ISSN: 1920-2997 http://rjgg.org

European prehistory in mirror of genetics: a contemporary view*

Oleg Balanovsky[1]

Introduction

Numerous branches of knowledge are currently affected by a certain eurocentrism, this being especially the case of the studies focused on human population genetics. These studies were (and remain) developed predominantly in European laboratories and are less popular on other continents. Population geneticists need human populations as subject matter of their research, and in most cases we study those which are in proximity of our homes. That is why Europe is the continent which is by far better studied genetically, and the European genetic landscape is much deeper understood and discussed than any other part of the world. For the same reason the controversies between different schools of thought regarding various aspects of the European gene pool became much more apparent

The two concepts

The variation of "classical" genetic markers (which became referred to like that when the DNA came to the fore to replace them) was best summarized in the book by Luigi Luca Cavalli-Sforza and his colleagues published in 1994. Unsurprisingly, the largest chapter of this book is the "European" one, which describes how the European genetic landscape had been formed during the Neolithic expansion from the Near East. It was one of the most minutely-elaborated concepts in population genetics at that time; nonetheless it was almost entirely rejected in the subsequent decade.

The European genetic landscape, as restored based on the analysis of classical markers, shows three principle features: 1) a general homogeneity (the Europeans are genetically very similar to each other, compared to populations of other continents); 2) the presence of only a few outliers (isolated peripheral populations such as Icelanders, Saami, or Sardinians); their peculiarities are the secondary, having arose after these populations were demographically split off and underwent the genetic drift from the main European corpus; 3) Clear geographic patterns of gradual genetic changes.

To identify these geographic patterns Cavalli-Sforza and his colleagues (Menozzi et al., 1978, Cavalli-Sforza et al., 1994) and independently Russian geneticists (Rychkov & Balanovskaya, 1992) developed the method of "synthetic maps". These maps are created by a complex mathematical algorithm but in a simpler way they consist of displaying the geographic distribution of an "ideal" genetic marker, which correlates with geographical patterns of the majority of real markers presenting the data (Menozzi et al., 1978; Rychkov & Balanovskaya, 1992; Balanovskaya & Nurbaev, 1997a). This synthetic map visually demonstrated gradual changes with a remarkable geographical pattern: from Anatolia via the Balkans over the rest of Europe i.e. from the Southeast to the Northwest (Fig 1). This picture was interpreted as a result of the gradual spread of farming (and farmers) across Europe which was known since Gordon Childe (1928) to follow the same trajectory.

This concept was additionally substanti-

Received: June 8 2009; accepted: June 8 2009; published June 30 2009
[1]Research Centre for Medical Genetics, Russian Academy of Medical Sciences, Moscow, Russia
*This study has been originally published in: Balanovsky O.P. (2009) Human genetics and Neolithic dispersal' The East European Plain on the eve of agriculture edited by Pavel M. Dolukhanov, Graeme R. Sarson and Anvar M. Shukurov. BAR -S1964. ISBN 978 1 4073 0447 2. P:235-46

ISSN: 1920-2997 http://rjgg.org

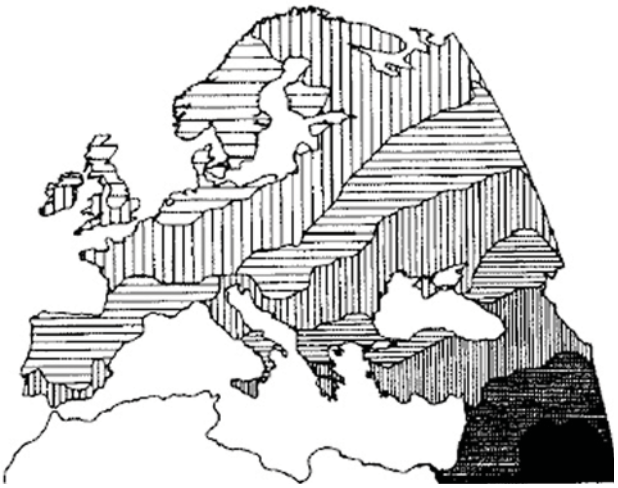

Figure 1. Synthetic map, summarizing genetic variation in Europe (from Cavalli-Sforza et al., 1994).

ated in two ways. First, the "isogenes" (lines connecting the same gene pools on the genetic map) have shown a remarkable agreement with isochrones (lines showing the early arrival of agriculture based on archaeological and radiometric evidence). Second, the concept and the mathematical model of the so-called demic diffusion was developed (Ammerman & Cavalli-Sforza, 1984). It implies a slow (generation by generation) migration of farmers which assimilated indigenous populations, and thereby gradually dissolved the initial "farming" gene pool. As a consequence, the geographic trajectory of migration becomes a geographic line of gradual genetic changes: from a "mainly farming" gene pool in Anatolia to a "mainly indigenous" one in the Europe's north-west and north-east (as the most distant from Anatolia).

This elegant, reasonable and sufficiently substantiated concept of the origin and composition of the European gene pool dominated population genetics in the 1980-1990s. Generally, four elements of this concept could be potentially criticised:

i) the data-set (classical markers);

ii) the methodology (synthetic maps);

iii) the logical foundation (attributing south-east-northwest pattern to Neolithisation) or

iv) controversial results obtained with a use of independent data, methods and logics. Critics used all four elements but with a varaible success.

The popular idea that classical markers are "worse" than new DNA markers has never been positively proven and should be considered rather as a scientific fashion. Yet some critics tend to reject the classical markers arguing that they are affected by natural selection, and therefore their variation would be the result of both historical and biological factors. However, many DNA markers can be equally affected by biological factors and therefore geographic distribution of a genetic marker reflects the history which to some degree was blurred by biological selection affecting this marker. And, second, the biological factors differently affect various markers and therefore disappear when averaging, in the case when numerous markers are considered (Yamazaki & Maryama, 1973; Lewontin & Krakauer, 1975; Balanovskaya & Nurbaev, 1997b).

The method of synthetic maps was attacked by Robert Sokal, who used an alternative method (autocorrelation analysis) for revealing geographical patterns in the genetic data (Sokal, Oden, 1978). Using computer simulation he demonstrated that synthetic maps compiled from interpolated maps produce gradual pattern even from randomly permutated data, hence the obtained patterns are artificial (Sokal et al., 1999). However, our recent simulations (Balanovsky et al., 2008) failed to recognise any difference between synthetic maps from interpolated surfaces (criticized by Sokal and colleagues), on the one hand, and from non-interpolated raw data (considered as control by Sokal and colleagues), on the other. It is equally remarkable that Sokal and colleagues did not express doubt that the gradual genetic pattern from Anatolia is the main feature of the European gene pool, yet they did question the methods applied for to identify this pattern.

Curiously enough, the applied logic conclusion (attributing the observed genetic pattern to the Neolithic expansion as the both followed the same trajectory) had never been criticized to the best of our knowledge, though population geneticists were well aware that a correlation never proves the cause-effect relationship. The interpretation in terms of the Neolithic expansion seemed so obvious, transparent, and natural, that this logical mistake became only apparent when controversial results started emerging from independent data.

The evidence, demonstrating the Palaeolithic time for the origin of the European gene pool was based on mitochondrial DNA (mtDNA). The principal difference between mtDNA and Y chromosomal markers on the one hand, and the classical and autosomal DNA markers on the other, resides in the presence or absence of recombination. Autosomal markers recombine and therefore each marker is inherited independently from all other markers. MtDNA and the main portion of the Y chromosome do not recombine. That is why every occurring mutation

ISSN: 1920-2997 http://rjgg.org

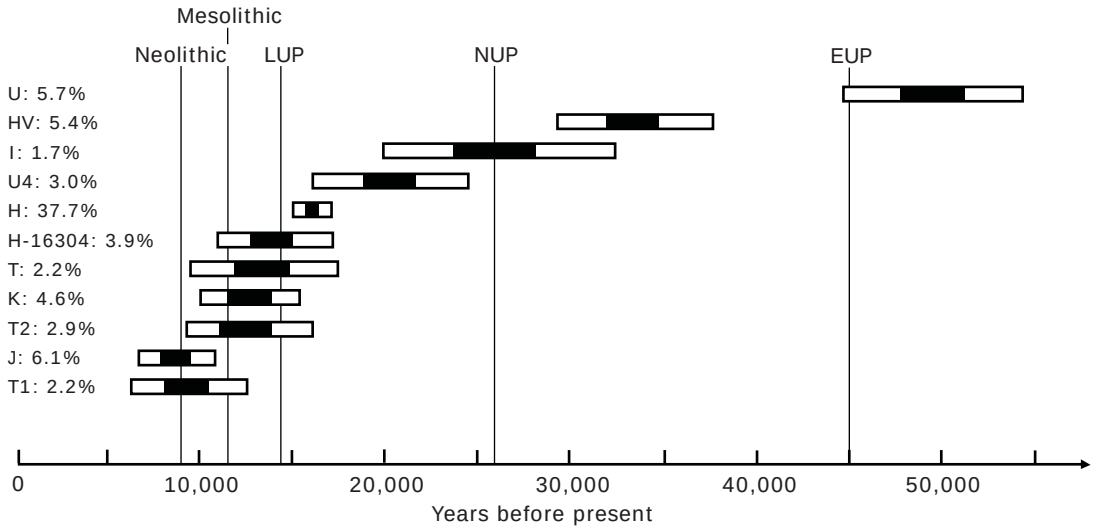

Figure 2. Ages of mitochondrial haplogroups in Europe (from Richards et al., 2000).
EUP - Early Upper Palaeolithic; MUP – Middle Upper Palaeolithic; LUP – Late Upper Palaeolithic.

gets transmitted from generation to generation alongside other mutations, which did occur earlier. In other words, the mutation (a mistake in the genetic text) became forever part of this text and is transmitted together with all the mistakes that had appeared in this text earlier. When comparing different texts (so-called haplotypes) it is possible to trace mutations back in time and reconstruct a "genealogy" of these texts. i.e. to draw their "family tree". This tree is commonly rooted in the most recent common ancestor (a "mitochondrial Eve") and each branch of the tree differs by its particular set of mutations. Next, each twig of a certain branch carries all mutations characteristic to this branch, together with a set of additional "twig-specific" mutations. These branches and twigs are called haplogroups (subhaplogroups) each of them unites a group of closely related haplotypes (which can be compared with a leaves of this tree). Assuming an average rate of mutations one can calculate the age of each haplogroup by multiplying the number of accumulated mutations by the mutation rate.

This methodology, applied to the European mitochondrial pool (Richards et al., 1996), demonstrated that most branches (haplogroups) found in Europe were much older that the Neolithic and most of them fell into the age range of the Upper Palaeolithic. Based on this evidence it was concluded, that European gene pool was formed by the initial peopling of the continent by anatomically modern humans (AMH) during the Upper Palaeolithic, and that it is still present in the most of present-day Europeans. As for the Neolithic expansion, it had, therefore, a limited impact on the European gene pool. Hence, this new concept imposed the "cultural diffusion" model of Neolithisation in contrast to "demic diffusion" model advanced by Ammerman and Cavalli-Sforza (1984).

The following decade witnessed a heated debate between two camps of geneticists, namely the "cultural diffusionists" and the "demists". Despite the ongoing debate, the methodological limits and benefits of both models are apparent. The source database for demic diffusion model was much richer (hundreds of markers studied in dozens of populations) while Richards and colleagues were restricted to one marker (mtDNA) studied in a limited set of populations. Arguably, as Barbujani and colleagues (1998) pointed out, the Palaeolithic origins of haplogroups found in Europeans do not necessarily imply that these haplogroups were present in Europe since the Palaeolithic. As haplogroups age is calculated based on its diversity, they could have accumulated diversity in other parts of the world arriving into Europe being already diverse. This problem of the pre-existing diversity met elegant solution in the following paper by Richards and colleagues (2000), which became the most recognized study of European genetics. In that paper the founder mtDNA lineages were identified which were deemed as the starting points for the entire European diversity accumulated in situ. Although different criteria for "founding" resulted in slightly different time assessments, all calculations demonstrated the Upper Palaeolithic age for most European clusters of lineages (haplogroups), while haplogroups whose appearance in Europe can be attributed to the Neolithic period make up only a quarter of the total European gene pool (Fig. 2).

ISSN: 1920-2997 http://rjgg.org

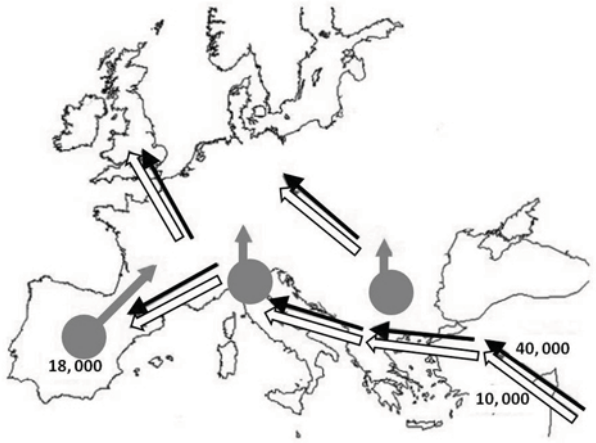

Figure 3. A scheme of the main demographic processes documented in the archeological record of Europe (from Barbujani, Bertorelle, 2001).
Numbers are approximate dates, in years before the present. Black arrows, Paleolithic colonization; grey arrows, Late Palaeolithic recolonization from glacial refugia (grey circles); white arrows, Neolithic demic diffusion.

The opponents of this concept did not miss the opportunity to point out that the deepest time assessment for an in situ European haplogroup was paradoxically older than the age of AMH appearance in Europe (Barbujani, Bertorelle, 2001). It should be noted that the time estimates are based largely on the calibration points used (mutation rate). But nevertheless the ability to present a time estimate was the strongest among of Richards et al. arguments whereas the contrary concept was based exclusively on the similarity between the genetic pattern and that of the Neolithic spread. This enabled one to pinpoint the main logical weak point in the "Neolithic" concept, namely, that the AMH initial settlement of Europe followed the same geographical trajectory which was later used by expanding Neolithic farmers. When Barbujani and Bertorelle (2001) summed up this discussion they admitted, that the gradual "out of Anatolia" geographic pattern, as established by classical and (later) other markers was correct. Yet this pattern could have originated from both, the Palaeolithic and Neolithic migrations, as the both were believed to follow the same Anatolian route (Fig. 3). The lesson learnt was that geographic patterns of genetic variation do not allow distinguish between these scenarios, and one needs non-recombining systems which are essential for time estimations.

Nearly simultaneously with the seminal publication summarising the mtDNA data (Richards et al., 2000), two papers on the second non-recombining system appeared, summing up the paternal perspective, i.e. Y chromosomal variations in Europe (Semino et al., 2000; Rosser et al., 2000). The both papers were based on extensive datasets. Although written in a different manner they established similar features.

Rosser and colleagues followed a phenomenological approach, describing patterns of Y chromosomal variation. They found very clear geographical clines in the distribution of all haplogroups and statistically calculated that the genetic similarity of populations was affected by their geographic proximity rather than linguistic similarity. In contrast, to that Semino and colleagues following an interpretative approach concluded that the observed geographical patterns could have been caused by the factors of similar geographic distribution in the Palaeolithic epoch. One may easily note, that in doing so they committed the same logical mistake as they interpreted geographical pattern "by association" with the known event of the same spatial pattern. And indeed, having reanalysed Semino's dataset, Chikhi et al (2002) came to the opposite conclusion and interpreted the clines as having been formed during the Neolithic. At that time, time the estimates for Y chromosomal haplogroups were much less informative and reliable than those for mtDNA. The reason for this is that it is hard to distinguish on the Y chromosome the pre-existing diversity (which founder population had brought from its homeland) and one accumulated in situ. (Founder analysis, which was the convenient instrument for mtDNA, proved to be too complex to be applied to the Y chromosome).

Since the 1990s, the studies on mitochondrial DNA and Y chromosome diversity became dominant in population genetics, which resulted in a specific "two-system" way of thinking. According to it, the greater part of migratory events was allegedly reflected in the both systems. Over the following years numerous studies were published on Y chromosomal and mtDNA variation in virtually all European countries. Most of them provided missing pieces for the European genetic puzzle but did refrain from making oversimplified and/or general conclusions. Those which did could be roughly classified into two groups: those describing the overall genetic landscape (based on the data of the totality of haplogroups) and deducing a particular genetic event from the distribution of particular haplogroup (the haplogroup-driving approach).

Mitochondrial landscape of Europe

From the perspective of mitochondrial DNA, the European gene pool consists of 7-10 most

ISSN: 1920-2997 http://rjgg.org

frequent haplogroups. All but one of them came from the Near East: in the majority of cases during the Early Upper Palaeolithic (EUP) in conjunction with the initial AMH dispersal and a smaller part during the Neolithic epoch (in the course of Neolithisation, Richards et al., 2000). The genetic landscape has been reshaped in the Mesolithic/Late Palaeolithic times, during the repopulation of Europe from the southern European refugia. The only European haplogroup that presumably had emerged in Europe (haplogroup V) became spread across the entire continent during Mesolithic/Late Palaeolithic recolonisation (Torroni et al., 2001). The western European origin of this haplogroup (the Franco-Cantabrian refugium) is presumed, based on its high frequency in this area as well as the occurrence of its phylogenetic predecessor (pre-V lineages). However, a recent accumulation of genetic data on previously poorly studied Eastern Europe, enabled the present author (Balanovsky, 2008) to suppose the occurrence of an additional East European centre of origin of this haplogroup. This finding is based on even higher frequency and yet again, on the presence of pre-V lineages in East European steppe area. Impossibility to distinguish between the western and eastern European homelands emphasised the important feature of the European mitochondrial landscape – its extreme homogeneity.

Indeed, when additional data from different European populations became available the genetic similarity in haplogroup frequencies (and identity in haplogroup spectra) has been immediately recognised (Simoni et al., 2000). As a result, mtDNA studies has appeared dealing with Europe as a whole, comparing it with the Near East or other areas, while attempts to trace genetic processes within Europe encountered problems (Helgason et al., 2000). This development was rather discouraging for archaeologists and linguists who were typically interested in a genetic support of existing hypothesis in their respective disciplines, although on a much smaller scale. Fortunately, the paper entitled "In search of geographic patterns in European mitochondrial DNA" (Richards et al., 2002) made the point that with the emergence of a larger dataset (with more than 3,000 individual mtDNAs) a spatial structuring became more apparent (e.g. the south-north difference was acknowledged among macro-regions of Europe: Mediterranean area, Central Europe, Scandinavia, and, surprisingly, the Basque Country).

Presently one can affirm that size of the dataset is the key factor. Having at our disposal the database six times larger than previously possessed, comprising 20,000 European mtDNAs (Balanovska, Zaporozhchenko, Pshenichnov, Balanovsky; MURKA Mitochondrial Database and Integrated Software, unpublished) we were able to recognise a much clearer patterning (Fig. 4). European populations altogether occupying all parts of this plot provide a geometrical illustration of the genetic variation in Europe.

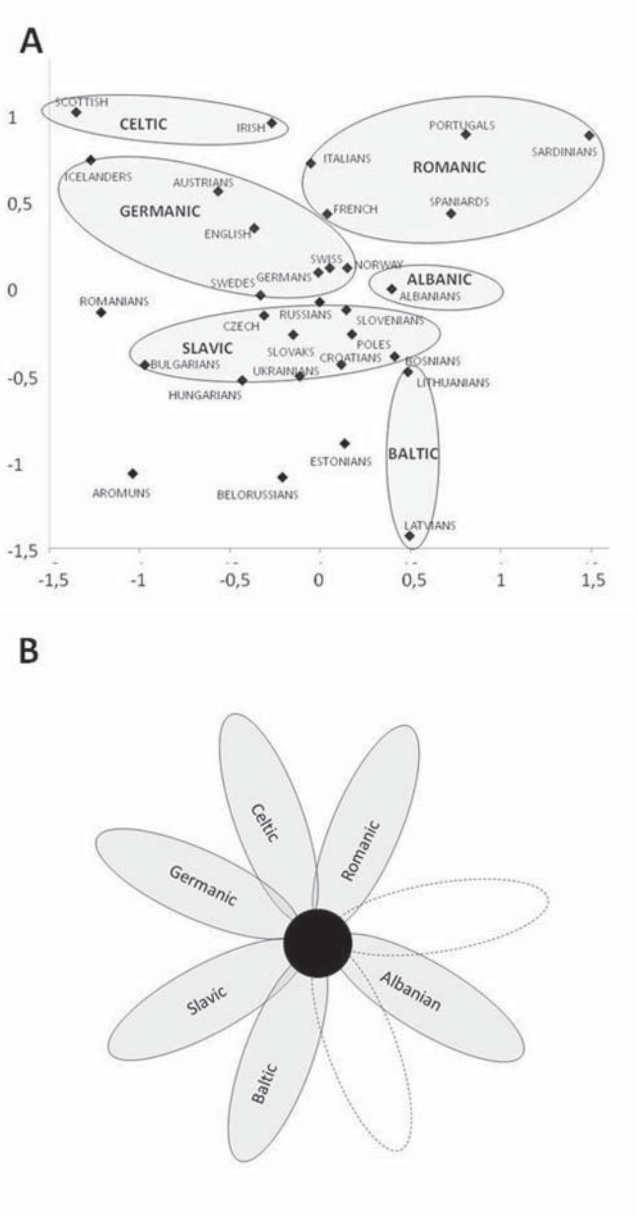

Figure 4. Genetic relationships of European populations from mitochondrial DNA perspective.
A. The multidimensional scaling plot (geometric distances between points display the genetic distances between corresponding populations). Ellipses mark populations belonging to the same linguistic group.
B. The idealized approximation of the plot (A) by flower-like structure. Black core – proto-Indo-European population; dotted ellipses – hypothetical extinct linguistic groups.

ISSN: 1920-2997 http://rjgg.org

The most remarkable feature is that the pattern distinctly resembles linguistic groups: populations cluster together according to their linguistic group of the Indo-European family. Three largest clusters are formed by Romanic, Slavic and Germanic speakers, while Baltic and Celtic speakers form smaller clusters and Albanians form a "cluster" of its own outside any other cluster. There are a few exceptions: Romanians, Aromuns and Sicilians lie outside the Romanic cluster while Estonians join the Slavic cluster. In both cases the geographical distance (remoteness for Romanians and Aromuns, proximity for Estonians) had probably a stronger impact on genetics than the linguistic affiliation.

Among the non-Indo-European populations of Europe the Basques found their place outside any other cluster but close to the Romance one (not surprisingly, considering the geographic proximity again). Finno-Ugric and Turkic speakers are not shown on this plot because of their extreme genetic variation, but on another plot they lie apart of Indo-Europeans.

This linguistic structuring of European mitochondrial DNA follows a remarkable "flower shape" pattern: all clusters looking like petals around the "core" of the flower. The possible explanation of this pattern is that genetic and linguistic differentiations were parallel processes or, in better words, – two aspects of the same process, related to the multiplication and differentiation of proto-Indo-European population in Europe. It is well known, that in many particular cases distribution of genes is opposed to distribution of language (especially in cases of the language replacement by the elite dominance model). However, in very general view, almost all Europe is populated by speakers of one linguistic family and almost all Europe is genetically homogenous. This allows speculations (like our flower-like interpretation of the genetic plot) which consider genetic and linguistic evolution as generally parallel processes, disregarding partial exceptions. Such speculations inevitably oversimplify both processes but could serve as a starting point for more detailed studies.

Therefore, one can accept as a working hypothesis the differentiation of proto-Indo-European language into linguistic groups being accompanied by genetic differentiation resulting in a clear clustering pattern (Fig. 4). This allows one to introduce time frames into the formation of the European mitochondrial landscape. It would coincide with origin and differentiation of European branches of IE family, i.e. covers the last 5-6 millennia (Starostin et al., their linguistic database is avalable at http://starling.rinet.ru/main.html). This does not necessarily imply the Neolithisation (for example, major changes during the Bronze Age is one of alternative explanations), but lends credence to the hypotheses advocating a relatively recent origin (or at least late major reshaping) of the mitochondrial pool in Europe.

One may note that the significance of the linguistic factor is quite obvious on the graph (Fig. 4 A). However, the idea of a single proto-population totally depends on the flower-like structure of this graph (Fig. 4 B) and should be therefore considered as one of the plausible hypotheses.

Y chromosomal landscape of the Europe

While the "homogeneity" is the principal feature of mitochondrial pool, the Y chromosomal pool is characterized by a high heterogeneity. As with mtDNA, there are seven Y chromosomal haplogroups dominating in Europe. But while frequencies of mitochondrial haplogroups are quite similar across Europe, Y chromosomal haplogroups follow a clear geographical pattern (Fig. 5). Neither classical markers, nor mitochondrial haplogroups demonstrated such obvious and elegant trends. Therefore, Y chromosome became an effective instrument in population genetics.

One should remember that European gene pool cannot be homogeneous and heterogeneous at the same time. The question is to what degree different markers are able to reveal the existing degree of variations. Having dozens of autosomal markers, classical population geneticists achieved reasonable resolution in assessing the variation between populations (Cavalli-Sforza et al., 1994). Mitochondrial DNA failed to reveal a difference between populations and successfully operates only at a higher hierarchical level: separating regions (Richards et al., 2002) and linguistic groups (present study, fig. 4). Y chromosome operates much better and separates even subpopulations within the same ethnic group (Balanovsky et al., 2008). Recent studies based on half of million autosomal markers became able to separate even individuals within subpopulations (Novembre et al., 2008).

This high differentiation power of Y chromosome (i.e. clear geographical clines of its haplogroups) was revealed already in the early large-scale studies (Semino et al., 2000; Rosser et al., 2000). These clines have been recently summarized in a panel of frequency distribution maps (Balanovsky et al., 2008). Two main haplogroups, accounting altogether almost for a

ISSN: 1920-2997 http://rjgg.org

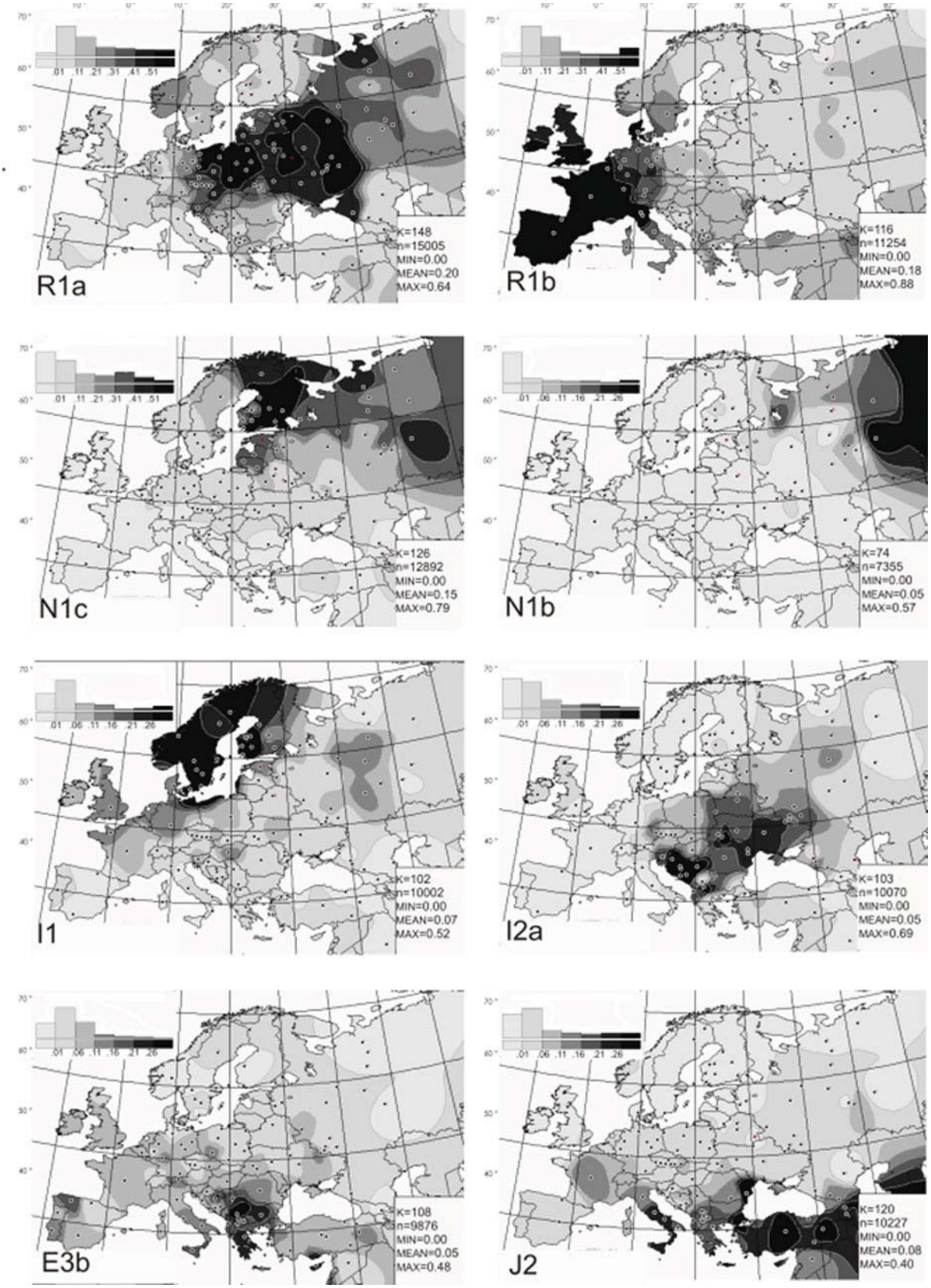

Figure 5. Geographic distribution of European Y chromosomal haplogroups (modified from Balanovsky et al., 2008).
K – number of studied populations; n – number of studied individuals; MIN, MEAN, and MAX- minimal, mean and average frequency on the map.

ISSN: 1920-2997 http://rjgg.org

half of the total European Y chromosomal pool are distributed along the west-to-east axis. Haplogroup R1b accounts for roughly 50% of the Y chromosomal pool in Western Europe and decreases eastward, while R1a reaches the same high frequency in the east (Fig. 5) and decreases westward.

Analysis of another type of Y chromosomal markers (microsatellite variation) also proved the western and eastern domains to be main features of the Y chromosomal pool (Roewer et al., 2005). As it was stressed above, the "interpretation by association" should be made with caution. That is why attributing these domains to Late Palaeolithic re-colonisation from two principal refugia (the south-western and south-eastern ones) can be considered as a possible but not yet proven hypothesis. (One of other possibly hypotheses is attributing these genetic domains to descendants of Late Neolithic Bell Beaker and Corded Ware cultures).

These two principal European haplogroups R1b and R1a are shared between Europe and other regions (Central Asia, Near East, India and North Africa). But two other haplogroups, I1 and I2a (according to previously used nomenclature the same haplogroups were labelled as I1a and I1b, respectively) are restricted to Europe, where they had likely originated.

While R1b and R1a occupy the west and the east, I1 and I2a predominate in the Europe's north and south, respectively. I1 which is frequent in Scandinavia and southern Baltic area has attracted less attention due to obviously late colonization of this region. In contrast, the distribution of haplogroup I2a (Balkan haplogroup) has been widely debated. As southeast European autochthonous haplogroup it could not be attributed to Neolithic immigrants (or any other immigrants) into Europe. We will discuss it in more details below.

Three remaining haplogroups (E, J, and N1c) are not evenly spread across the entire Europe but are restricted to distinct areas. For this and other reasons they are believed to mark later migration waves into Europe which did not cover the entire continent.

The haplogroup N1c (N3 or TAT, according to previous nomenclatures) is restricted to north-east Europe (mainly Finnic speakers) and Siberia. During the last decade it remained unclear whether is marks an eastward migration from Europe or the opposite westward migration trend. In 2007 Rootsi and colleagues have shown that this haplogroups could be deeply rooted in East Asian phylogeny and therefore the occurrence of this haplogroup in Europe may be attributed to the Asian influence. Authors supposed step-by-step migration from North China to Eastern Europe, which started in early Holocene and underwent a secondary expansion on its long way. Derenko and colleagues (2007) studied microsatellite variation associated with this haplogroup in more detail and tried to estimate its age. They identified two variants, one of which migrated into Europe 6-10 ky ago, while the second (less frequent) variant was shown to come by the way of a smaller and more recent migration, 2-4 ky ago. Although these time estimations should be taken with great caution, the both studies (Rootsi t al., 2007; Derenko et al., 2007) agree that north-east Europe had a significant (or even predominant) genetic legacy in South Siberian/Central Asian populations.

This creates a problem for "two systems" approach, because from mitochondrial perspective Siberian/East Asian haplogroups appeared in low frequencies and only in the eastern edge of Europe and did not account for a significant portion of the gene pool anywhere else in Europe. (When low frequency of typical East Asian haplogroup F was found on Croatian isles and in even lower frequencies on Croatian mainland, this was considered as a paradox and a special paper (Tolk et al., 2001) tried to explain it by possible medieval gene flow caused by trade routes of Venice). That is why a significant Asian presence in Europe, concluded from haplogroup N1c remains one of the main inconsistencies between Y chromosomal and mitochondrial genetic systems. From our point of view, this problem could be resolved if one takes into consideration the fact that genetic boundary between Europe and Asia lies much eastern than Ural Mountains. The western Central Asia (the Altai Mountains in particular) could be therefore considered as a genetically intermediary in present time and primary "European" zone in the past. This view explains why Y chromosomal haplogroup N (despite its origin in East Asia 20-30 ky ago) in pre Neolithic or Neolithic times could be the characteristic haplogroup for Caucasoid populations in Eurasian steppe west from the Altai and also for Mongoloid populations east from the Altai. From the western part of this area the haplogroup N1c could spread northward and northwestward by a number of migrations suggested for this area. This view also explains why these migrations did not bring East Eurasian mitochondrial haplogroups into Europe: the source area having mainly Western Eurasian haplogroups even in the contemporary gene pool. It was more the case in earlier times before Turkic speakers brought East Eurasian haplogroups by

ISSN: 1920-2997 http://rjgg.org

their expansion into this area two millennia ago onward.

Two last Y chromosomal haplogroups to be discussed are E and J. They predominate in North Africa and Near East, and in Europe they are found mainly in Mediterranean area. Not all sub-branches of these two haplogroups reached Europe, but mainly one branch of haplogroup E (namely, E-V13) and two branches of haplogroup J (J-M241 and J-410).

While J-410 follows the separate pattern, the other two haplogroups (and also haplogroup I2a, mentioned above) are concentrated in the Balkans and have not been found in neighbouring regions with any significant frequencies. The ages of these haplogroups estimated from their STR diversity are: E-V13 from 4 to7,5 ky; J-M241 from 3,5 to 6 ky; I2a from 5,5 and 10 ky (Battaglia et al., 2008). This roughly coincides with time of Neolithic transition in this part of Europe. The model suggested by Battaglia and colleagues states that Neolithic cultural package was adopted by local Mesolithic populations of the Balkans which started grow in numbers, expanding farming across the entire Balkan peninsula and later transmitting this package to other Mesolithic populations of Europe. This model explains why these three haplogroups are restricted to the Balkans, why they exhibit decreasing frequency towards other part of Europe and why their age is similar to that of the Neolithic transition.

However, the internal logic of this model is opposite to those applied by Richards and colleagues in relation to mitochondrial DNA. Indeed, Richards and colleagues proved that mitochondrial haplogroups whose diversity was accumulated in Europe in situ are of Palaeolithic age; and from this fact they concluded that present-day Europeans are descendants of Palaeolithic population of Europe (Richards et al., 2000). Eight years later, Battaglia and colleagues proved that Y chromosomal haplogroups whose diversity was also accumulated in Europe in situ are of Neolithic age; but from this contrasting fact they concluded also that present day Europeans are descendants of Palaeolithic population of Europe (Battaglia et al., 2008). Both studies are reasonably substantiated and their conclusions look correct. However this example illustrates that genetic studies need more robust and universal logic, at least when dealing with such complex process like the Neolithisation. In this particular case the possible logical compromise lays in the fact that concept of Battaglia and coauthors actually implies both, cultural diffusion and demic diffusion models. Although the authors did not formulate this explicitly, their concept implies, that cultural diffusion took place between regions (Analolia and Balkans, Balkans and Central Mediterrania) while the demic diffusion occurred within regions.

Ancient DNA

Analysis of ancient DNA (aDNA) provides direct data on the former European gene pool, which are free from assumptions and speculations which often accompany deductions of past genetic processes, based on the contemporary genetic pattern. This advantage of aDNA may trigger a revolution in population genetics and if this did not happen so far, this was due to limited quality and quantity of available aDNA evidence.

Problems with the quality (authenticity) of aDNA data are dramatic because of the possible contamination by modern DNA. For this reason some aDNA results may be false, and many early aDNA papers were criticized exactly from this point of view. This problem could be partially resolved in few high-standard laboratories only, which have special equipment to minimize risk of contamination. Cross-checking, i.e. independent analysis of the same ancient sample in different aDNA labs is the second condition. The third one is implementing the "modern DNA free" style of excavation into the practice of the archaeological fieldwork. Having these three conditions met, one can reach reasonably degree of authenticity of the aDNA results.

The quantity problem consists in the scarcity and limited sample sizes of the aDNA data. Again, this problem could be solved only partially, by increasing the number of aDNA studies and average sample size per study. Fortunately, both factors tended to increase in the last decade, still trace amounts and high fragmentation of ancient DNA samples hinder its high-throughput analysis.

Because of these limitations aDNA at least presently cannot be the main source of genetic knowledge about the Neolithisation. But it is already one of the important sources on this problem. Indeed, analyses of Neandertal mitochondrial DNA (Krings et al., 1997; Ovchinnikov et al., 2000), though being criticized for probable mistakes in sequencing, put an end to a lengthy discussion of the possible assimilation of the Neandertal populations by anatomically modern humans. Specificity of Neandertal mitochondrial type (Currat, Excoffier, 2004) and absence of this type in present day Europeans (Behar et al., 2007) allow to root European gene

ISSN: 1920-2997 http://rjgg.org

pool in AMH colonization of the Europe, disregarding the previous epochs.

Direct genetic data on first Neolithic groups in Europe are of course most promising source for choosing between the demic and cultural diffusion models of Neolithisation. Such data are now available for Neolithic population of the Iberian peninsula (Sampietro et al., 2007) and Neolithic population of the Central Europe (sites of Linear Band Ceramic with age of 7.0 – 7.5 ky; Haak et al., 2005). Iberian Neolithic population was shown to be genetically similar to the present day Iberian population. In contrast, LBK population in Central Europe was shown to be genetically distinct from the present day population of that or any other region of Europe). The most remarkable feature of the Neolithic population was mitochondrial haplogroup N1a found in 6 of 24 individuals. This haplogroup is virtually absent in present-day Europe. The mathematical simulation has shown that if this Neolithic population was source of present-day Europeans they could not have lost this haplogroups by stochastic genetic drift.

It was therefore concluded (Haak at el, 2005) that Neolithic LBK population did not become parental for present-day European gene pool, but became dissolved in pre-existing European populations. This conclusion is therefore in agreement with cultural diffusion model in assuming that since Neolithic farmers arrived in Europe, the farming was adopted by aboriginal populations and first farmers did not leave any considerable genetic legacy in their new homeland.

The study by Haak and colleagues did answer the question: "what happened with first farmers after their arrival in Europe". To address the another question, "where these first farmers came from", the consequent study was performed (Haak, pers. comm.). Based on extended dataset (44 individual mtDNAs from different sites of early LBK culture) it was found that this population is genetically similar to present day populations of Northern Mesopotamia, southern Caucasus and eastern Anatolia. Although the genetic composition of this area could be disturbed after the Neolithic period by subsequent migrations, it is reasonable to suppose that inner areas of the Near East were homeland for migrating groups who finally brought these mitochondrial lineages into the LBK population of the Central Europe.

Of course, this data give rise to many new questions, and currently available aDNA data are not sufficient to address them. The moderate optimism is based on increasing number and quality of aDNA data which might allow better chronological and geographical resolution of genetic processes in the near future.

Conclusions

The increasingly accumulating genetic data on extant and extinct (aDNA) European populations are most frequently discussed in terms of two opposite concepts: demic diffusion and cultural diffusion models of Neolithisation. In hands of Cavalli-Sforza and his colleagues the genetic mirror reflected Neolithic expansion across Europe (demic diffusion); but in hands of present-day writers this mirror reflects mainly the Palaeolithic legacy of Europeans and cultural diffusion model is needed to explain spread of farming.

Understanding the genetic history of Europe implies clarifying relative significance and patterns of each of the following processes:

1. the initial dispersal of AMH in Europe (Upper Palaeolithic);
2. the restructuring of the genetic landscape during the Mesolithic repopulation of the Europe from two-four refugia;
3. the importance of the Neolithic expansion viewed as the spread of early farming communities or spread of Neolithic cultural package;
4. the role of post-Neolithic human movements within Europe;
5. the "oriental" influence in different epochs – from Palaeolithic to Medieval times.

To address these questions population genetics operated with autosomal (classical) markers in the past and autosomal (DNA) markers may became the new standard in the future, while the present day studies are based on mitochondrial DNA and Y chromosomal variation.

Analysis of mtDNA demonstrated that most of European haplogroups came from the Near East during the Upper Palaeolithic times and Neolithic migration of Near Eastern farmers did not contribute much into the European gene pool. The south-east – northwest cline within Europe, as established by many genetic markers, is not considered anymore as the trace of Neolithic expansion, because Palaeolithic colonists used likely the same geographical route.

Y chromosomal data reveal distinct domains of prehistoric movements within Europe. Particularly, two different haplogroups predominate in Western versus Eastern Europe, and one may speculate about two secondary homelands, associating them with Mesolithic refugia or centres of later expansions.

ISSN: 1920-2997 http://rjgg.org

Southeast Europe (the Balkans) which deserves special attention as gates into Europe, is populated by three different Y chromosomal haplogroups exhibiting similar patterns: being autochthonous for Europe these haplogroups started to expand in time frames comparable with the Neolithisation; it was supposed that this expansion might took the form of Balkan's Mesolithic population adopting farming from their Anatolian neighbours.

Analysis of ancient DNA indicated that first Central European farmers (LBK) were of Near Eastern origin but did not left recognisable descendants. The early farmers in Iberia (and possible in other areas of late Neolithisation) were of aboriginal European genetic type.

Genetic mirror shows a controversial picture: even in this summary "indigenous" Balkan populations adopted farming without immigrant farmers, but "immigrant" gene pool was found in first farmers even north of Balkans (in Central Europe). Nevertheless most lines of reasoning show that Neolithisation did not change drastically the European gene pool and consequently did not involve large-scale population movements. Since, if one would like to obtain further information about these (relatively minor) movements from genetic data it is necessary to be equipped with a large genetic databases and a good dose of scepticism not to rush to conclusions.

References

Ammerman A.J. and L.L Cavalli-Sforza (1984) Neolithic Transition and the Genetics of Populations in Europe, Princeton. N.J.: Princeton University Press.

Balanovskaya E.V. and S.D Nurbaev (1997a) "A computer-aided methods for gene-geographic study of a population gene pool in space of principal components", Genetika (Moscow) Vol 33: 1456-1469.

Balanovskaya E.V. and S.D Nurbaev (1997b) "The selective structure of gene pool: possibilities of studying", Genetika (Moscow) Vol. 33: 1572-1588.

Balanovsky O, S. Rootsi, A. Pshenichnov, T. Kivisild, M. Churnosov, I. Evseeva, E. Pocheshkhova, M. Boldyreva, N. Yankovsky, E. Balanovska and R. Villems R (2008) "Two sources of the Russian patrilineal heritage in their Eurasian context" American Journal of Human Genetics Vol. 82(1):236-50.

Balanovsky O. (2008) "Gene pool of the high latitudes in Eurasia or where Saami came from?", in Way to North: the enviroments and the earliest inhabitans of Arctic and Subarctic. Moscow, Institut of Geography. P.. 277-282 (in Russian).

Barbujani G, G. Bertorelle and L.Chikhi (1998) "Evidence for Paleolithic and Neolithic gene flow in Europe", American Journal of Human Genetics Vol 62(2):488-92.

Barbujani G. and G. Bertorelle (2001) "Genetics and the population history of Europe", Proceedings of the National Academy of Sciences USA Vol 98(1):22-5.

Battaglia V, S. Fornarino, N. Al-Zahery, A. Olivieri, M.Pala, M.N. Myres, R.J. King, S.Rootsi, D. Marjanovic, D.Primorac, R.Hadziselimovic, S.Vidovic, K.Drobnic, N.Durmishi, A.Torroni, A.S. Santachiara-Benerecetti, P.A. Underhill and O. Semino (2008) "Y-chromosomal evidence of the cultural diffusion of agriculture in southeast Europe", European Journal of Human Genetics Advance online publication 24 December 2008; doi: 10.1038/ejhg.2008.249

Behar D.M, S. Rosset, J. Blue-Smith, O. Balanovsky, S.Tzur, D. Comas, R.J. Mitchell, L. Quintana-Murci, C.Tyler-Smith, R.S. Wells and Genographic Consortium. (2007) "The Genographic Project public participation mitochondrial DNA database" PLoS Genetics Vol 3(6):e104.

Cavalli-Sforza L.L., P. Menozzi and A. Piazza A. (1994) History and Geography of Human Genes. Princeton: Princeton University Press. 1059 p.

Chikhi L, R.A. Nichols, G. Barbujani and M.A. Beaumont (2002) "Y genetic data support the Neolithic demic diffusion model", Proceedings of the National Academy of Sciences USA Vol 99(17):11008-13.

Childe V.G. 1928. The Dawn of European Civilization. New York, Knopf

Currat M, and L. Excoffier (2004) "Modern humans did not admix with Neanderthals during their range expansion into Europe" PLoS Biology Vol 2(12):e421.

Derenko M, B. Malyarchuk, G. Denisova, M. Wozniak, T. Grzybowski, I.Dambueva, and I. Zakharov (2007) "Y-chromosome haplogroup N dispersals from south Siberia to Europe", Journal of Human Genetics. Vol 52(9):763-70.

Haak W, P. Forster, B. Bramanti, S. Matsumura, G. Brandt, M. Tänzer, R. Villems, C. Renfrew, D. Gronenborn, Alt, K.W. and J. Burger (2005) "Ancient DNA from the first European farmers in 7500-year-old Neolithic sites", Science Vol 310(5750):1016-8.

Helgason A, S. Sigurethardóttir, J. Nicholson, B. Sykes, E.W. Hill, D.G. Bradley, V. Bosnes, J.R. Gulcher, R. Ward, and K. Stefánsson (2000) "Estimating Scandinavian and Gaelic ancestry in the male settlers of Iceland", American Journal of Human Genetics Vol 67(3):697-717.

Krings M, A. Stone, R.W. Schmitz, H. Krainitzki, M. Stoneking,and S. Pääbo (1997) "Neandertal DNA sequences and the origin of modern humans", Vol Cell 90(1):19-30.

Lewontin R.C., and J. Krakauer J. (1975) "Testing the heterogeneity of F values" , Genetics. Vol 80: 397-398.

Menozzi P.,A. Piazza A and L.L. Cavalli-Sforza L.L. (1978) "Synthetic maps of human gene frequencies in Europe", Science Vol 201: 786-792.

Novembre J, T. Johnson, K. Bryc, Z. Kutalik, A.R. Boyko, A. Auton, A. Indap, K.S. King, S. Bergmann, M.R. Nelson, M Stephens, and C.D. Bustamante (2008) "Genes mirror geography within Europe", Nature Vol 456(7218):98-101.

Ovchinnikov IV, Götherström A, Romanova GP, Kharitonov VM, Lidén K, Goodwin W. (2000) "Molecular analysis of Neanderthal DNA from the northern Caucasus", Nature 404(6777):490-3.

Richards M, H. Côrte-Real, P. Forster, V. Macaulay, H. Wilkinson-Herbots, A. Demaine, S. Papiha, R. Hedges, H.J. Bandelt, Sykes B. (1996) "Paleolithic and neolithic lineages in the European mitochondrial gene pool", American Journal of Human Genetics. 59(1):185-203.

Richards M, Macaulay V, Hickey E, Vega E, B. Sykes, V. Guida, C. Rengo, D. Sellitto, F Cruciani, T. Kivisild, R. Villems, M. Thomas, S. Rychkov, O. Rychkov, Y. Rychkov, M. Golge, D. Dimitrov, E. Hill, D. Bradley, V. Romano, F. Cali, G. Vona, A. Demaine, S. Papiha, C. Triantaphyllidis, G. Stefanescu, J. Hatina, M. Belledi, A. Di Rienzo, A. Novelletto, A. Oppenheim, S. Norby, N. Al-Zaheri, S. Santachiara-Benerecetti, R. Scozari, A. Torroni, and H.J. Bandelt (2000) "Tracing European founder lineages

ISSN: 1920-2997 http://rjgg.org

in the Near Eastern mtDNA pool", American Journal of Human Genetics Vol 67(5):1251-76.

Richards M, Macaulay V, A. Torroni, and H.J. Bandelt. (2002) "In search of geographical patterns in European mitochondrial DNA", American Journal of Human Genetics Vol 71(5):1168-74.

Roewer L, P.J. Croucher, S. Willuweit, T.T. Lu, M. Kayser, R. Lessig, P. de Knijff, M.A. Jobling, C. Tyler-Smith and M. Krawczak M. (2005) "Signature of recent historical events in the European Y-chromosomal STR haplotype distribution", Hum Genet. 116(4):279-91.

Rootsi S, L.A. Zhivotovsky, M. Baldovic, M. Kayser, I.A. Kutuev, R.Khusainova, M.A. Bermisheva, Gubina, M. S.A. Fedorova, A.M. Ilumäe, E.K. Khusnutdinova, M.I. Voevoda, L.P. Osipova, M. Stoneking, A.A. Lin, V. Ferak, J. Parik, T. Kivisild, P.A. Underhill, and R. Villems. (2007) "A counter-clockwise northern route of the Y-chromosome haplogroup N from Southeast Asia towards Europe", European Journal of Human Genetics Vol 15(2):204-11.

Rosser Z.H., T. M.E. Zerjal, M Hurles,. Adojaan, D. Alavantic, A. Amorim, W. Amos, M. Armenteros, E. Arroyo, G. Barbujani, L. Beckman,. J. Beckman, E Bertranpetit, D.G. Bosch, G. Bradley, H.B.Brede, Cooper, P. Côrte-Real, R. de Knijff,. Decorte, Y.E. Dubrova, O. Evgrafov, A. Gilissen, S. Glisic, M. E.W. Gölge, Hill, A Jeziorowska, L. Kalaydjieva, M. Kayser, T. Kivisild, S.A. K. Kravchenko, A. Krumina, V. Kucinskas, J. Lavinha, L.A. Livshits, P. Malaspina, S. Maria, McElreavey, T.A. Meitinger, A.V. Mikelsaar, R.J. Mitchell, K. Nafa, J. Nicholson, S. Nørby, A. Pandya, J. Parik, P.C.Patsalis, L. Pereira, B. Peterlin, G. Pielberg, M.J. Prata, C. Previderé, L. Roewer, S. Rootsi, D.C. Rubinsztein, J. Saillard, F.R. Santos, G. Stefanescu, B.C. Sykes, A. Tolun, R. Villems, C. Tyler-Smith, and M.A. Jobling. (2000) "Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language", American Journal of Human Genetics Vol 67(6):1526-43.

Rychkov Yu.G., and E.V Balanovskaya. (1992) "Genetic geography of USSR population", Genetika (Moscow) Vol . 28. (1) : 52-75.

Sampietro M.L., O. Lao, D. Caramelli., M. Lari., R. Pou, M. Martı, J. Bertranpetit., and C. Lalueza-Fox (2007) "Palaeogenetic evidence supports a dual model of Neolithic spreading into Europe", Vol Proc Biol Sci. Vol 274(1622):2161-2167.

Semino O, G. Passarino, P.J. Oefner, A.A. Lin, S. Arbuzova, L.E. Beckman, G. De Benedictis, P. Francalacci, A. Kouvatsi, S. Limborska, M. Marcikiae, A. Mika, B. Mika, D. Primorac, A.S. Santachiara-Benerecetti, L.L. Cavalli-Sforza, and P.A. Underhill (2000) "The genetic legacy of Paleolithic Homo sapiens sapiens in extant Europeans: a Y chromosome perspective", Science Vol 290(5494):1155-1159.

Simoni L, F. Calafell, D. Pettener, J. Bertranpetit, and G. Barbujani (2000) "Geographic patterns of mtDNA diversity in Europe", American Journal of Human Genetics Vol 66(1):262-278.

Sokal R.R., and N.L Oden. (1978) "Spatial autocorrelation in biology: I. Methodology", Biological Journal of Linnean Society Vol. 10. P. 199-228.

Sokal R.R., N.L Oden, and B.A. Thomson (1999) "A problem with synthetic maps", Human Biology Vol 71 (1): 1-13.

Tolk H.V., L. Barac, M. Pericic, I.M. Klaric, B. Janicijevic, H. Campbell, I. Rudan, T. Kivisild, R. Villems, and P. Rudan (2001) "The evidence of mtDNA haplogroup F in a European population and its ethnohistoric implications", European Journal of Human Genetics Vol 9(9):717-23.

Torroni A, H.J. Bandelt, V. Macaulay, M. Richards, .M Cruciani, C. Rengo, V. Martinez-Cabrera, R. Villems, T. Kivisild, E. Metspalu, J. Parik, H.V. Tolk, K. Tambets, P Forster, B. Karger, P. Francalacci, P. Rudan, B. Janicijevic, O. Rickards, M.L. Savontaus, K. Huoponen, V. Laitinen, S. Koivumäki, B. Sykes, E. Hickey, A. Novelletto, P. Moral, D. Sellitto, A. Coppa, N. Al-Zaheri, A.S. Santachiara-Benerecetti, O. Semino, and R. Scozzari (2001) "A signal, from human mtDNA, of postglacial recolonization in Europe", American Journal of Human Genetics Vol 69(4):844-52.

Yamazaki T. and T. Maryama (1973) "Evidence that enzyme polymorphisms are selectively neutral", Nature Vol 245: 140-141.