Chemistry: Bulgarian Journal of Science Education | *Volume 28 Number 5, 2019* | *Природните науки в образованието*

ON THE GENETIC TIES BETWEEN EUROPEAN NATIONS

Jordan Tabov, Nevena Sabeva-Koleva, Georgi Gachev
Institute of Mathematics and Informatics – Bulgarian Academy of Sciences (Bulgaria)

Abstract. The gathered data from genetic research permits analysis of geographic distribution of haplogroups of European nations. In this paper we conjecture that the predominant processes playing main part in modeling the contemporary quantitative distribution of these haplogroups in Europe are diffusion type processes.

Keywords: gene studies; origin of European nations

Introduction

Displacement of large groups of people on the planet Earth results in a "gene displacement", i.e. in "displacement of genetic information" through the children and grandchildren of the people. The close examination of those processes could lead to a better understanding of a "genetic ties" between different nations and their cultural, political and economic influence. There is a large database concerning genetic studies about European people in EUPEDIA. It gives rise to a number of hypotheses which can be modified and verified when the database is enlarged. We present here some useful steps in this direction.

Group and individual displacements of people

In the past the main perception for origin of peoples was "biblical" one, i.e. it was connected to Noah's sons.

During Christian era the perception for the Deluge and Noah's ark was universal for European nations. Thus, all people are assumed to have their origin from Noah's sons – Jaffet, Shem and Ham and from their grandchildren.

It is well established as a common understanding that all European nations are Yafet's descendants. The Yafet's sons are considered as progenitors of certain nations or groups of nations. For example, the progenitors of Slavic nations are connected to the sixth Jaffet's son – Meshech, but in some sources it is linked to the seventh son – Tiras.

The idea of Noah's ark assumes the same origin for all people. As a consequence, it comes the idea of migration – all people started their journey from one and the same place (usually this is the pick of Ararat) and after many resettlements they arrive to the place they occupy at the moment of research.

Although the displacement of entire nation of long distance is difficult to imagine, the displacement of organized groups of people (troops for example) is a well-established fact.

For example, the crusaders marched hundreds and thousands of kilometers. They also have been accompanied by their wives and children.

Most probably certain people moved in a short distance of about 30-40 kilometers. This occurs when an inhabitant of a village took a wife or husband from a nearby village.

Thus, a short transfer of genes has occurred. It is clear that for long periods similar transfers resulted in a long-range transfer of genes. This usually occurs in a several stages. There are other reasons for short displacements of individuals, for example, representatives of different handicrafts such as blacksmiths and potters from a village moved to another village to find better market for their goods. In general, the process of "gene transfer" as a result of short distance displacement of people can be characterized as a diffusion type process.

Along gathering genetic data for individuals an interest of finding genetic ties between large groups of people ("populations") had occurred; genetic ties between different nations and their common roots are investigated; attempts to track the movement (migration) of the people was made; a new notion of "genetic distance" between different nations was introduced.

From numerous publications in that direction let us mention[1-4] Nazarova (2009); Balanovsky (2012); Delev (2018); Tabov & Sabeva-Koleva (2018); Chiaroni et al. (2009); Karmin et al. (2015); Karlsson et al. (2006); Trombetta et al. (2011).

Haplogroups of European nations

According to EUPEDIA the European population features the following haplogroups:

I1 I2*/I2a I2b R1a R1b G J2 J*/J1 E1b1b T Q N

Among Bulgarians their distribution[1] in percentage are given by:

4 20 2 17 11 5 11 3 23.5 1.5 0.5 0.5

Thus, it can be said that "average Bulgarian" has "conditional halo group" composed by components of distinct haplogroups and these components have different weight. Most common haplogroups for Bulgarian population are: E1b1b, I2*/I2a, R1a, R1b and J2. For Macedonia the corresponding percentages are:

3 23 1,5 13,5 12,5 4 14 2 21,5 1,5 0,50,5,

and for Serbia:

8,5 33 0,5 16 8 2 8 0,5 18 1 1,5 2

These complex data show that the problem for "kinship" among nations is more complicated than the similar problem for individuals.

Genetic distance

Tabov & Sabeva-Koleva (2018) propose a method for compering “genetic closeness“ between the nations: for any nation they consider the percentages in the above sequence of haplogroups as coordinates of a point in 12-dimensional Euclidean space. The distance between the points is defined as *genetic distance.*

This definition is European specific: for populations in which the sum of these 12 Y-chromosome haplogroups is less than 95%, we obtain greater deviations compared to the values of most Y- chromosome haplogroups, therefore when we apply this definition for genetic distance, we ignore them. Thus,

$$\text{Distance}(\text{Country1}, \text{Country2}) = \sqrt{\sum_{i=1}^{12} (x_{1,i} - x_{2,i})^2}$$

Using the genetic distance, we can visualize the genetic distances between different nations through the lengths of the segments connecting the corresponding points. For example, for the triples of countries (Bulgaria, Macedonia and Serbia) and (Bulgaria, Macedonia and Greece) we obtain the diagrams in **Fig. 1**.

Two-dimensional visualization of genetic distance for four countries is not always appropriate. The reason is that the ratio between the distances of the points, corresponding to the percentages of Y-chromosomes of different nations, when projected in a plane usually changes significantly. This leads to speculations when such results are interpreted and discussed.

In the case of Bulgaria, Macedonia, Greece and Serbia though the plane projection is a good approximation and is given in **Fig. 2**.

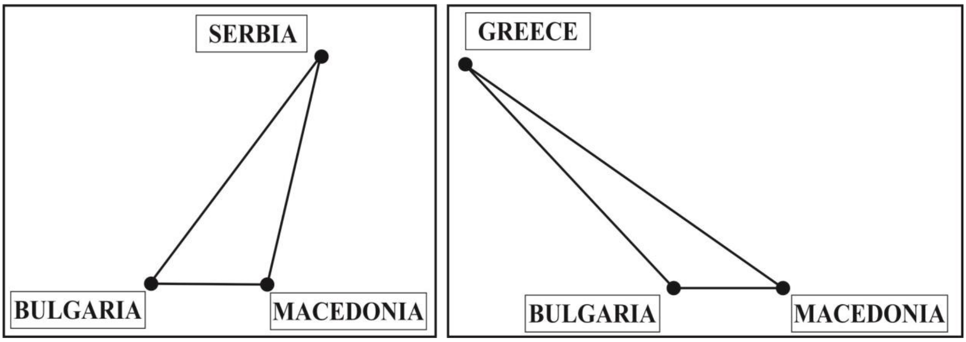

Figure 1. Visualization genetic closeness of the triples of countries (Bulgaria, Macedonia, Serbia) and (Bulgaria, Macedonia, Greece).

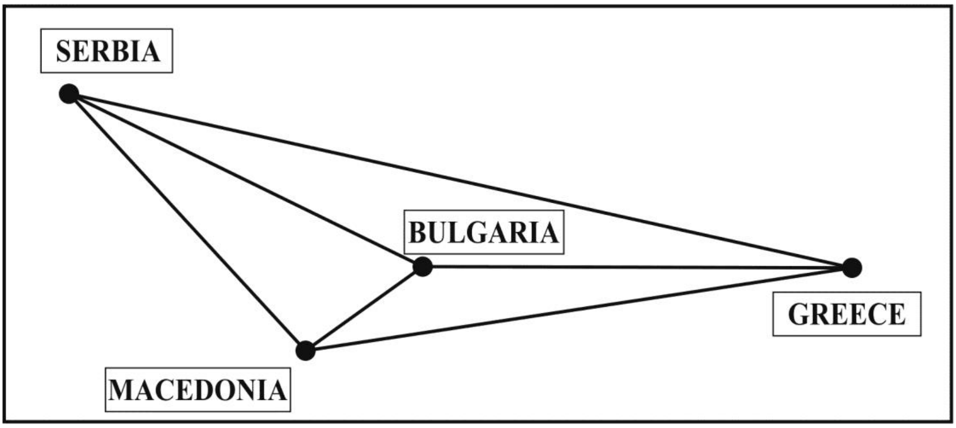

Figure 2. Visualization of genetic similarity of (Bulgaria, Macedonia, Serbia and Greece)

In our talk (Tabov & Sabeva-Koleva, 2018), given in 2018 at a conference, we offered the following hypothesis:

Hypothesis: The change of ratio of haplogroups in Europe is a diffusion type process: it can be conjectured that the main part of this process has been played by "exchange of haplogroups" between neighboring nations. Most probably large scale migration has played less influence in this process. The results we are going to present below provide further arguments and clarification of this hypothesis.

Genetic-geographical correlation

The first argument is connected to the correlation when "genetic distance" is compared to "geographical distance". If such a correlation exists this leads to the conclusion for significant diffusion type exchange of genes. Indeed, for all 38 European nations (we exclude small ones by territory and population) the correlation has coefficient of 0.53; and without Iceland, that has specific location, the coefficient becomes 0.61.

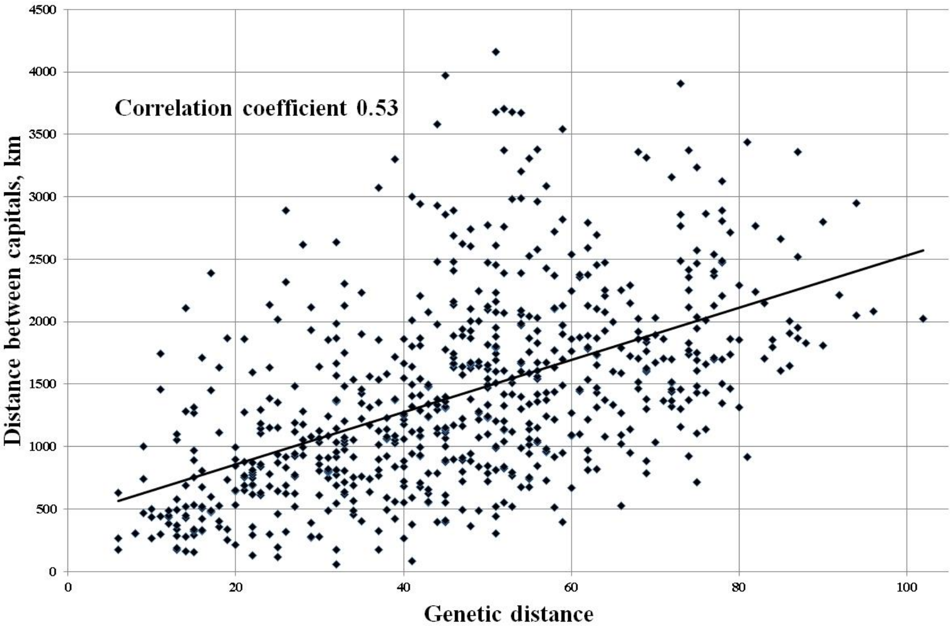

Figure 3. Correlation of "genetic distance" and "geographical distance" for all 38 European nations (we exclude small ones by territory and population) the correlation has coefficient of 0.53

These values of the coefficient of correlation between "genetic distances" and "geographic distances" of European nations/countries (Figs. 3 and 4) makes the diffusion type distribution of genes across Europe main tool that shaped the contemporary "genetic picture" of male population of our continent.

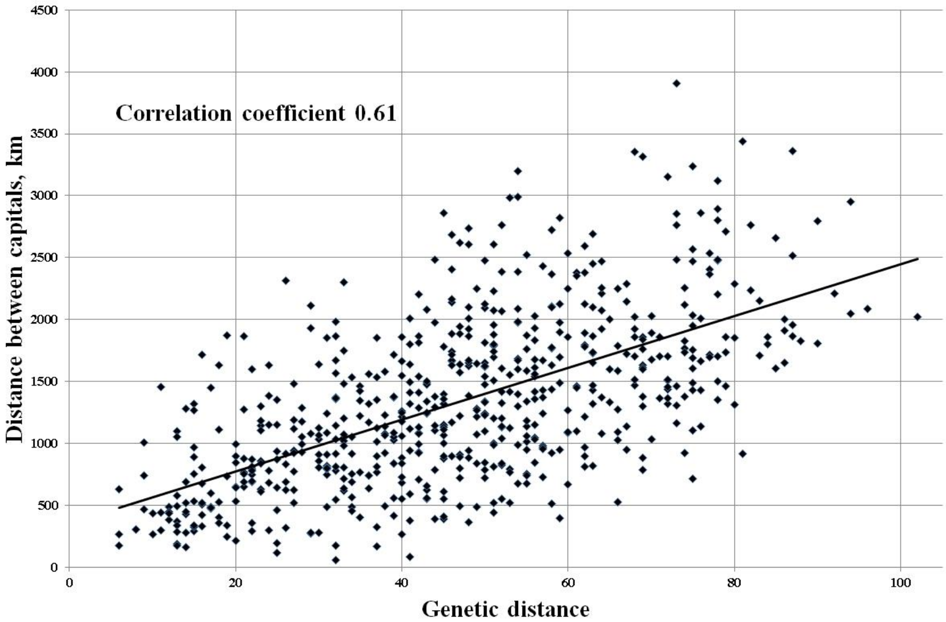

Figure 4. Correlation between "genetic distance" and "geographical distance" for 37 European nations/countries (we exclude small ones by territory and population such as Vatican, Monaco, etc. Iceland is also excluded since it has specific location); the correlation coefficient equals 0.61

Distribution of all 12 European haplogroups in Europe

The second argument supporting the diffusion type distribution of male genes in Europe is the qualitative distribution (in percentages) of each of the 12 European haplogroups. We consider three of them: I2*/I2a, E1b1b and I1; the "presence" (in percentages) of these haplogroups in Bulgarian population are 21%, 23,5% and 4%, respectively. The first are "Balkan" – the highest percentage for both of them is achieved in Balkan countries.

Fig. 5 is indicative: the percentages of haplogroup I2*/I2a are higher in Bosnia and Herzegovina; they are significant in neighboring countries – Serbia, Croatia, Montenegro, Rumania, Slovenia, Bulgaria and Moldova; they are to be found in most countries, neighbors of the above – Hungary, Slovakia, Ukraine, Belarus, Greece. This is in line with the logic of outside diffusion penetration to the Balkan Peninsula. A small deviation from the above conclusions is the significant percentage of I2*/I2a in Russia; one possible explanation for this could be the migration of Bulgarians in Russia during and after Russia-Turkish wars during XIX century.

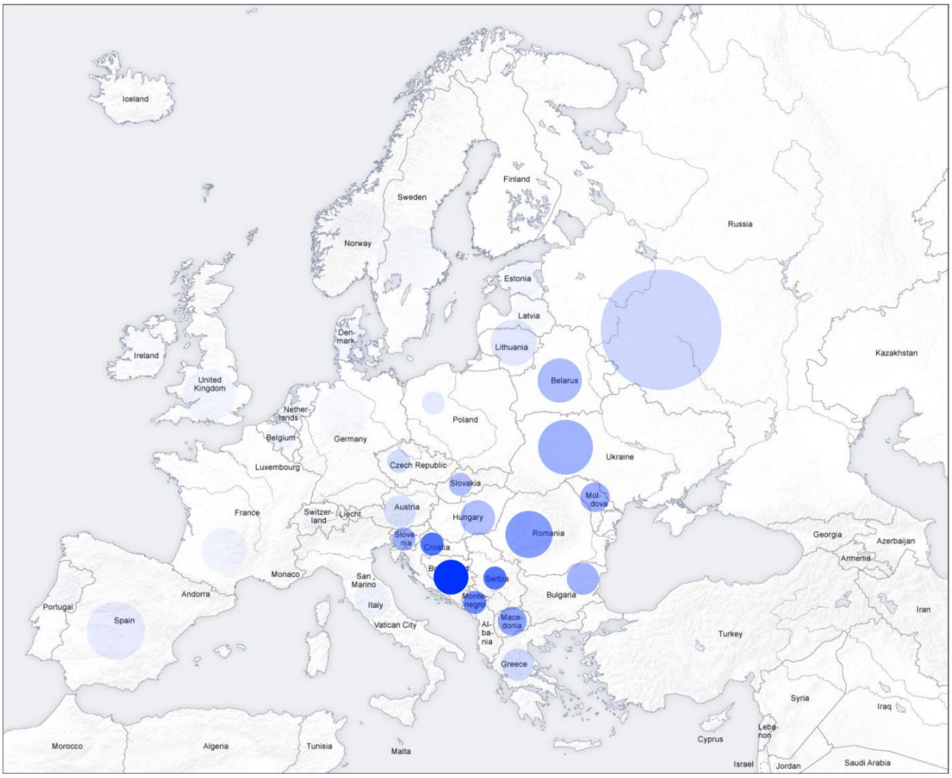

Figure 5. Haplogroup I2*/I2a: quantitative distribution in Europe (the countries having bigger percentage of I2*/I2a are marked with darker disks)

Fig. 6 explains the transition from “geographic picture” of the percentage distribution of I2*/I2a in European countries to formal-mathematical definition of correlation (in percentages) in European countries and the distance to the capital of the given state to Sarajevo (the capital of Bosnia and Herzegovina – the country with highest percentage I2*/I2a). The coefficient of 0.67 is significant and this confirms the conclusions from **Fig. 5.**

Fig. 7 is an analog of Fig. 5 for haplogroup E1b1b. The reasoning and the conclusions are similar. The percentages of E1b1b are higher in Albania, Montenegro, Bulgaria and Macedonia; they are significant in neighboring countries – Serbia, Bosnia and Herzegovina, Greece, Romania, Moldova, Italy. This corresponds to the diffusion penetration from the countries of Balkan Peninsula outward. There is a small deviation here: the significant percentage of E1b1b in Portugal.

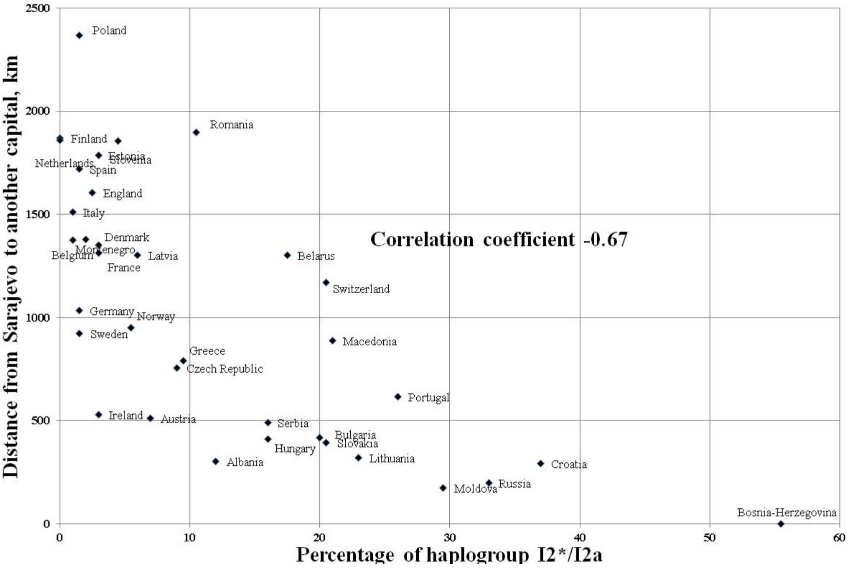

Figure 6. Haplogroup I2*/I2a: correlation of quantities (in percentages) in Europe an states and the distance from the capital of the corresponding state to Sarajevo (the capital of Bosnia and Herzegovina, where I2*/I2a has higher percentage)

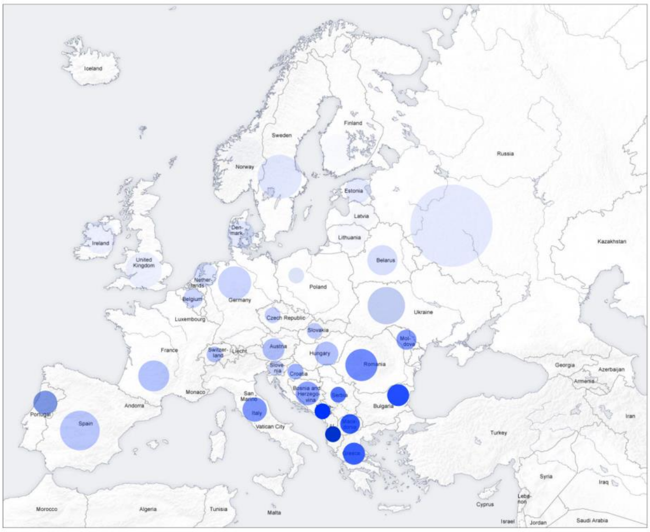

Figure 7. Haplogroup E1b1b: quantitative distribution in Europe (darker discs represent countries with higher percentage of E1b1b)

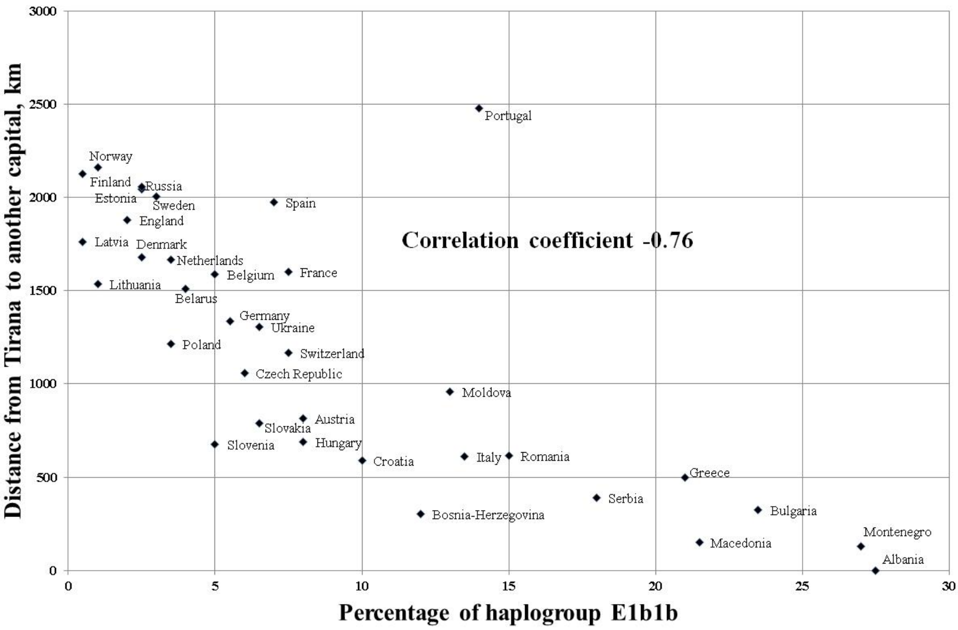

Figure 8. Haplogroup E1b1b: correlation of quantities (in percentages) in European countries and the distance from the capital of corresponding country to Tirana (the capital of Albania – the country with highest E1b1b percentage)

Fig. 8 illustrates the formal mathematical definition of correlation of quantities (in pecentages) of E1b1b in European countries and the distance from the capital of corresponding country to Tirana (the capital of Albania – the country with highest E1b1b percentage). The coefficient of 0.76 is high enough and confirms the conclusions of Fig. 7: the main factor leading to the current state in Europe is the diffusion type distribution of E1b1b.

For various reasons the "geographic picture" of the percentage distribution of haplogroup I1 in Europe is interesting and important. **Fig. 9** is convincing – from the country with the highest percentage (Sweden) towards other countries this percentage decreases. **Fig. 10** illustrates by means of formal mathematics the correlation of the quantities (in percentages) in European countries and the distance from the corresponding country to Stockholm (the capital of Sweden). The coefficient of the correlation 0.71 is significant and supports the conclusions from **Fig. 9**.

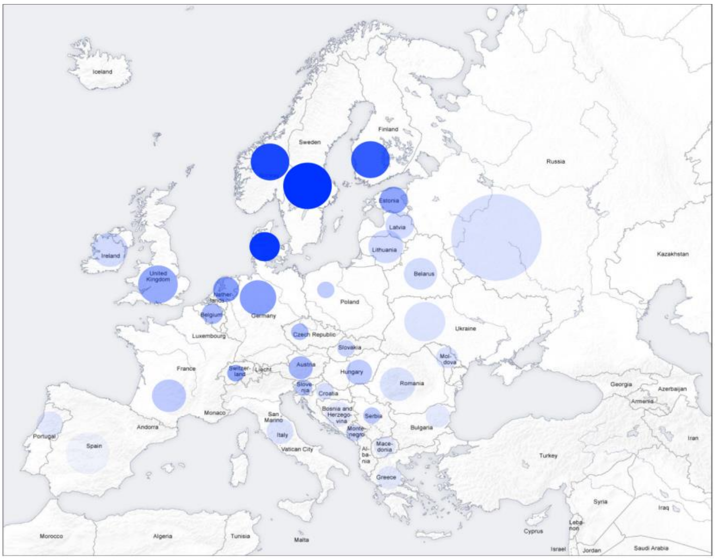

Figure 9. Haplogroup I1: quantitative distribution in Europe (darker discs represent countries with higher percentage of I1)

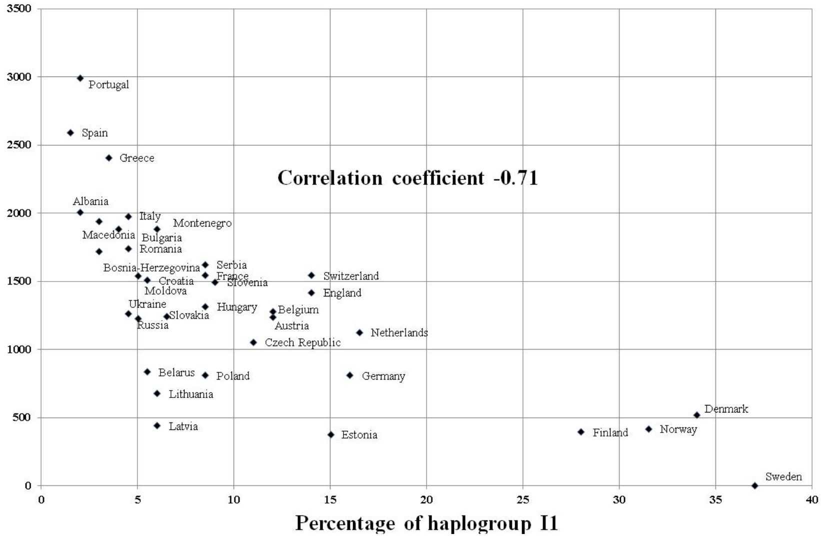

Figure 10. Haplogroup I1: correlation of the quantities (in percentages) in European countries and the distance to Stockholm (the capital of Sweden – the country with higher percentage of **I1**)

Conclusions and hypothesis

The above analysis of the quantitative distribution (in percentages) of the three European haplogroups I2*/I2a, E1b1b and I1 on the territory of Europe support the hypothesis of diffusion type "exchange of genes" between European nations. They support the formulation of more general hypothesis for the "genetic history" of European population:

***Hypothesis**:* Long time ago (probably few thousand years) 15 – 30 regions in Europe has been "nests" for one or 2 – 3 haplogroups, i.e. the population of Europe in a particular region is referred to one or 2 – 3 haplogroups. Along the communication, traveling and migration a diffusion type exchange of genes between neighboring regions has occurred, thus the genes pervasion has intensified.

With this process the percentage of people having "local haplogroups" decreased at the expense of "outside groups".

During the Middle Ages an opportunity for "batch" exchange of genes through arm forces (for example crusaders) occurred.

An interesting example in this direction is "Goth's" haplogroup I1 with hypothetical nest in today Southern Sweden.

According to EUPEDIA this haplogroup in Bulgaria is 4%; in other words it can be conjectured that Goth's "contribution" to the genesis of the Bulgarian nation is about 4%.

APPENDIX

Data table for the percentage of haplogroups for 39 European nations1)

Country	I1	I2*/I2a	I2b	R1a	R1b	G	J2	J*/J1	E1b1b	T	Q	N
Albania	2	12	1,5	9	16	1,5	19,5	2	27,5	1	0	0
Austria	12	7	2,5	19	32	7,5	9	1	8	1	0,5	0,5
Belarus	5,5	17,5	1	51	5,5	1,5	2,5	1	4	0	0	10
Belgium	12	3	4,5	4	61	4	4	1	5	1	0,5	0
Bosnia-Herzegovina	3	55,5	0	15	3,5	1,5	4	0,5	12	1	2	2
Bulgaria	4	20	2	17	11	5	11	3	23,5	1,5	0,5	0,5
Croatia	5,5	37	1	24	8,5	2,5	6	1	10	0,5	1	0,5
Czech Republic	11	9	4	34	22	5	6	0	6	1	1,5	0,5
Denmark	34	2	5,5	15	33	2,5	3	0	2,5	0	1	1
England	14	2,5	4,5	4,5	67	1,5	3,5	0	2	0,5	0,5	0
Estonia	15	3	0,5	32	8	0	1	0	2,5	3,5	0,5	34
Finland	28	0	0,5	5	3,5	0	0	0	0,5	0	0	61,5
France	8,5	3	3,5	3	58,5	5,5	6	1,5	7,5	1	0,5	0
Germany	16	1,5	4,5	16	44,5	5	4,5	0	5,5	1	0,5	1
Greece	3,5	9,5	1,5	11,5	15,5	6,5	23	3	21	4,5	0	0
Hungary	8,5	16	2	29,5	18,5	3,5	6,5	3	8	0	0	0,5
Iceland	29	0	4	23	42	0	0	0	0	0	1	1

Ireland	6	1	5	2,5	81	1	1	0	2	0	0	0
Italy	4,5	3	2,5	4	39	9	15,5	3	13,5	2,5	0	0
Latvia	6	1	1	40	12	0	0,5	0	0,5	0,5	0,5	38
Lithuania	6	6	1	38	5	0	0	0	1	0,5	0,5	42
Macedonia	3	23	1,5	13,5	12,5	4	14	2	21,5	1,5	0,5	0,5
Malta	1	10	1	3,5	32,5	6,5	21	8	9	4,5	1	0
Moldova	5	21	3	30,5	16	1	4	4	13	1	0	1,5
Montenegro	6	29,5	1,5	7,5	9,5	2,5	9	0,5	27	0	2	1,5
Netherlands	16,5	1	6,5	4	49	4,5	3,5	0,5	3,5	1	0	0
Norway	31,5	0	4,5	25,5	32	1	0,5	0	1	0	1	2,5
Poland	8,5	5,5	2	57,5	12,5	1,5	2,5	0	3,5	0,5	0,5	4
Portugal	2	1,5	3	1,5	56	6,5	9,5	3	14	2,5	0,5	0
Romania	4,5	26	2,5	17,5	12	5	13,5	1,5	15	0,5	0,5	0,5
Russia	5	10,5	0	46	6	1	3	0	2,5	1,5	1,5	23
Serbia	8,5	33	0,5	16	8	2	8	0,5	18	1	1,5	2
Slovakia	6,5	16	1,5	41,5	14,5	4	2	1	6,5	0,5	0,5	3
Slovenia	9	20,5	1,5	38	18	1,5	2,5	0	5	1	0	0
Spain	1,5	4,5	1	2	69	3	8	1,5	7	2,5	0	0
Sweden	37	1,5	3,5	16	21,5	1	2,5	0	3	0	2,5	7
Switzerland	14	1,5	8	3,5	50	7,5	3	0,5	7,5	0,5	1,5	1
Ukraine	4,5	20,5	0,5	44	8	3	4,5	0,5	6,5	1	0,5	5,5

NOTES

1. https://www.eupedia.com/europe/european_y-dna_haplogroups.shtml
2. https://en.wikipedia.org/wiki/List_of_Y chromosome_haplogroups_in_populations_of_the_world
3. https://web.archive.org/web/20120621052330/http://www.familytreedna.com/understanding-haplogroups.aspx
4. https://ru.wikipedia.org/w/index.php?title=Гаплогруппа_E1b1b (Y-ДНК)&oldid=97696115

REFERENCES

Balanovsky, O. (2012). *Variability of the gene pool in space and time: the synthesis of data on the mitochondrial DNA genogeography and Y-chromosome: DSc thesis*. Moscow: Moscow University [In Russian].

Chiaroni J., Underhill, P.A. & Cavalli-Sforza, L.L. (2009). Y chromosome diversity, human expansion, drift, and cultural evolution. *Proc. Nat. Acad. Sci. USA, 106*(48), 20174 – 20179.

Delev, E. (2017). *Population genetics of the Bulgarians, founders of the European civilization*. Sofia: Bomi [In Bulgarian].

Karlsson, A.O., Wallerström, T., Götherström, A. & Holmlund, G. (2006). Y-chromosome diversity in Sweden: a long-time perspective. *Eur. J. Human Genetics, 14*, 963 – 970.

Karmin, M., Saaq, L., Vicente, M., Wilson Savres, M.A., Järve, M., Talas, U.G., Rpptsi, S., Ilumäe, A.M., Mägi, R., Mitt, M., Pagani, L., Puurand. L., Faltyskova, Z., Clemente, F., Cardona, A., Metspalu, E., Sahakyan, H., Yunusbayev, B., Hudhasov, G., DeGiorgio, M., Looqväli, E.L. .... & Kivisild, T. (2015). A recent bottleneck of Y chromosome diversity coincides with a global change in culture. *Genome Res., 25*, 459 – 66.

Nazarova, F. (2009). The closeness of the Finnish Ugric, Slav and Germanic populations according to anthropological and genetic data. *J. Evol. Biology Res., 1*(1), 18 – 26.

Tabov, J. & Sabeva-Koleva, N. (2018). A method for assessing the genetic proximity of neighboring peoples. In: *Bulgarian historical readings: Bulgarian image.* Varna: Andari [In Bulgarian].

Trombetta, B., Cruciani, F., Sellitto, D. & Scozzari, R. (2011). New topology of the human Y chromosome haplogroup E1b1 (E-P2) revealed through the use of newly characterized binary polymorphisms. *PLoS ONE, 6*, art. no. e16073.

✉ **Prof. Jordan Tabov**
Dr. Nevena Sabeva-Koleva
Dr. Georgi Gachev
Institute of Mathematics and Informatics
Bulgarian Academy of Sciences
Acad. G. Bonchev St., Block 8
1113 Sofia, Bulgaria
E-mail: tabov@math.bas.bg
nevenasybeva@gmail.com
gachev@gmail.com