Biomath Communications

https://biomath.math.bas.bg/biomath/index.php/conference

Statistical Analysis of Genetic Data for the Balkans plus Historical Evidence

Jordan Tabov, Georgi Gachev

Institute of Mathematics and Informatics,
Bulgarian Academy of Sciences, Bulgaria
jordan@tabov.com
gachev@gmail.com

Abstract

Initial statistical analysis of genetic data on the Balkan nations showed the extent of their genetic links to each other. A careful review of the data highlighted an interesting feature: relatively large genetic differences between Greek regions. This is clearly expressed in the cases of Northern Greece, whose population is genetically much closer to the population of Bulgaria than to the population of Central and Southern Greece, whose population is genetically closer to the Albanian population than to the population of Northern Greece. These conclusions are based on numerical values from the results of statistical analysis of genetic data from EUPEDIA. The article also presents several historical testimonies that offer an explanation of the established regional genetic features of the population of today's Greece.

Keywords: Haplogroups, Multidimensional Model, Euclidean Distances, Genetic Distances, Genetic Proximity, Cluster, Balkan Countries

MSC2020: 92D25

1 Genetic distance

According to EUPEDIA the 12 basic haplogroups specific to the European population and their distribution in percentage among the Bulgarians are [1]:

I1	I2*/I2a	I2b	R1a	R1b	G	J2	J*/J1	E1b1b	T	Q	N
4	20	2	17	11	5	11	3	23.5	1.5	0.5	0.5

 Received: July 20, 2022, Accepted: December 30, 2022, Published: January 4, 2023
Citation: Jordan Tabov, Georgi Gachev, Statistical Analysis of Genetic Data for the Balkans plus Historical Evidence, Biomath Communications 9 (2022), 2212309, https://doi.org/10.55630/bmc.2022.12.309

Thus, it may be said that "the average Bulgarian" has "compound haplogroup" which consists of weighted partial inclusion of all basic haplogroups. Most common haplogroups for Bulgarian population are: E1b1b, I2*/I2a, R1a, R1b and J2.

Earlier papers [2, 3] propose a method for comparing "genetic closeness" between the nations. The main concept is that each nation is represented by a point in multidimensional Euclidean space. Each dimension corresponds to the content of a single haplogroup in the population. The exact position of a point is a 12-tuple of the percentage content of every of the 12 haplogroups among members of a single nation.

This metric space makes it possible to introduce a "genetic distance" between any two nations: we will define it as the geometric distance between any two nation points. The population of Europe is characterized by the previously mentioned 12 haplogroups. Then the "genetic distance" in Europe will be metric:

$$Distance(CountryA, CountryB) = \sqrt{\sum_{i=1}^{12}(CountryA_i - CountryB_i)^2} \quad (1)$$

This definition is European specific: for population 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.

The genetic distances between European nations, calculated by formula (1), further will be called "e-genetic" (i.e. European genetic).

This "e-genetic distance" makes it possible to study the genetic links between European nations. Research in this area has been done for a long time, among those publications that are relatively directly related to our analyses and results presented below, we will mention [2–10].

2 Cluster analysis of e-genetic distances between the European nations

Significant correlation is found when comparing "e-genetic" distance between each two European nations with the corresponding geographical distance [3]. This leads to the natural hypothesis that the distribution of the haplogroups on the territory of Europe is mainly a result of diffusion distribution and not of "large-scale migrations". It is in line with the announced by Angel Galabov's observations for the Bulgarians [4, 5].

Further analysis of the "e-genetic" distances between the regional population groups of Europe makes it possible to associate these groups in "genetically close" clusters. One such cluster is the "Balkan", formed by the peoples of the

	Albania	Bosnia-Herzegovina	Bulgaria	Croatia	Greece	Macedonia	Montenegro	Romania	Serbia
Balkan genetic centre	21	31	9	18	22	6	11	6	9
Albania		51	18	38	13	16	23	23	30
Bosnia-Herzegovina			39	22	53	37	32	33	25
Bulgaria				24	18	6	15	11	16
Croatia					38	23	25	16	13
Greece						17	27	22	31
Macedonia							13	9	15
Montenegro								17	14
Romania									12

Table 1: b-genetic distances between 9 Balkan countries: Albania, Bosnia-Herzegovina, Bulgaria, Croatia, Greece, Macedonia, Montenegro, Romania and Serbia [11], and their b-genetic distances to the "Balkan genetic centre".

following 8 Balkan countries: Bosnia-Herzegovina, Bulgaria, Croatia, Greece, Macedonia, Montenegro, Romania and Serbia. Frequently 2 more countries are considered belonging to the "Balkan" region: Slovenia and Albania. From the perspective of the "haplogroup comparison" however, Slovenia is closer to another group of countries. Its e-genetic distances to Slovakia and Ukraine are 8 and 14, while the Bulgaria and Greece are much farther on 31 and 40 correspondingly (see also [6]). Since about 8% of Albanian haplogroups do not belong to the set of 12 "European" haplogroups, the e-distance from Albania to above mentioned eight Balkan countries is inaccurate. Intending to acquire a better comparison, the following approach is proposed in [11].

To have better approximation and improved view of Albania the "13th coordinate" is introduced. These are missed percentage values to sum of 100%. Thus for each country is added 13th "haplogroup" of unknown type just to fill the gap. Also this extends the formula (1) with another, 13th coordinate, the distance modified this way will be called b-genetic (i.e. Balkan genetic) distance. Thus, the countries in the Balkan cluster become 9: Albania, Bosnia-Herzegovina, Bulgaria, Croatia, Greece, Macedonia, Montenegro, Romania and Serbia. The table of the b-genetic distances between countries in the cluster is presented in Table 1. A detailed study of the Western part of this cluster is published in the paper [6].

	Albania	Bosnia-Herzegovina	Bulgaria	Croatia	Northern Greece	Central Greece	Southern Greece	Macedonia	Montenegro	Romania	Serbia
Balkan genetic centre	18	30	6	21	10	21	21	3	12	7	14
Albania		47	18	38	16	12	9	16	23	23	32
Bosnia-Herzegovina			34	16	39	51	50	32	29	28	18
Bulgaria				24	8	18	19	6	15	11	18
Croatia					27	41	40	23	25	16	10
Northern Greece						16	16	10	21	12	22
Central Greece							11	19	26	26	35
Southern Greece								19	27	25	35
Macedonia									13	9	17
Montenegro										17	17
Romania											12

Table 2: b-genetic distances between 11 Balkan countries and regions of countries: Albania, Bosnia-Herzegovina, Bulgaria, Croatia, Greece (divided in 3 parts: Northern Greece, Central Greece and Southern Greece), Macedonia, Montenegro, Romania and Serbia and their b-genetic distances to the "Balkan genetic centre".

Table 1 clearly shows the proximity of the nations of the Western Balkans with those of the Eastern Balkans and the remoteness of all Balkan nations from the "Northern Slavs", which in turn belong to another cluster, this is also announced by A. Galabov [5].

3 Greece divided in three parts

The initial statistical analysis of genetic data on the Balkan nations shows the extent of their genetic links to each other. A careful review of the data highlights an interesting feature: the genetic specificity of the human populations in 3 individual Greek regions.

From Table 1 it can be seen that, in general, genetically closest to the residents of a country from the considered "Balkan cluster" are the populations of the neighboring countries. For example, the b-genetic distances from Serbia to the "neighbors" Romania, Croatia, Montenegro, Macedonia and Bulgaria are smaller than the b-genetic distances from Serbia to the other countries in the cluster.

The b-genetic distances from Northern Greece, Central Greece and Southern Greece to the rest of the Balkan nations show a clear deviation from this rule. As can be seen from Table 2, Northern Greece is at a b-genetic distance of 16 from Central (and Southern) Greece, and at a b-genetic distance of 8 from Bulgaria and of 12 from Romania.

The b-genetic proximity of Southern Greece to Albania is also surprising: the b-genetic distance between Southern and Central Greece is 11, which is 2 more than the b-genetic distance between Southern Greece and Albania.

All these "anomalies", or deviations from the "rule", are of natural interest and must be carefully studied. Steps in this direction are given below.

The comparison of the data in the tables of Table 1 and Table 2 provide an opportunity to see and assess the impact that the "fragmentation" of Greece has on the "location" of the "Balkan Genetic Centre" with respect to the individual countries and regions covered by the cluster. Briefly, Albania, Bosnia-Herzegovina, Bulgaria and Macedonia have moved closer to it (with 3, 1, 3, 3 respectively), while Croatia, Montenegro, Romania and Serbia have moved away from it (with 3, 1, 3, 3, respectively).

4 Commentary on the b-genetic proximity of Bulgaria and Northern Greece

The territory of Northern Greece includes today's Southern Macedonia, Aegean Thrace, Epirus and Thessaly.

At the end of 10^{th} century and during the 11^{th}, the name "Bulgaria" was widely used for the region discussed at the 8^{th} Eucomenical Council: Epirus, Thessaly, and Dardania. This can be seen on a number of historical maps, e.g. on the map in Figure 1. Here the territory of the eastern part of the Balkans is presented as Byzantine, in the southern part of the territory "Bulgaria" is the region of Epirus, which after the fall of Samuel's kingdom under Byzantine rule has become part of the administrative unit "Theme of Nicopolis", Figure 2.

For the Bulgarians as a predominant ethnic group in the Theme of Nicopolis there is information from a text from Skylitzes about the uprising of Petar Delyan [14, p. 227]. According to him, after the conquest of the Bulgarian kingdom, the Byzantine emperor Basil II left the taxes for the Bulgarians as they were under Samuel, because when the tax collector changed them, the local population of the Theme of Nicopolis (with the exception of Nafpaktos) could not come to terms with this and revolted.

From this story it is clear that 1) the local population on the Theme of Nicopolis has been Bulgarian, and 2) that this Theme has been part of the kingdom of Samuel. K. Jireček interprets this episode as follows: the inhabitants of the province near Nikopolis joined the Bulgarians, "their fellow tribesmen" [14, p. 227]. K. Jireček also presents information from Malatera, who has lived

Figure 1: Fragment of the map "South-Eastern Europe c. 1000" [12].

Figure 2: Fragment of the map "The Byzantine Empire 1045 AD" [13].

in the 11th century, that the Epirotes have been Bulgarians, and that the town of Arta (in the southern part of Epirus) has been a Bulgarian town [14, p. 124].

5 Commentary on Albania's b-genetic proximity to Central and Southern Greece: Albanians in Attica and Athens in the 1830s

The genetic proximity of the populations of Albania and Central (and Southern) Greece seems unexpected, but there is also information about it in the scientific literature. Valuable information about this was left by Jakob Falmerayer, professor of philology and history, and since 1835 a member of the Bavarian Academy of Sciences. He traveled extensively in the Orient for research purposes. He knew many languages and therefore had the opportunity to get a

des Mittelalters bildet. Nur ſo viel muß hier ſchon geſagt werden, daß von der Schneide des ſlaviſchen Zagora (Helikon) bis an die Spitze Attika's, und vom Iſthmus bis auf die Nordſeite des böotiſchen Sees Topolja (Kopais) nicht etwa nur die altgriechiſche, ſondern auch die ſlaviſch-neugriechiſche Bevölkerung des Mittelalters bis auf die letzte Spur verſchwunden iſt. Jedermann, der Attika und Bö-

Figure 3: Falmerayer's text about the population of Attica [15, p. 49].

In der großen Frage, welche durch meine Lehre zuerſt in Anregung gebracht wurde, iſt demnach über die heutigen Bewohner von Athen und Attika unwiderruflich entſchieden. Als Nicht-Griechen liegen ſie ſchon jenſeits der Gränzen aller weitern Prüfung, Zweifel und Erörterung. Und überhaupt hat etwa nicht allein Attika, ſondern der ganze große Landſtrich, welcher, um mit der Sprache unſerer Zeit zu reden, zwiſchen dem Hexamilon, Kap Colonnäs, Berg Zagora, See Topolja und dem Euripus eingeſchloſſen iſt, mit der althelleniſchen Welt nichts mehr gemein. Alles dieſes, einſt Herz und Mittelpunkt der Hellenen, iſt heute Neu-Albanien.*) Athen wurde wieder barbariſch, die Fackel der Wiſſenſchaften, durch die italieniſche Herrſchaft gepflegt, erloſch durch die albaneſiſchen Coloniſten völlig.

Figure 4: Falmerayer's text on Attica – that this is "New Albania" [15, p. 51].

good idea of the population and the history of the places visited. Among his most significant scientific works are: On the Origin of Today's Greeks [15] and History of the Morea in 2 volumes [16, 17].

According to Falmerayer, from the Slavic Zagora (modern-day Elikon) all the way to the coast of Attica and from the Isthmus to the northern part of Lake Boeotia, Topolia (modern-day Kopaes), "the last pale reflection not only of ancient Greek Hellenism but even of Slavic-Hellenic from the Middle Ages disappears" [15, p. 49] – Figure 3.

In his time – in the first half of the 19th century – in Attica and Boeotia "the culture, customs and mother tongue of the population were Albanian" [15, p. 34]. "Today there is an Albanian Athens," Falmerayer writes, "as thousands of years ago, before the Greek people were formed from a mosaic of tribes, there was an Athens of the Pelasgians." "Not only Attica, but the whole area between today's Examilia, Colonas, Mount Zagora, Lake Topolia and Evripos has nothing to do

with the ancient Greek world. All these lands, where the heart of the Greeks once has beaten, today represent New Albania." [15, p. 51] – Figure 4.

The observations of Fallmerayer are confirmed by other authors, see [18].

6 Conclusion

The examples of the unexpectedly large genetic proximity of the populations in the couple (Bulgaria, Northern Greece) and of the populations in the couple (Albania, Central Greece) show how the accumulation of data from genetic research and its mathematical processing increase our ability to look critically at a number of old prejudices and to seek and find a more accurate interpretation of the contradictory information from historical sources.

References

[1] EUPEDIA, "Distribution of European Y-chromosome DNA (Y-DNA) haplogroups by country in percentage", [03/01/2023].

[2] J. Tabov, N. Sabeva-Koleva, "A method for assessing the genetic proximity of neighboring peoples", *Bulgarian Historical Readings: Bulgarian Image*, 2018.

[3] J. Tabov, N. Sabeva-Koleva, G. Gachev, "On the Genetic Ties between European Nations", *Chemistry: Bulgarian Journal of Science Education*, 28:615-627, 2019.

[4] A. Galabov, "Genetically, Bulgarians are most similar to the population of Northern Greece and Italy", Dnes newspaper, Yantrapress, October 8, 2013, [03/01/2023].

[5] A. Galabov, "Bulgarians are not Slavs by genes", Presa newspaper, February 2, 2015, [03/01/2023].

[6] L. Kovacevic, K. Tambets, A.-M. Ilumäe, A. Kushniarevich, B. Yunusbayev, A. Solnik, et al., "Standing at the Gateway to Europe – The Genetic Structure of Western Balkan Populations Based on Autosomal and Haploid Markers", *PLoS ONE*, 9(8):e105090, 2014.

[7] B. Trombetta, F. Cruciani, D. Sellitto, R. Scozzari, "A New Topology of the Human Y Chromosome Haplogroup E1b1 (E-P2) Revealed through the Use of Newly Characterized Binary Polymorphisms", *PLoS ONE*, 6(1):e16073, 2011.

[8] S. Karachanak, V. Grugni, S. Fornarino, D. Nesheva, N. Al-Zahery, V. Battaglia, et al., "Y-Chromosome Diversity in Modern Bulgarians: New Clues about Their Ancestry", *PLoS ONE*, 8(3):e56779, 2013.

[9] Family Tree DNA, "Understanding Haplogroups: How are the haplogroups named?", Archived from the original on June 21, 2012, [03/01/2023].

[10] E. Delev, *Population genetics of the Bulgarians, founders of European civilization*, BOMI, 2017.

[11] G. Gachev, J. Tabov, "Digital Analysis of Genetic Proximity of the Balkan Population", *International Journal "Information Theories & Applications"*, 27(3):256-272, 2020.

[12] J. B. Bury, "South-Eastern Europe c. 1000", The Atlas To Freeman's Historical Geography, Third Edition, 1903, [03/01/2023]. https://commons.wikimedia.org/wiki/File:South-eastern_Europe_c._1000.jpg

[13] Stelian Brezeanu, *O istorie a Imperiului Bizantin*, Bucharest, 1981; Andrei Nacu, "The Byzantine Empire 1045 AD", December 10, 2017, [03/01/2023]. https://en.wikipedia.org/wiki/File:Map_Byzantine_Empire_1045.svg

[14] Konstantin Jireček, *History of the Bulgarians*, 1876; Iztok-Zapad, 2015.

[15] Jakob Philipp Fallmerayer, *Welchen Einfluss hatte die Besetzung Griechenlands durch die Slaven auf das Schicksal der Stadt Athen und der Landschaft Attika?, oder, Nähere Begründung der im ersten Bande der "Geschichte von Morea während des Mittelalters" aufgestellten Lehre über die Entstehung der heutigen Griechen*, J. G. Gotta'schen Buchhandlung, Stuttgart und Tübingen, 1835. https://books.google.com/books?id=StVEAAAAcAAJ

[16] Jakob Philipp Fallmerayer, *Geschichte der Halbinsel Morea während des Mittelalters: ein historischer Versuch. Teil 1: Untergang der peloponnesischen Hellenen und Wiederbevölkerung des leeren Bodens durch slavische Volksstämme*, J. G. Gotta'schen Buchhandlung, Stuttgart und Tübingen, 1830. https://books.google.com/books?id=iDdCAAAAcAAJ

[17] Jakob Philipp Fallmerayer, *Geschichte der Halbinsel Morea während des Mittelalters: ein historischer Versuch. Teil 2: Morea, durch innere Kriege zwischen Franken und Byzantiner verwüstet und von albanesischen Colonisten überschwemmt, wird endlich von den Türken erobert, von 1250-1500 nach Christus*, J. G. Gotta'schen Buchhandlung, Stuttgart und Tübingen, 1836. https://books.google.com/books?id=-vpNAAAAcAAJ

[18] Wikipedia, "Arvanites", [03/01/2023].