without the correction, or 183 ± 19 "generations" with the correction, that is 4575 ± 470 years to a common ancestor (the mutation rate constant for the 67-marker haplotype equals 0.12 mutation per haplotype, or 0.00179 mutation per marker per "generation"). Again we have an almost absolute fit. An average time span to the common ancestor for the 148 haplotypes (rather, for 148 individuals) is 4500 ± 60 years. This is, of course, a formal calculation of the margin of error, however, the accuracy of the calculations is notable in terms of cross-fit between different series of haplotypes and different haplotype formats.

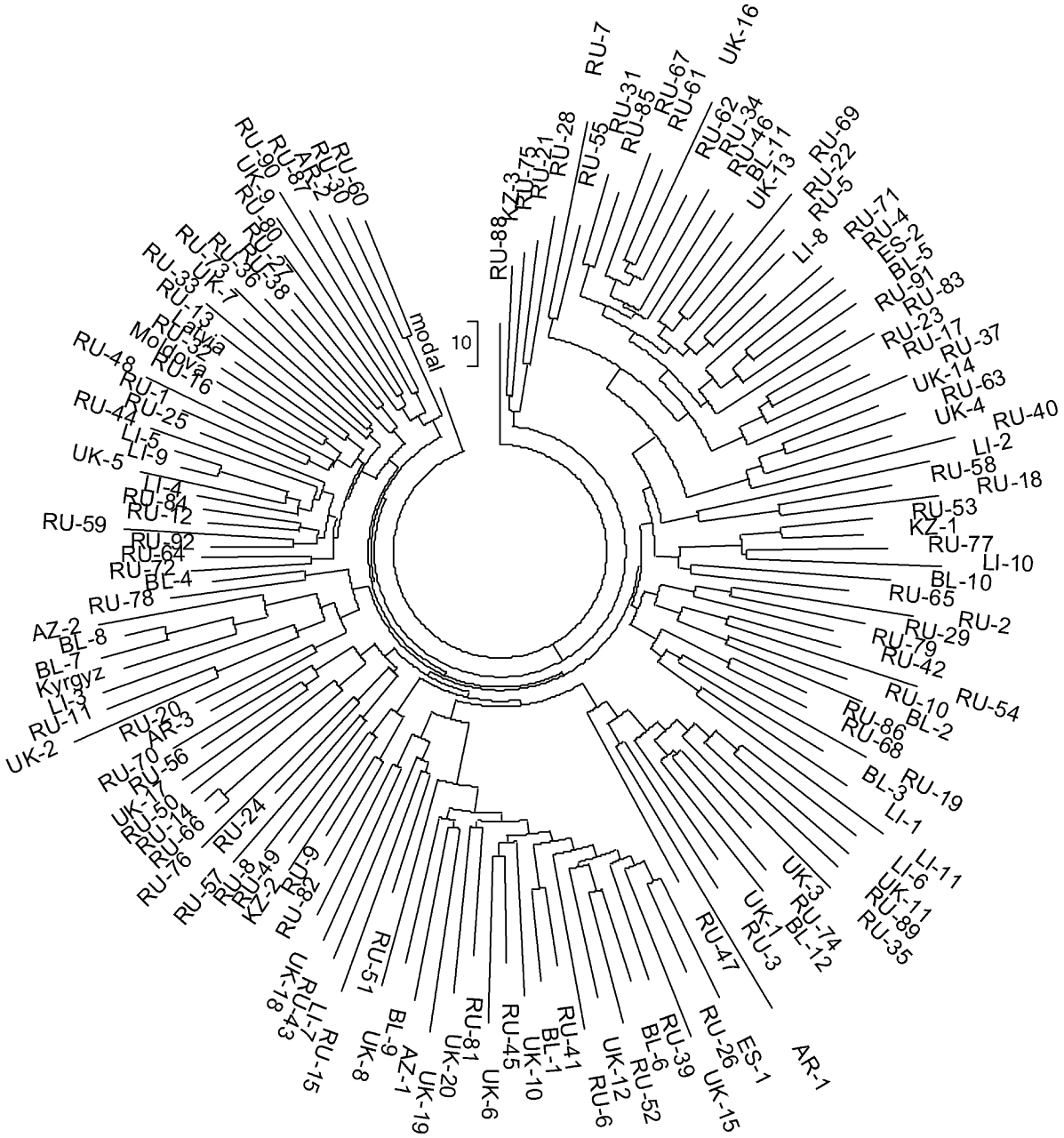

Fig. 3. Tree of 148 haplotypes of haplogroup R1a1 collected in the "post-Soviet" territory. The 67-marker haplotypes were taken from YSearch database and local Russian sources. Indexes RU, UK, BL, LI, KZ, AR, ES, and AZ correspond to Russian (92 haplotypes), Ukrainian (20), Belorussian (12), Lithuanian (11), Kazakh (3), Armenian (3), Estonian (2), Azerbaijan (2), and Kyrgyz, Latvian, and Moldovan (one haplotype in each region). A number of haplotypes on the tree reflects two factors: (1) relative population size of the ethnic group, and (2) a number of people from the ethnic group who placed a commercial order for DNA testing.

History of determinations of a time span to a common ancestor of haplotypes of haplogroup R1a1 on the "post-Soviet" territory from 2008 through 2010 and using different series of haplotypes

Date	Number of haplotypes in the dataset	Total number of mutations	Time span to common ancestor, years	Source
June 2008	26	178	4400 ± 550	[5]
November 2008	44	326	4825 ± 550	[6]
January 2009	58	423	4725 ± 520	[1]
February 2009	255	1320	4475 ± 460	[4]
March 2009	98	711	4700 ± 500	[1]
June 2009	110	804	4750 ± 500	[1]
November 2010	148	1037	4500 ± 470	this study
–"–	148	2023	4475 ± 460	–"–
–"–	148	2748	4575 ± 470	–"–

Overall, a series of haplotypes, numbered in the hundreds, forms a robust system that is quite resistant to statistical variations (table). The table shows that an average time span to a common ancestor of the R1a1 haplogroup from the "post-Soviet" territory, with ethnic Russians as the most numerous population, equals 4600 ± 150 years. This margin of error is noticeably lower compared with that which was estimated from a number of mutations in haplotype series and the assumed margins of error in mutation rate constants.

It was discovered in the course of the studies that haplotypes of the majority of Indians listed in databases as belonging to R1a1 haplogroup, which in turn can be projected to include more than 100 million Indians, are practically identical to the haplotypes of the ethnic Russians of the same haplogroup R1a1 [5, 7]. In other words, half of the Russians and between one-quarter and one-third of the Indians are descendants of the same common ancestor. For the R1a1 "Indo-European" Indians he lived around 4050 years before present, which is the same time as when a common ancestor of the Iranian R1a1 tribe lived [8]. The base (ancestral) haplotype shown above is identical for the Russians, Indians, and Iranians, but a common ancestor for the Russians on the Russian Plain is "older" by about 500 years [7]. This is directly related to a solution of the most interesting historical puzzle that was addressed about 200 years ago. We will leave it for the reader to answer, but we will hint that the puzzle was related to the Aryans, as the people called themselves in the Indian Vedas. Until lately science did not know how to describe the Aryans in "scientific terms". What objectively measurable parameter could have been applied to them? There were some "Aryans", who allegedly came to India from the North, so they knew of snow, cold weather, they knew of the birch tree, and ash tree and beech tree and they knew of wolves, bears, and horses. Now we know that the Aryans were people of the R1a1 haplogroup, whose descendants live nowadays across the world, among them being Eastern Slavs – Russians, Ukrainians, and Belarus. Furthermore, among the Eastern Slavs the share of the R1a1 haplogroup is the highest one, and the time when their common ancestors lived on the Russian Plain preceded the time of the passage of the Aryans to India and the Iranian Plateau, which took place around 3500 years before present, in the middle of the 2nd millennium BC.

A striking illustration for these studies of a joint history of common ancestors of the Russians and the Indians is as follows. Let us take a look at the 67-marker haplotype of the author of this paper, a Slav of the R1a1 haplogroup:

13 24 16 11 11 15 12 12 10 13 11 30 16 9 10 11 11 24 14 20 34 15 15 16 16 11 11 19 23 15 16 17 21 36 41 12 11 11 9 17 17 8 11 10 8 10 10 12 22 22 15 10 12 12 13 8 15 23 21 12 13 11 13 11 11 12 13,

and three quite typical 67-marker haplotypes of Indians randomly taken from the Indian database FTDNA. Mutational differences between them are shown in bold:

13 24 **17 10** 11 14 12 12 10 13 11 **32** 16 9 10 11 11 24 14 20 **31** 12 15 15 16 11 **10** 19 23 **16** 16 **17 20 33 34 13** 11 11 8 17 17 8 **11** 10 8 11 10 12 22 22 15 10 12 12 13 8 14 23 21 13 13 11 13 11 11 12 **13**,

13 24 16 11 11 14 12 12 10 13 11 **31** 16 9 10 11 11 24 14 20 **33** 12 15 15 16 **10** 12 19 23 15 17 18 **18** 35 **41 15** 11 11 8

17 17 8 12 10 8 11 10 12 22 22 15 10 12 12 13 8 **13** 23 21 **12 12** 11 13 **10** 11 12 12,

13 **23** 16 11 **12 15** 12 12 10 13 11 **30** 16 9 10 11 11 24 14 20 **30** 12 **16 16** 16 11 12 19 23 15 16 18 **21** 35 **39 12** 11 11 8 17 17 8 12 10 8 11 10 12 22 22 **16** 10 12 12 13 8 14 **24 22** 13 13 11 13 11 11 12 12.

The four haplotypes are obviously similar to each other. The number of mutations between the Indian haplotypes (pair-wise) equals 27-30, and that between the Slavic haplotype of the author and each of the Indian haplotypes equals 25-30. In other words, the Slavic haplotype is closer to the Indian haplotypes than the Indian haplotypes between each other. In fact, those differences are within the margins of error, and all four haplotypes are equally similar to each other.

This can be compared to a typical Western European base haplotype of haplogroup R1b, which (and its variations) include around 60% of Western and Central Europeans, and up to 90% (and higher) population in the British Isles [9]:

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

The number of mutations between the European R1b base haplotype and the Indian (and Russian) haplotypes shown above is around 50. This is of no surprise, since their common ancestors are separated by at least 30 thousand years. There are almost no haplotypes of the R1b haplogroup in India and Iran. It looks like there were no bearers of the R1b haplogroup, that is, ancestors of the majority of present day Western Europeans, among the Aryans 3500 years ago. It should also be noted that the upper castes of India in the present time consist of up to 72% of bearers of haplogroup R1a1, particularly among Brahmins [10]. At the same time not a single Brahmin among 367 tested belonged to haplogroup R1b [10].

Similar approaches to the analysis of haplotypes of ethnic Russians of haplogroup I ("Western Slavs") and N1c ("Finno-Ugric", or "Northern Slavs") resulted in the haplotype trees shown below [11, 12]. In paper [3], which listed those haplotypes, their analysis was not conducted as well as in all other papers typically aimed at population genetics, in which quite different aspects of the populations are addressed and a quite different methodology is employed. Population genetics commonly undertakes a comparative analysis of populations via studying frequencies of alleles and their changes under effects of evolution-driven processes, aiming at adaptation and specialization in populations and eventually identifying regulatory processes and transitions from genotypes to phenotypes in the populations. On the contrary, DNA genealogy is essentially a historical science, which is developed on the basis of chemical and biological kinetics and DNA sequencing. It is focused on a chronological component. Briefly, DNA genealogy is a merge of DNA sequencing and the methods of chemical kinetics, aiming at the analysis of a dynamics of changing of those sequences in populations. There is no genetics in DNA genealogy; it considers only non-gene regions in the DNA.

Kinetics of accumulation of mutations in haplotypes is similar in kind with kinetics of parallel and consecutive chemical or biological reactions, in fact, reversible reactions, since mutations from the ancestral haplotype in the DNA can occur in the both directions, that is, by increasing or decreasing of alleles (number of tandem repeats). This looks like a rather complicated system, particularly for time spans in thousands and tens of thousands of years. Analysis of such complicated systems requires approaches of chemical kinetics, which are not in use in population genetics. Those approaches include logarithmic regularities of reducing the numbers of the initial, ancestral, base haplotypes with time in each branch of a haplotype tree, of accumulation of mutations in haplotypes, consideration of reverse mutations, which progressively accumulate with time, and of the symmetry of "forward" and "backward" mutations, etc. [1, 13-15].

It should be noted that the overall haplogroup I in the cited article [3] included 117 haplotypes; however, the haplotype tree (Fig. 4) clearly shows the presence of quite distinct branches, each one having its own history and common ancestor who lived in quite different historical times. Granted, those branches descended from one common ancestor, presumably of haplogroup I, who lived at least 15,000 years before present; however, they formed downstream haplogroups, I1 and I2, which in turn produced their own DNA-genealogical branches. The branch on the right belongs to haplogroup I2 with a common ancestor who lived 3000 $\pm$ 380 years ago, who had the following haplotype: 16 14 15 13 31 24 11 11 13 15 10 13 20 15 17 23 10, in the same format which was employed above in this paper for a description of R1a1 haplotypes, the so-called Y-filer format. This base haplotype of the ethnic Russians of haplogroup I2 differs from the R1a1 base haplotype of ethnic Russians (see above) by 14 mutations, which corresponds to the "lateral" time difference between the respective common ancestors in the range of tens of thousands of years.

The ancient branch of haplogroup I2 of ethnic Russians in Fig. 4, which is the most "fluffy" and remote from the "trunk" of the tree, has a common ancestor who lived 10,500 $\pm$ 1100 years before present. This branch contains 203 mutations in the 20 haplotypes, which is on average 0.597 $\pm$ 0.084 mutation per marker (cf. 0.304 $\pm$ 0.017 mutation per marker for haplogroup R1a1 above). Clearly, this branch is much more ancient and the calculations above show how much more ancient.

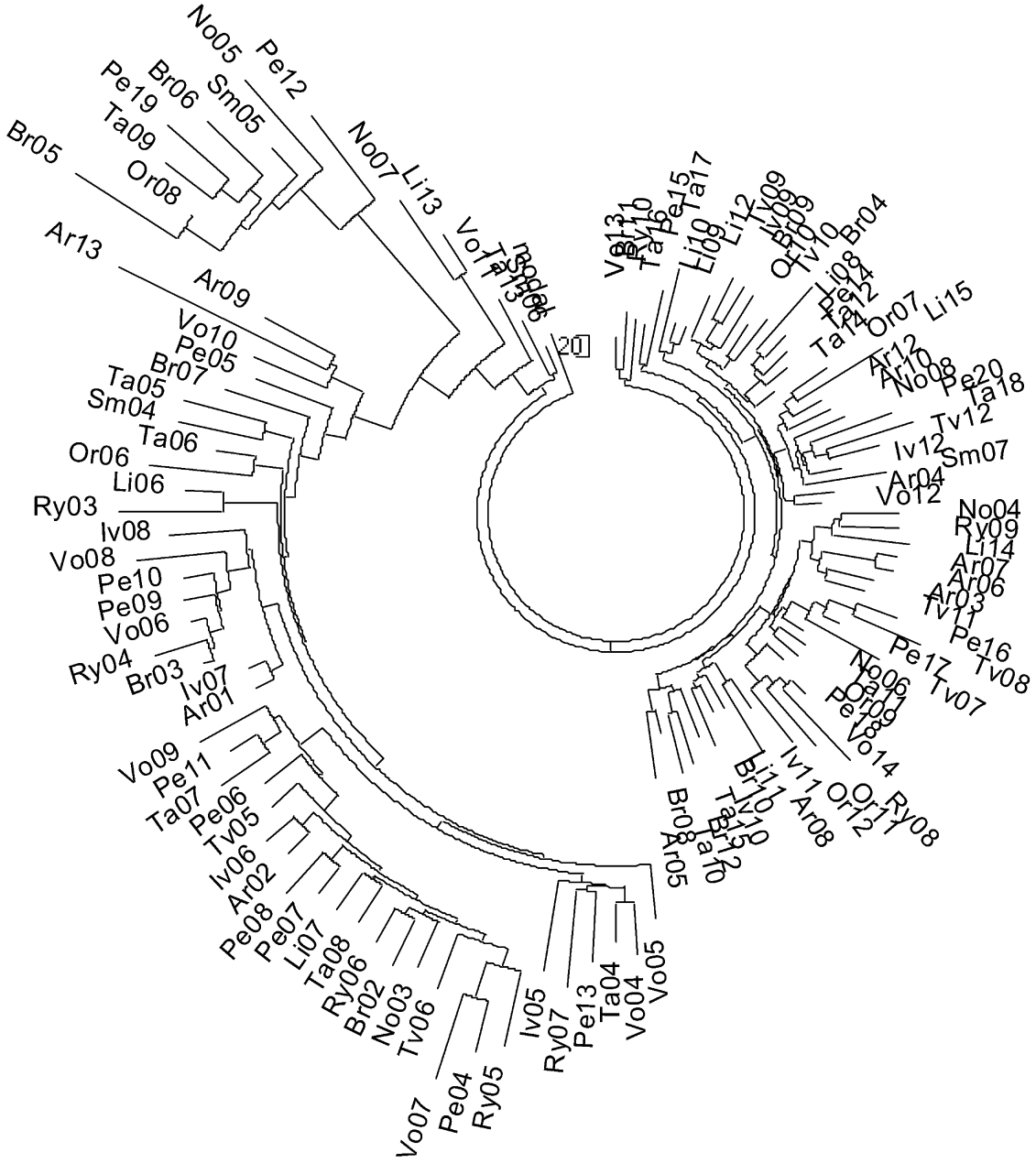

Fig. 4. Tree of 117 ethnic Russian 17-marker haplotypes of haplogroup I collected in the 12 regions of the Russian Federation listed in Fig. 1. The tree was composed using haplotypes listed in [3]. A "young" branch of haplogroup I2 is shown on the right-hand side (3000 ± 380 years to a common ancestor), a branch of haplogroup I1 (3650 ± 800 years to a common ancestor), in the middle part of the left-hand side branch, and the ancient branch of haplogroup I2 (10,500 ± 1100 years to a common ancestor) are shown on the left.

Let us compare data obtained with 17-marker haplotypes again with those for more extended haplotypes of haplogroup I2. Figure 5 shows a haplotype tree for haplotypes I2 of Eastern Europe. It has the same configuration as that for ethnic Russians (Fig. 4), albeit turned in a different direction, which does not matter for data analysis. The "age" of the "young" branch equals 2650 ± 320 years, which is somewhat less than that for ethnic Russians (3000 ± 380 years), but within the margin of error of the calculations. The "age" of the ancient branch is 10,800 ± 1200 years, which is practically equal to that for ethnic Russian of haplogroup I2 (10,500 ± 1100 years). Thus, the I2 tribe in Eastern Europe is the same as that for ethnic Russians, with the same common ancestors, irrespective of the contemporary state borders. The base (ancestral) haplotype for them is the same, and in the 25-marker haplotype it looks as follows: 13 24 16 11 14 15 11 13 13 13 11 31 17 8 10 11 11 25 15 20 32 12 14 15

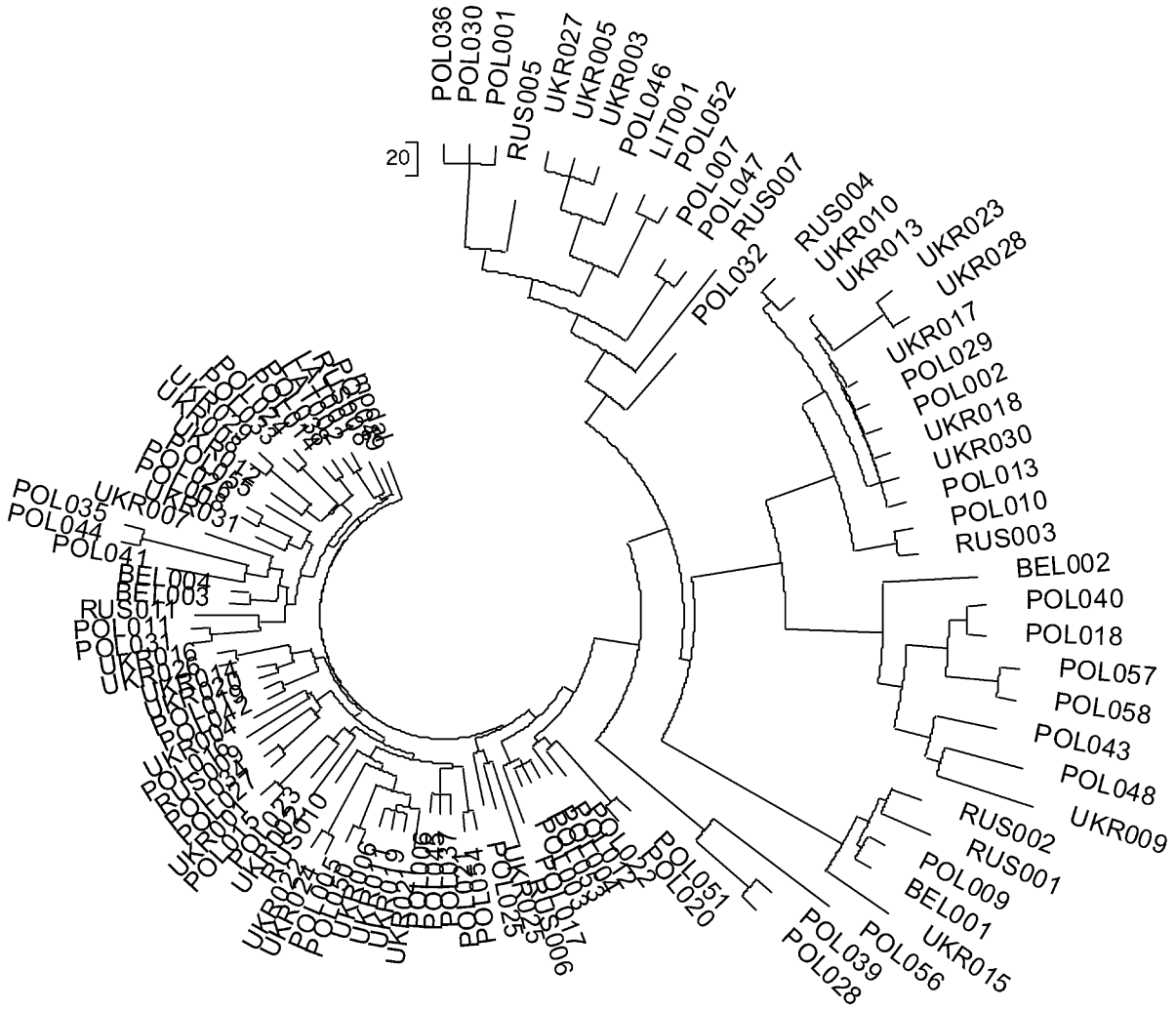

Fig. 5. The 25-marker haplotype tree of haplogroup I2 for Eastern Europe. The tree consists of 96 haplotypes taken from the YSearch database. Indexes correspond to Russia, Ukraine, Belarus, Poland, Lithuania, and Latvia (the latter is represented only in the "young" branch).

15. It is exactly the same as that for the 17-marker haplotypes, with the respective makers in the latter shown in bold: **13 24 16 11 14 15** 11 13 **13 13 11 31 – 17** 8 10 11 11 25 **15 20** 32 12 14 15 15.

Common ancestors of haplogroup I2 in Scandinavia are quite different from those in Eastern Europe, but their "age" is the same. The haplotype tree of I2 in Scandinavia is shown in Fig. 6. The branch on the right-hand side has a common ancestor who lived 3025 ± 470 years before present, which is within the margin of error with that for a common ancestor for the "young" branch of ethnic Russians (3000 ± 380 years before present), and the I2 Eastern Europeans in general (2650 ± 320 years before present). However, the base haplotype of this Scandinavian branch is quite different: 13 22 14 10 13 14 11 14 11 12 11 28 – 15 8 9 8 11 23 16 20 28 12 14 15 15. It differs from the Eastern European one (including the ethnic Russians) by 26 mutations in 25 markers (!), which sets their common ancestors apart by as many as 29,000 years. This means that the mutational pattern, which we see in haplotypes of the "young" branches in both cases, is just a "tip of the iceberg", and shows what biologists call "passing a bottleneck of the population". We see only two lines of rather recent descendants of an ancient common ancestor. Those descendants (miraculously) passed the bottleneck and started their lines again some three thousand years ago both in Scandinavia and in Eastern Europe, being already quite apart from each other in terms of their base haplotypes. It is the mutational difference between the two lines that makes it possible to calculate that their common ancestor lived almost thirty thousand years ago, apparently in Europe. Those are the times of Neanderthals.

Another fragment of the same ancient lineage is the I2 branch with an "age" of around 10,000 years (10,500 ± 1100 years for the ethnic Russians and 10,800 ± 1200 in

Eastern Europe). This ancient branch in Scandinavia (the branch on the left-hand side in Fig. 6) is also quite different compared to that in Eastern European I2. Their base haplotypes are as follows: 13 23 14 10 14 15 11 14 11 12 11 28 – 16 8 9 8 11 24 16 20 28 12 14 14 15 and 13 23 15 10 15 15 11 13 11 13 12 29 – 16 8 9 11 11 24 14 20 27 12 14 15 16.

The first, Scandinavian branch has on average 0.509 ± 0.065 mutation per marker, which translates to 9575 ± 1140 years to a common ancestor. The second, Eastern European branch, has 0.556 ± 0.045 mutation per marker, that is 10,800 ± 1200 years to a common ancestor. There are 13 mutations between their base haplotypes, which places their common ancestor at about 15,000 years ago. Those are also fragments of an ancient lineage of I2, and before it there was the I lineage, which was the most ancient haplogroup in Europe.

It should be noted that the apparently "young" branch of haplogroup I1 in Fig. 4 is indeed the youngest one in the overall haplogroup I. Its base haplotype in the Y-filer format is 14 14 14 12 29 22 10 11 13 16 10 11 20 14 15 22 10, and its common ancestor lived 3650 ± 800 years before present. Practically the same base haplotype in the 25-marker format (the same alleles are shown in bold,

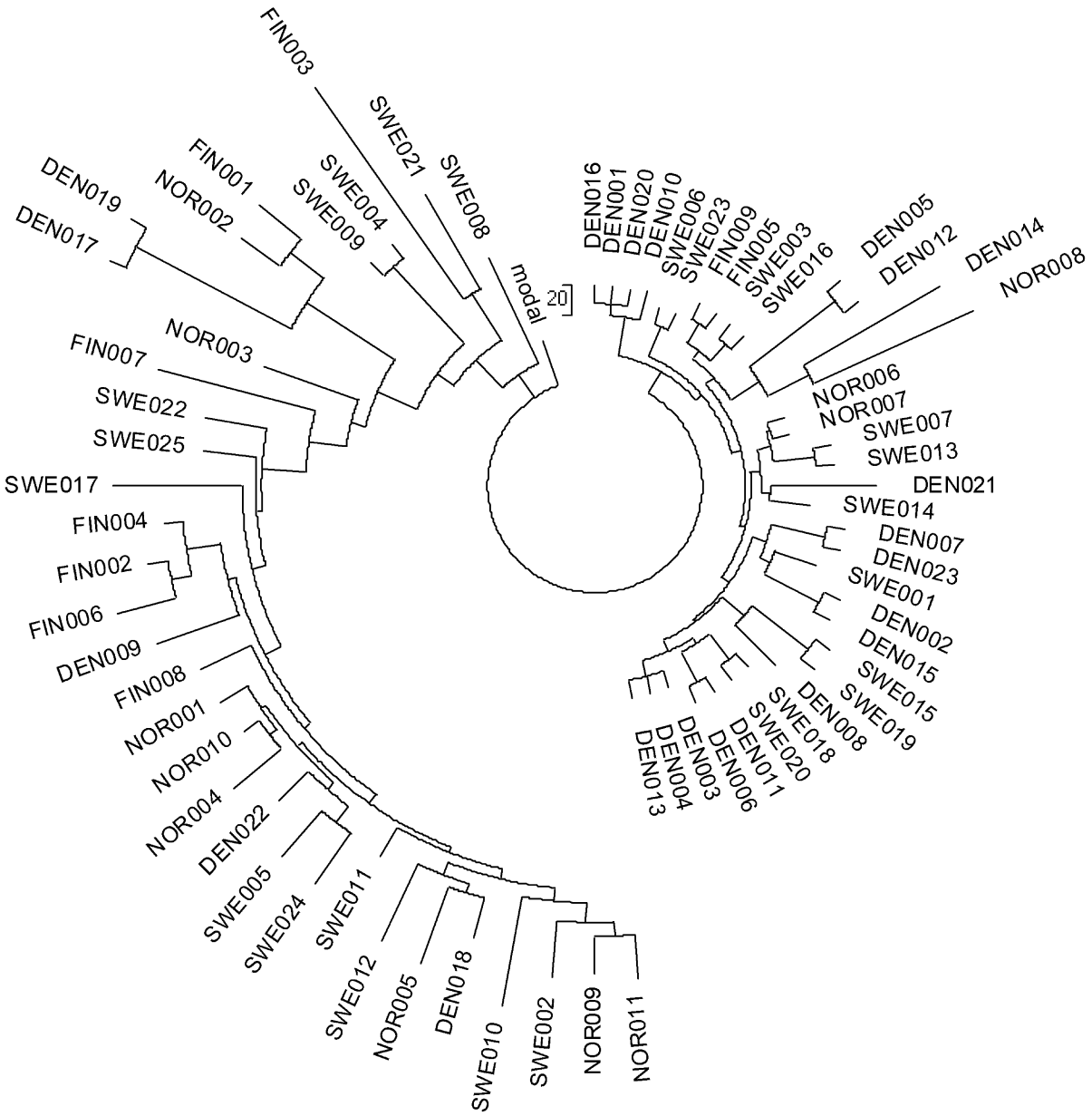

Fig. 6. The 25-marker haplotype tree of haplogroup I2 for Scandinavia. The tree consists of 68 haplotypes taken from the YSearch database. Indexes correspond to Sweden, Norway, Denmark, and Finland.

and other alleles differ by only fractions of one) is **13-22-14-10-**13-**14**-11-14-**11-12-11-**28 – **15**-8-9-8-11-23-**16-20**-28-12-14-15-16, and it was found in Western (857 haplotypes), Central (284 haplotypes), and Eastern European (74 haplotypes) series of I1 haplotypes, and separately in Germany (276 haplotypes) [16] with "ages" of their common ancestors of 3425 ± 350, 3425 ± 350, 3225 ± 360, and 3225 ± 330 years, respectively. On top of this there is a small series of quite different European I1 haplotypes, from France, England, Switzerland, Germany, and Poland, which fall apart in their mutations so much that it should take as many as 21,400 years for such a variation. This is a minimal "age" for a common ancestor of haplogroup I1 in Europe. A common ancestor of both haplogroups I1 and I2, that is a common ancestor of the upstream haplogroup I, should have lived more than 30 thousand years ago, and probably more than 40 thousand years ago.

Let us move to ethnic Russians of the Finno-Ugric origin, haplogroup N1c. They are the same ethnic Russians as those described above and they were selected for testing using the same criteria as all other ethnic Russians. They are Slavs, as well as the "Western Slavs" and the "Eastern Slavs". On the analogy, they can be called the "Northern Slavs". The main difference in their origin is that ancient bearers of haplogroup N1c had migrated thousands years ago from South Siberia, apparently from the Altai region, and the language of their predecessors was assigned by linguists to the Altai language family, and then to the Ural group of languages. Based on the hypothesis on the Ural-Altai family of languages, which existed since the 18th century, it can be suggested that bearers of haplogroup N1c are rather Ural folks than the Finno-Ugric ones.

Figure 7 shows a haplotype tree of ethnic Russians of haplogroup N1c. Analysis showed that all the haplotypes

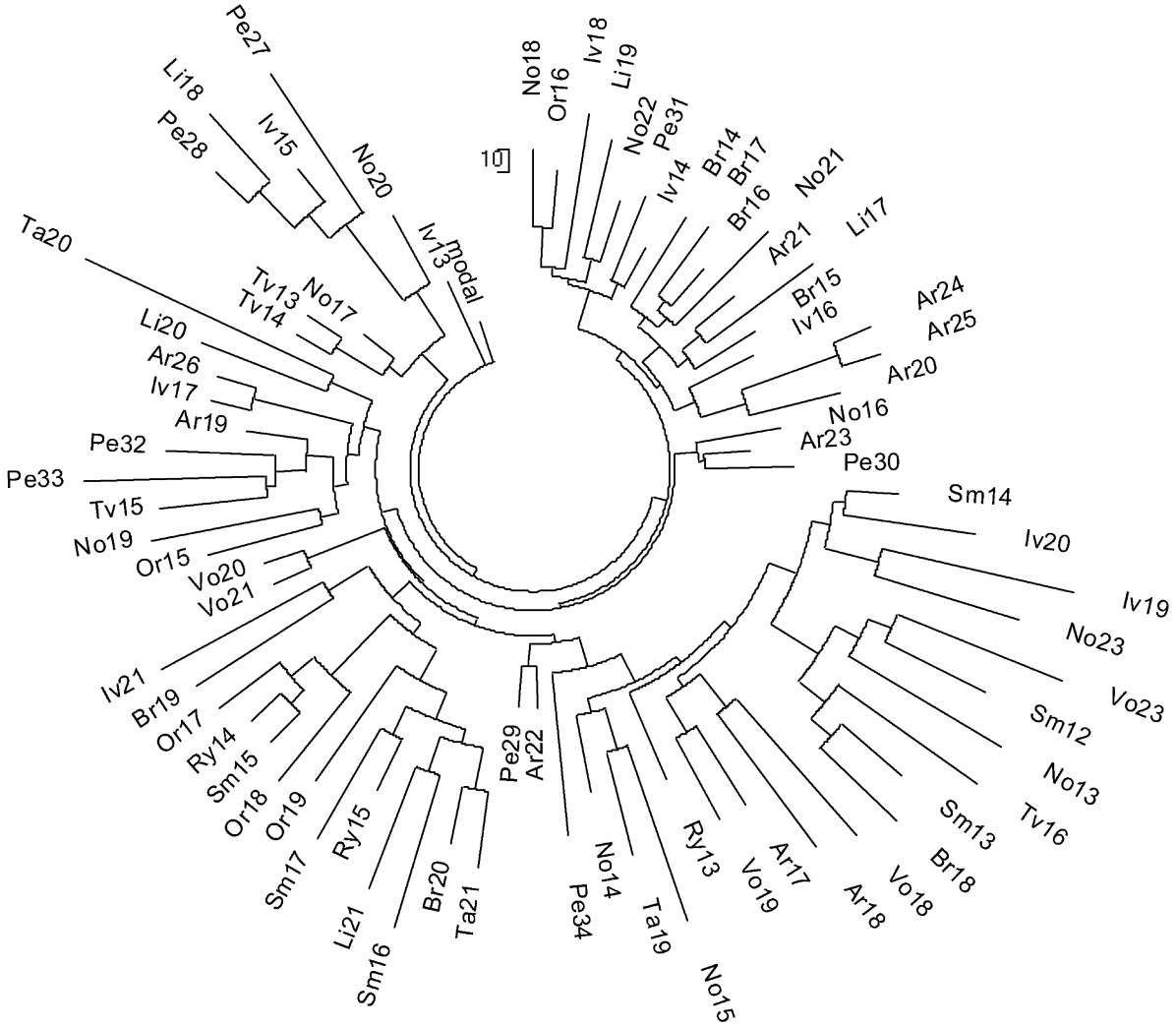

Fig. 7. Tree of 76 ethnic Russian 17-marker haplotypes of haplogroup N1c collected in the 12 regions of the Russian Federation listed in Fig. 1. The tree was composed using haplotypes listed in [3].

descended from one common ancestor whose 17-marker base haplotype is 14 11 13 14 30 23 11 14 14 14 10 10 19 14 17 22 11. It differs from those of R1a1 and I2 by 18 and 23 mutations, respectively. It shows again that the common ancestors of these haplogroups lived tens of thousands of years apart and apparently on different territories, and that phylogenetically haplogroups N1c and R1a1 are closer to each other than haplogroups N1c and I2.

All 76 haplotypes in Fig. 7 contain 317 mutations from their base haplotype shown above [12], pointing to a relatively small number of mutations per marker (0.245 ± 0.028), which in turn translates to 3525 ± 400 years to a common ancestor of ethnic Russians of haplogroup N1c. This was the middle of the second millennium BC, when the Aryans (haplogroup R1a1) were moving to Hindustan and the Iranian Plateau. Scientists later called them "Indo-Aryans", "Avesta Aryans", or "Iranians" for the sake of classifications.

Let us consider what is described above. A few examples have been described in this paper, but the author has been studying dozens and probably hundreds of examples of DNA genealogy trees which he has considered in recent years employing approaches of biochemistry, molecular biology, and chemical and biological kinetics. His work at Harvard University and in a biomedical company as Professor and Chief Scientist has made this work possible, along with his principal professional task aimed at drug design and developing new approaches to cure cancer and other acute inflammatory pathologies [17, 18]. This overview provides just an outline of a possible role that DNA genealogy can play in further understanding of origin, dynamics, and migrations of human tribes, clans, peoples, genealogical lineages, as well as the languages of mankind both in territory and time. Unlike anthropology, history, and archaeology, which often deal with ancient artifacts recovered from the ground of Earth, DNA genealogy deals with molecules extracted from our bodies. After a series of rather routine technical operations, often conducted by a skilled technician using an almost fully computerized system, a researcher obtains data on mutations, which have accumulated in certain fragments of the DNA, and can compare patterns of those mutations in individuals and groups of people, populations, ethnic groups, and races. This can give unique information on movements of the ancestors of both present-day and excavated bearers of DNA molecules, again in territory and time, down to times of 50,000-80,000 years before present and on any territory. In principle, this search is not restricted by the mentioned dates and can (and will) be extended to hundreds of thousands and millions of years in depth, but currently science does not have the means to do the job properly and knowledgeably.

Another challenge in DNA genealogy is posed by what geneticists call population bottlenecks. Any human population suffered from some kind of a disaster at some points in the past, and countless number of times populations may not have survived. In reality it means that a population, let it be a family, a tribe, a village, or a settlement, is reduced in number of individuals so much that either their existence ceased, the lineage(s) die out (have not passed the population bottleneck), or it reduces to several or even one male, whose offspring eventually survives and multiplies in numbers. In those cases a surviving male becomes the common ancestor of the surviving lineage. Calculations of a time span to a common ancestor typically lead to that surviving individual, since mutations in the DNA of our contemporaries are counted from such a person. He becomes, in terms of DNA genealogy, a common ancestor of the given population.

Then, in terms of DNA genealogy, "a common ancestor" means not necessarily one person, but one haplotype (the base haplotype). In reality there could be several individuals having the same haplotype, such as father and son(s), brothers, a group of relatives, etc. It could have been several practically unrelated people, only one of which survived in future generations, and becomes the base haplotype from which mutations have started to multiply.

Methods of DNA genealogy allow an understanding as to when that common ancestor lived, hence when that population bottleneck had occurred on the absolute time scale. Even with modest statistics provided, namely, by considering only tens, better if hundreds of haplotypes, sometimes even thousands of DNA specimens, hence, haplotypes time spans to common ancestors can be calculated with the accuracy of plus-minus 10% at 95% confidence. Thus, if a common ancestor of a given population lived, say, 5000 years ago, then having a hundred of 25-marker haplotypes of his descendents who live today, a time span to the common ancestor can be calculated as 5000 ± 530 years. This would mean that the time period when the common ancestor lived would fit between 4470 and 5530 years before present, with 95% confidence.

Such an accuracy (or a lack of it for some) results from a number of "experimental points", that is the total number of markers in haplotypes under consideration. A hundred 25-marker haplotypes contain 2500 markers total, and each one of them mutates in accord with its mutation rate constant in the course of those thousands of years. The 2500 markers will produce ~770 mutations during those 5000 ± 530 years. The mutation rate constant for the 25-marker haplotypes equals 0.046 mutation per haplotype per 25 years, or 0.00183 mutation per marker per 25 years, as has been explained above. These values were established and calibrated [1]. If there are less than 770 mutations in a series of 100 of 25-marker haplotypes, then a common ancestor lived more recently than ~5000 years ago, and a respective time span can be calculated rather reliably. For example, an analysis of 750 of 19-marker haplotypes of Basques (Pyrenees) of haplogroup R1b1b2 has shown that all of them contain 2796

mutations from their base haplotype and all the mutations had a degree of symmetry of 0.5. It means that exactly 50% of all the mutations increased numbers of alleles, and 50% decreased their number, that is mutations which occurred quite randomly. This number of mutations is translated to 3625 ± 370 years before present for the lifetime of a common ancestor of all the 750 Basques. In reality, bearers of that haplogroup (R1b1b2) arrived in the Pyrenees around 4800 years ago and brought the Bell Beaker archaeological culture to continental Europe. Apparently, the Basques passed through a population bottleneck between 4800 and 3625 ± 370 years ago, since their common ancestor was about a thousand years "younger" compared to the arrival time of the R1b1b2 bearers to Pyrenees.

There is another approach in DNA genealogy in which mutations in haplotypes are not counted. Instead, one can count just a number of base haplotypes in the haplotype dataset. Those base haplotypes are identical to each other. Their numbers are determined by the laws of chemical kinetics, or physical chemistry, and by the first-order kinetics in this particular case when the whole haplotype dataset is derived from the same common ancestor. The same law of first-order kinetics is applicable in radioactive decay processes, though the latter are simpler compared to the mutation processes in the DNA. The thing is that a stable isotope, which typically is a product of a radioactive decay, cannot turn back into a radioactive material; the decay is an irreversible process. On the contrary, a mutation can proceed back or forth with equal probability, because mutations are random events. In other words, a mutated marker can proceed further with a second mutation in the same direction, or come back as if there were no mutation in the first place. This feature, that is reverse mutations, or back mutations, complicates the calculations compared with those for the first-order process of radioactive decay. Obviously, back mutations lead to an apparent reduction in the mutation number, and, if overlooked, lead to underestimations of a time span to a common ancestor. However, this problem is successfully solved in DNA genealogy, and handy correction tables have been published [1, 13, 14].

Coming back to the Basques dataset, 750 of their haplotypes contain 16 base ones, that is after 3625 ± 370 years of random mutations in the 17-marker haplotypes in full accord with the theory of probability 16 haplotypes from 750 did not enter into a mutated state, they essentially sit on the tail of the mutation distribution curve. If to apply formal law of first-order kinetics to that system, then from the 750 haplotypes 16 would stay the same after $\ln(750/16)/0.0285 = 135$ generations passed after the life time of the common ancestor of the lineage, that is after about apparent (not corrected for back mutation) 3375 years. As explained above, a generation of 25 years long is "built in" to the mutation rate constant of 0.0285 mutation per haplotype per generation. When we introduce a correction for back mutations [1], again in full accord with the theory of probability, the number of generations is increased from 135 to 156, and the time span to the common ancestor becomes 3900 years. One can see that the 3900 years fits into the 95% confidence interval 3625 ± 370 years to a common ancestor. The fit between mutation counting (the linear method) and base haplotype counting (the logarithmic method) shows that the system is a correct one for DNA genealogy, since it has only one common ancestor.

This particular example indicates that counting mutations or counting base haplotypes are equally proper in DNA genealogy. Of course, counting base haplotypes makes sense only if there is not too long a distance to a common ancestor of the population, otherwise there would be no base haplotypes left among present-day descendants. In a recent study [1] values of mutation rate constants are listed for 22 typical haplotype formats employed in DNA genealogy from the 5-marker to 67-marker ones. Since then the table has been extended to 32 typical haplotypes.

To show how calculations can be massive, we will cite a study of 857 of 25-marker haplotypes of haplogroup I1 in England, which has 21,425 "experimental points" [1, 7]. Those haplotypes contained 4868 mutations from their base haplotype. The term "base" instead of the more simplified "ancestral" is employed because in many cases it is not quite known that the coalescent (extrapolated) haplotype is necessarily the ancestral one, particularly when the common ancestor is an ancient one. This might be just an approximation, a stretch. Hence the term "base" haplotype.

As many as 4868 mutations allow us to determine an average number of mutations per marker with accuracy of 2-3% with 95% reliability. This is why for asymmetrical mutations an accuracy of calculations (that is the average square deviation) of the average number of mutations per marker is determined as a reciprocal square root from the number of mutations, and for symmetrical mutations it is defined quite similarly [1]. For 4868 mutations this deviation equals 2.87% with 95% reliability. However, since an accuracy of average mutation rate constants in haplotypes is commonly not better than 5%, for the 95% confidence a time span to a common ancestor can be determined with accuracy no better than 5%, hence the 10% confidence interval.

Based on this, it was determined for the aforesaid English I1 haplotypes that the common ancestor of the 857 individuals lived 3425 ± 350 years before present. These are relatively "young" European populations (in terms of their common ancestor), though some tribes had common ancestors who lived – according to DNA genealogy data – some 16-23 thousand years ago. For a number of Asian populations their common ancestors lived 16-21 thousand years ago, for some African tribes they lived 28-37 thousand years ago. These are numbers

calculated from the DNA of their descendants who live now, hence, their lineage had survived.

These examples are given here to show that DNA genealogy is not just an emerging field of science built upon shaky ground, which is often observed for many new fields of science. For the last several years DNA genealogy has practically completed with its computational basis, specified its principal definitions and – in general – its paradigm, and calculations are now performed with good reliability and accuracy. It was experimentally shown, using genomes of many human populations, as well as with genomes of chimpanzee, that mutations represent a kind of a "molecular clock", with a "ticking" (that is, mutation) rate in the DNA that has been pretty much constant during the last two million years at least [19].

It was experimentally shown using thousands of father–son pairs that mutations in haplotypes are indeed equally likely in the "both directions", and that the tandem mutations, that is insertions or deletions of repeating blocks of certain nucleotides, and the tandem, that is repeating blocks of nucleotides, called markers, can be decreased or increased in length, that is changes of a number of alleles, with the same probability. It was shown using father–son pairs that double or other "multistep" mutations in haplotypes and markers happen quite rarely, in fractions of 1 or 2%, and practically do not affect time span values to common ancestors of populations and a series of haplotypes [20].

Identification of a geographical location, that is where the common ancestor lived, is a much more complicated problem to solve since the timing of his life does not point to a geography. To figure out where in which territories common ancestors of the given population lived, one needs to employ independent data of archaeology, linguistics, and anthropology, with a full understanding that this data could be irrelevant if haplotypes of those excavated people did not survive to our times.

This is exactly why a union between anthropology, archaeology, linguistics, and DNA genealogy is so important. In exchange, DNA genealogy provides to those disciplines a distinct mark of the tribe, a population, which is a well-defined mutation (SNP) in a certain position in the Y chromosome, which physically accompanied each member of the tribe. This mutation is not assimilated in populations, as assimilated languages, cultures, religions, physical features, anthropological parameters, or their indexes.

This mutation, SNP, stays the same in populations; it can help to distinguish a member of the population after thousands and tens of thousands of years of evolution. It can trace migrations of certain tribes and their individual representatives, dead or alive, trace remnants of tribe members at archaeological excavations, and it can help to understand how archaeological cultures have been connected to each other – not only using material artifacts, but via people's DNA, it can add important components to the dynamics of human populations and their lineages.

It seems that in a number of cases descendants of individuals in ancient settlements 45-50 thousand years ago have survived to the present time. However, this can be verified only via DNA analysis of excavated bones. Technically it can be made by determination of nucleotide sequences of fragments of DNA in the bones, which was sealed there and survived over many millennia. Those studies are rare at the present time, they are very complicated and difficult, and very expensive, but they are coming up more often.

It should be emphasized once more that DNA genealogy of present-day people reaches only population bottlenecks, connecting our contemporaries to their direct ancestors. These bottlenecks are not only the results of epidemics, wars, natural cataclysms in the past, though all those factors greatly affected the structure of human populations in the past. It is hard to imagine how much the black plague in the middle of the 14th century, which killed a quarter of the European population, had changed the landscape of haplogroups and haplotypes in Europe. How many lineages and whole haplogroups were terminated? Many genealogical lineages have started in the middle of the 14th century. They were started by survivors, some of which became "common ancestors" of new lineages extended into our modern times.

Genocide is the worst enemy of DNA genealogy, as well as of mankind. How many genealogical lineages had been terminated by massacres in Armenia in the first quarter of the 20th century, genocide of the Jews, gypsies, other ethnic groups, religious and political groups and other populations? We can only guess. How many genealogical lineages have been broken by the world wars, how many of them are being terminated violently right now...? Any war makes an irreversible contribution to destroying genetic material. The Gaul wars of Julius Caesar led, according to Plutarch, to the death of more than one million people in Central Europe, and a million more were taken into slavery. We can only guess how this changed the haplogroup and haplotype landscape in Europe.

However, population bottlenecks occur not only as a result of mortal events, but also after movements to new territories, migrations of haplotype bearers, and by similar rather peaceful events. In mass migrations, or a migration of even a few individuals, they "carry" mutated haplotypes of their common ancestors to a new place. In terms of DNA genealogy, a common ancestor of the migrants might be the same for a new territory, having a time span to it counted in millennia. For example, a common ancestor of descendants of English Pilgrims is about the same both in England and in the US. A common ancestor of the Russians of all the major haplogroups is about the same in Russia and in the US. This, of course, does not mean that Eastern Slavs, for instance, lived in North America 4500 years before present, concurrently

with those on the Russian Plain. Nevertheless, similar conclusions are a typical mistake of beginners in DNA genealogy. Another example – Iceland was populated only since the 9th century AD, but common ancestors of Icelanders in all major haplogroups lived thousands of years ago, as in continental Europe.

A time span to a common ancestor of haplogroup R1b1 calculated from mutations in haplotypes in Central Asia (particularly in the Altai region) stretches to around 16 thousand years before present; that calculated from haplotypes collected among ethnic Russians points to 6775 years ago; on the Caucasus (subclade R1b1b2-L23) 6000 years ago; in the middle East, in Lebanon, and among the Jews of haplogroup R1b1b2 5500-5200 years ago; in the Pyrenees 4800 years ago; in France 4200 years ago; in Ireland 3800-3400 years ago. Furthermore, subclades of the R1b1 haplogroup follow the same route going from upstream to downstream ones. This trek reflects a proper direction and route of migrations of bearers of haplogroup R1b1 between 16,000 and 3-4 thousand years before present. More than that, it also allows us to connect this R1b1 tribe with certain archaeological cultures, among them the Beaker Culture, which nicely fits to the migration route from Pyrenees to France to Northern Europe to British Isles.

Hence, a new name for the new science: Molecular History. This is making (defining) historical reconstructions based on molecular characteristics of the DNA descendants of ancient lineages and/or excavated DNA. Since ancient ancestors in the course of their migrations carried their languages to territories far away, studying those migrations of the DNA (haplogroups and haplotypes) which took place hundreds, thousands, and more years ago, we can obtain new knowledge on migrations of languages thousands of years ago. A comparison of these data with those obtained by linguists using different in kind methods can examine and verify their more traditional interpretations and concepts, and sometimes (or always?) come to quite unexpected, new concepts and ideas.

Thus biochemistry has turned with its unexpected and unpredictable edge to other fields of research, particularly in history, linguistics, and archaeology. It is very desirable that DNA genealogy contributes to the peaceful and good relationships between peoples on our planet. It shows again and by quite a direct way that all of us, independently of racial and religious preferences, are descendants of the same common ancestors. Analysis of haplotypes of all populations in the world shows that all people are relatives, after all. Furthermore, we can calculate already quite reliably the degree of those relationships, and biochemistry plays a significant role in the degree of that reliability.

In conclusion, the author wants to emphasize that he does not ascribe to himself a role of the pioneer in this new field of science. A modest role of the author is in bringing about some quantitative measures and tools for analysis of mutation patterns in haplotypes, in quantitative analysis of haplotype trees and their branches, in more accurate calculations of time spans to common ancestors of populations, in calibrations of mutation rate constants, and in creating "a measuring stick" for a new line of studies in history based on DNA of descendants of ancient participants in those historical events. The primary profession of the author by his education is quantitative descriptions of time-dependent events, that is chemical kinetics in physical chemistry, and it helps in analyzing kinetics of mutations in DNA. This paper certainly was not tasked to give a comprehensive review of the history of creation and development of sciences from which the current DNA genealogy is stemming. It should be noted in all fairness that DNA genealogy has descended from molecular biology, genetics, and population genetics, which was and is being developing by many specialists in the area. In the development of population genetics of humans in the 1990s a major role was played by scientists such as (though, of course, personal preferences of the author of this study can be noted): L. Cavalli-Sforza, M. Feldman, D. Goldstein, M. Hammer, M. Jobling, T. Karafet, M. Kayser, P. de Knijff, A. Nebel, M. Nei, A. Oppenheim, O. Semino, M. Stoneking, M. Thomas, P. Underhill, B. Walsh, R. Wells, L. Zhivotovsky, and many others who are just impossible to mention here in entirety (see, for example, [21-36]).

These and dozens of other researchers have created a wealth of knowledge that made it possible to assign many SNPs to principal human tribes and to introduce a concept of haplogroups and haplotypes in human DNA, to reconstruct a hierarchy of subclades in those haplogroups and assign them to downstream families of mankind and their DNA-lineages. The SNPs, or *snips*, are being carefully considered and classified to achieve a maximum differentiation between human families, and at the same time to show clear connections between them. It can be mentioned here that there are 20 main tribes (haplogroups) on the Earth, which are named according to letters of the Latin alphabet, from A to T; however, their downstream subclades are counted already in many hundreds, and they are getting to a thousand, albeit the number of snips are in the many millions. However, since the volume of this paper is limited, we have to stop at this optimistic "many millions" point.

REFERENCES

1. Klyosov, A. A. (2009) *J. Genet. Geneal.*, **5**, 186-216.
2. Klyosov, A. A. (2009) *Hum. Genet.*, **126**, 719-724.
3. Roewer, L., Willuweit, S., Kruger, C., Nagy, M., Rychkov, S., Morozowa, I., Naumova, O., Schneider, Y., Zhukova, O., Stoneking, M., and Nasidze, I. (2008) *Int. Legal Med.*, **122**, 219-223.
4. Klyosov, A. A. (2009) *Proc. Russian Acad. DNA Geneal.*, **2**, 232-251.

5. Klyosov, A. A. (2008) *Proc. Russian Acad. DNA Geneal.*, **1**, 400-477.
6. Klyosov, A. A. (2008) *Proc. Russian Acad. DNA Geneal.*, **1**, 947-957.
7. Klyosov, A. A. (2009) *J. Genet. Geneal.*, **5**, 217-256.
8. Klyosov, A. A. (2009) *Proc. Russian Acad. DNA Geneal.*, **2**, 1217-1229.
9. Klyosov, A. A. (2010) *Proc. Russian Acad. DNA Geneal.*, **3**, 249-299.
10. Sharma, S., Rai, E., Sharma, P., Jena, M., Singh, S., Darvishi, K., Bhat, A. K., Bhanwer, A. J. S., Tiwari, P. K., and Bamezai, R. N. K. (2009) *J. Hum. Genet.*, **54**, 47-55.
11. Klyosov, A. A. (2009) *Proc. Russian Acad. DNA Geneal.*, **2**, 801-815.
12. Klyosov, A. A. (2009) *Proc. Russian Acad. DNA Geneal.*, **2**, 370-389.
13. Klyosov, A. A. (2008) *Proc. Russian Acad. DNA Geneal.*, **1**, 252-348.
14. Klyosov, A. A. (2008) *Proc. Russian Acad. DNA Geneal.*, **1**, 812-835.
15. Adamov, D. S., and Klyosov, A. A. (2009) *Proc. Russian Acad. DNA Geneal.*, **2**, 422-442.
16. Klyosov, A. A. (2010) *Proc. Russian Acad. DNA Geneal.*, **3**, 96-158.
17. Klyosov, A. A., Witczak, Z. J., and Platt, D. (2006) *Carbohydrate Drug Design*, Cambridge University Press, Cambridge.
18. Klyosov, A. A., Witczak, Z. J., and Platt, D. (2008) *Galectins*, John Wiley and Sons.
19. Sun, J. X., Millikin, J. C., Patterson, N., and Reich, D. E. (2009) *Mol. Biol. Evol.*, **26**, 1017-1027.
20. Klyosov, A. A. (2010) *Proc. Russian Acad. DNA Geneal.*, **3**, 1853-1860.
21. Cordaux, R., Bentley, G., Aunger, R., Sirajuddin, S. M., and Stoneking, M. J. (2004) *Forensic Sci.*, **49**, 1-2.
22. Goldstein, D. B., Linares, A. R., Cavalli-Sforza, L. L., and Feldman, M. W. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 6723-6727.
23. Hammer, M. F., Redd, A. J., Wood, E. T., Bonner, M. R., Jarjanazi, H., Karafet, T., Santachiara-Benerecetti, S., Oppenheim, A., Jobling, M. A., Jenkins, T., Ostrer, H., and Bonne-Tamir, B. (2000) *Proc. Natl. Acad. Sci. USA*, **97**, 6769-6774.
24. Heyer, E., Puymirat, J., Dieltjes, P., Bakker, E., and de Knijff, P. (1997) *Hum. Mol. Genet.*, **6**, 799-803.
25. Jobling, M. A., and Tyler-Smith, C. (1995) *Trends Genet.*, **11**, 449-456.
26. Karafet, T. M., Zegura, S. L., Posukh, O., Osipova, L., Bergen, A., Long, J., Goldman, D., Klitz, W., Harihara, S., de Knijff, P., Wiebe, V., Griffiths, R. C., Templeton, A. R., and Hammer, M. F. (1999) *Am. J. Hum. Genet.*, **64**, 817-831.
27 Kayser, M., Roewer, L., Hedman, M., Henke, L., Hemke, J., Brauer, S., Kruger, C., Krawczak, M., Nagy, M., Dobosz, T., Szibor, R., de Knijff, P., Stoneking, M., and Sajantila, A. (2000) *Am. J. Hum. Genet.*, **66**, 1580-1588.
28. Nebel, A., Filon, D., Weiss, D. A., Weale, M., Faerman, M., Oppenheim, A., and Thomas, M. (2000) *Hum. Genet.*, **107**, 630-641.
29. Nebel, A., Filon, D., Brinkmann, B., Majumder, P. P., Faerman, M., and Oppenheim, A. (2001) *Am. J. Hum. Genet.*, **69**, 1095-1112.
30. Nei, M. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 6720-6722.
31. Semino, O., Passarino, G., Oefner, P. J., Lin, A. A., Arbuzova, S., Beckman, L. E., de Benedictis, G., Francalacci, P., Kouvatsi, A., Limborska, S., Marcikiae, M., Mika, A., Mika, B., Primorac, D., Santachiara-Benerecetti, A. S., Cavalli-Sforza, L. L., and Underhill, P. A. (2000) *Science*, **290**, 1155-1159.
32. Takezaki, N., and Nei, M. (1996) *Genetics*, **144**, 389-399.
33. Underhill, P. A., Shen, P., Lin, A. A., Jin, L., Passarino, G., Yang, W. H., Kauffman, E., Bonne-Tamir, B., Bertranpetit, J., Francalacci, P., Ibrahim, M., Jenkins, T., Kidd, J. R., Mehdi, S. Q., Seielstad, M. T., Wells, R. S., Piazza, A., Davis, R. W., Feldman, M. W., Cavalli-Sforza, L. L., and Oefner, P. J. (2000) *Nature Genet.*, **26**, 358-361.
34. Walsh, B. (2001) *Genetics*, **158**, 897-912.
35. Wells, R. S., Yuldasheva, N., Ruzibakiev, R., Underhill, P. A., Evseeva, I., Blue-Smith, L., Jin, L., Su, B., Pitchappan, R., Shanmugalaksmi, S., Balakrishnan, K., Read, M., Pearson, N. M., Zerjal, T., Webster, M. T., Zholoshvili, I., Jamarjashvili, E., Gambarov, S., Nikbin, B., Dostiev, A., Aknazarov, O., Zallous, P., Tsoy, I., Kitaev, M., Mirrakhimov, M., Chariev, A., and Bodmer, W. F. (2001) *Proc. Natl. Acad. Sci. USA*, **98**, 10244-10249.
36. Zhivotovsky, L. A., and Feldman, M. W. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 11549-11552.