
Patrilineal and Matrilineal Genetic Diversity in Azerbaijan

Akper Alinazar Aliev¹, Anar Shamistan Ibrahimov¹, Vaqif Bilas Shadlinski²

¹Gene by Gene Ltd. (Azerbaijan DNA Project at FTDNA), Houston, the United States

²Department Human Anatomy, Azerbaijan Medical University, Baku, Republic of Azerbaijan

Email address:

anar.ibrahimov@khazar.org (A. S. Ibrahimov), medkrim@list.ru (V. B. Shadlinski)

To cite this article:

Akper Alinazar Aliev, Anar Shamistan Ibrahimov, Vaqif Bilas Shadlinski. Patrilineal and Matrilineal Genetic Diversity in Azerbaijan.

Biomedical Statistics and Informatics. Vol. 6, No. 3, 2021, pp. 59-64. doi: 10.11648/j.bsi.20210603.14

Received: June 24, 2021; **Accepted:** July 9, 2021; **Published:** August 23, 2021

Abstract: This work analyzed a large volume of literature on the genetic diversity of Azerbaijanis. Data from these sources were systematized and grouped, and a general description of trends in the distribution of Y-DNA and mt-DNA haplogroups was given. For the first time, the data of Azerbaijan DNA Project at FTDNA were used and on their basis the general statistical characteristics of the diversity of patrilineal and matrilineal gene pools were described. All information received within the Azerbaijan DNA Project at FTDNA was combined and brought together with data from various scientific sources. We made an attempt to summarize data on the distribution of patrilineal haplogroups in Azerbaijanis from some of the most informative sources with data from FTDNA. The subclades in question were designated according to the Yfull classification. The results of the work were analyzed using the M. D. Kendall coefficient of concordance. Kendall's coefficient of consistency was $W=0.83$, which indicates a high degree of consistency between the data of different authors at a significance level of $\alpha=0.01$. Taking these results into account, we have compiled a diagram of the diversity of Y-DNA haplogroups of the generalized sample of Azerbaijanis, which is currently the most presentable for characterizing patrilineal haplogroups.

Keywords: DNA Genealogy, Haplogroups, Genography of Azerbaijanis

1. Introduction

The genetic diversity of the population of the Caucasus in general and Azerbaijan in particular has recently been actively studied and discussed in various scientific circles [1-5]. Quite extensive factual material has been accumulated that requires analysis and interpretation. However, close acquaintance with these works made it possible to establish that many works duplicate each other. For example, the team of authors F. Cruciani et al. [6] in their work devoted to the phylogenographic analysis of haplogroup E3b, refer to materials from the work of a group of scientists R. Malaspina et al. [7], which contain information about the prevalence of some haplogroups among Azerbaijanis. This can be found in the article of Iranian scientists, who used the data of their compatriots published earlier as materials [8].

Some researchers interpret the results obtained in different ways. So, in [1], based on the results of testing 125 people, only the dominant haplogroups (J1 * -M267 and J2-M172) are indicated, and the rest of the variety (about a quarter of the entire material) is not specified. The main representative

group was Azerbaijanis from Derbent (Republic of Dagestan, Russia). More complicated is the situation in [8], where the sample consisted of 100 residents of Tabriz (East Azerbaijan province, Iran), where the dominant Y-haplogroups were J and R (~ 40% each). For these reasons, we decided to conduct a preliminary study of the diversity of haplogroups in the patrilineal and matrilineal parts of the gene pool of Azerbaijanis.

For this purpose, it was decided to analyze all available publications on the relevant topic, as well as to study open data on mt-DNA (mitochondrial DNA) and Y-DNA (Y-chromosome) of Azerbaijanis on the Family Tree DNA website [9]. It was planned to compare the results obtained with information on the anthropology of Azerbaijanis in general. It was planned to compare the results obtained with information on the anthropology of Azerbaijanis in general. These are mainly scientific works of the 20th century, which were obtained before the appearance of studies in molecular biology and which still continue to be the basis for forming the corresponding stereotypes of the anthropological appearance of Azerbaijanis.

2. Materials and Methods

The work analyzed the data of 12 authors on the diversity of haplogroups in the patrilineal part of the gene pool (based on SNP markers of the Y chromosome), as well as on the polymorphism of the HVR1 and HVR2mt-DNA segments, which were obtained during the study of biomaterials (saliva, blood, etc. etc.) taken from the Azerbaijanis [1-4, 6, 7, 10-13,

17]. The total number of men surveyed by these authors was 881. The number of subjects whose genetic samples were examined for the diversity of mitochondrial DNA were significantly fewer – only 327 people. Regarding mt-DNA polymorphism, it was possible to group data from only two Iranian sources [2, 10]. Information about the authors, the number of examined and the diversity of haplogroups is presented in tables 1 and 2.

Table 1. Distribution of Y-DNA haplogroups of Azerbaijanis according to various authors.

Author (-s) of the publication	Number of examined	Distribution of haplogroups (%)											
		E	F	G	I	J	K	L	T	N	Q	R	Others
R. S. Wells et al. [12]	21	4,8 (1)	9,5 (2)	4,8 (1)		48 (10)	9,5 (2)					24 (5)	
P. Malaspina et al. [10]	58					38							
M. E. Weale et al. [11]	29	7 (2)					10 (3)			3 (1)		21 (6)	59 (17)
T. Zerjal et al. [13]	19	5,25 (1)		5,25 (1)		58 (11)		5,25 (1)		5,25 (1)		21 (4)	
I. Nasidze et al. [3]	72	5,6 (4)	11 (8)	18 (13)	2,8 (2)	30,5 (22)						21 (15)	11 (8)
F. Cruciani et al. [6]	97	4 (4)											96 (93)
L. Andonian et al. [8]	100					40				10		40	10
S. Rootsi et al. [4]	297			13 (38)									
V. Grugni et al. [17]	63	11		8		27,1		3,2	7,9	1,6	4,8	36,4	
R. A. Shalyakho et al.[1]	125			11		36						30	23

* – The number of people tested (from those works where this information was indicated) is indicated in brackets.

Table 2. Distribution of mt-DNA haplogroups of Azerbaijanis according to Iranian authors.

A source	Number of examined	Distribution of haplogroups (%)														
		A	D	H	Hv	J	I	K	L	M	N	T	U	W	X	Others
M. Asgh-arzadehet al. [2]	133	0,75 (1)	0,75 (1)	25,57 (34)	9,77 (13)	11,28 (15)	4,5 (6)	6,02 (8)	0,75 (1)	0,75 (1)	2,26 (3)	11,28 (15)	20,3 (27)	2,26 (3)	0,75 (1)	3,01 (4)
A. Bahm-animehret al. [10]	294			22,7					2	4		16,5	20,1			34,7

* – The number of people tested (from those works where this information was indicated) is indicated in brackets.

In addition to this material, data from the Azerbaijani DNA project, which are freely available on the website of the American company FTDNA [9], were considered. This material includes information about the Y-haplogroups of 100 men from different regions of Azerbaijan, as well as the results of mitochondrial tests of 61 subjects (men and women). Subclades of Y-haplogroups were designated according to the classifications adopted for October 2017 in Yfull [14].

All the collected factual material was processed by simple methods of variation statistics [15]. In addition, the degree of consistency of the results of the work of various authors on the study of the genome of the Azerbaijani population was determined using the concordance coefficient of M. D. Kendall [16]. This method is used to reveal the consistency of the opinions of experts (in our case, independent authors) on several factors (haplogroups).

3. Results and Discussion

The study of the materials presented in the tables revealed some difficulties. In addition to problems with the nomenclature and the lack of data on the classifications used by individual authors, in many cases it was necessary to deal with incomplete identifications of haplogroups. For example, in early works such relict haplogroups of Y-DNA as F and K are often found (moreover, with an unrealistically large

percentage for the region). In connection with these circumstances, it was necessary to randomize the data used somewhat. Thus, some haplogroups were excluded from the study, and the carriers of these markers were assigned to the “others” group. In those cases when the authors in the studied sample were interested in only one (or two or three) haplogroups, and did not consider the distribution of others, all subjects without data on belonging to any haplogroup were also recorded in the “others” column.

After this kind of grouping, the relative percentages for individual haplogroups were calculated. Average percentages were calculated within large haplogroups, without specifying subclades or haplotypes. The sum of the relative percentages for haplogroups for Y-DNA was 108.08%, and for mt-DNA - 100.33%. Therefore, we scaled (from the total obtained) these data so that, as a result, all haplogroups for both Y-DNA and mt-DNA as a whole would be 100%. Of course, we were aware of the conventionality of such an approximation. According to statistical messages, in this case, it would be more efficient to search for a more realistic probability of assessing the distribution of any haplogroup (in%) by another method (for example, it was quite possible to use logistic regression). But in the case under consideration, this issue was not so fundamental - we just needed to approximately visualize the ratio of haplogroups of Azerbaijanis according to mosaic data of a number of authors. A difference between scaled values of less than 0.5% is unlikely to be significant in

such a situation.

Based on these messages, we compiled tables of data on the diversity of Y-DNA and mt-DNA haplogroups in Excel. For clarity, we presented the distribution of Y-DNA and mt-DNA haplogroups (according to our tables from Excel) in the form of pie charts (Figure 1). It should be noted that the picture obtained by us as a whole turned out to be comparable with the data published on various Internet resources in connection with the genealogy of Azerbaijanis. For the patrilineal part of the gene pool, it is enough to look at the share of such haplogroups as J, R and G, as if titular for

the population of Azerbaijan, in order to recognize this similarity.

However, in order to assess how adequately these diagrams reflect the genetic diversity of Azerbaijanis, we decided to compare them with similar results from the Azerbaijan DNA project. Based on the data indicated in the corresponding site, we calculated the percentage ratios of haplogroups according to data from the Azerbaijani DNA project. The results obtained are presented in the form of diagrams (Figure 2).

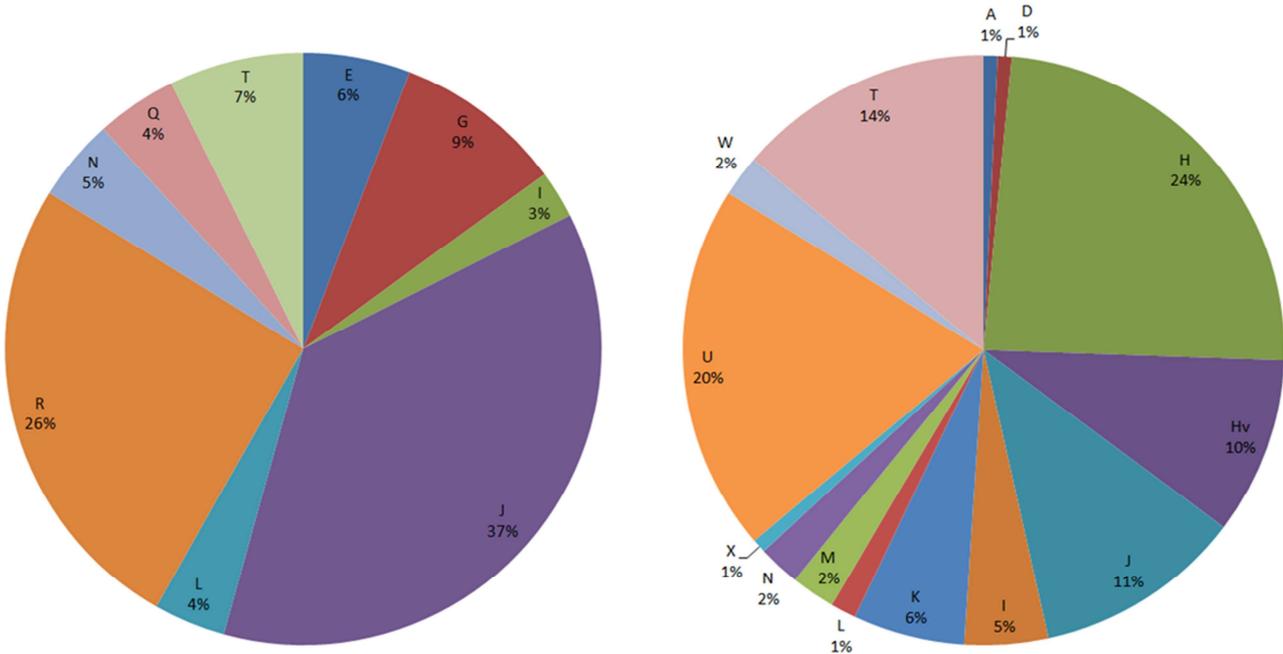


Figure 1. Distribution of Azerbaijanis by Y-DNA haplogroups (left) and by mt-DNA haplogroups (right), based on the summary data of various authors.

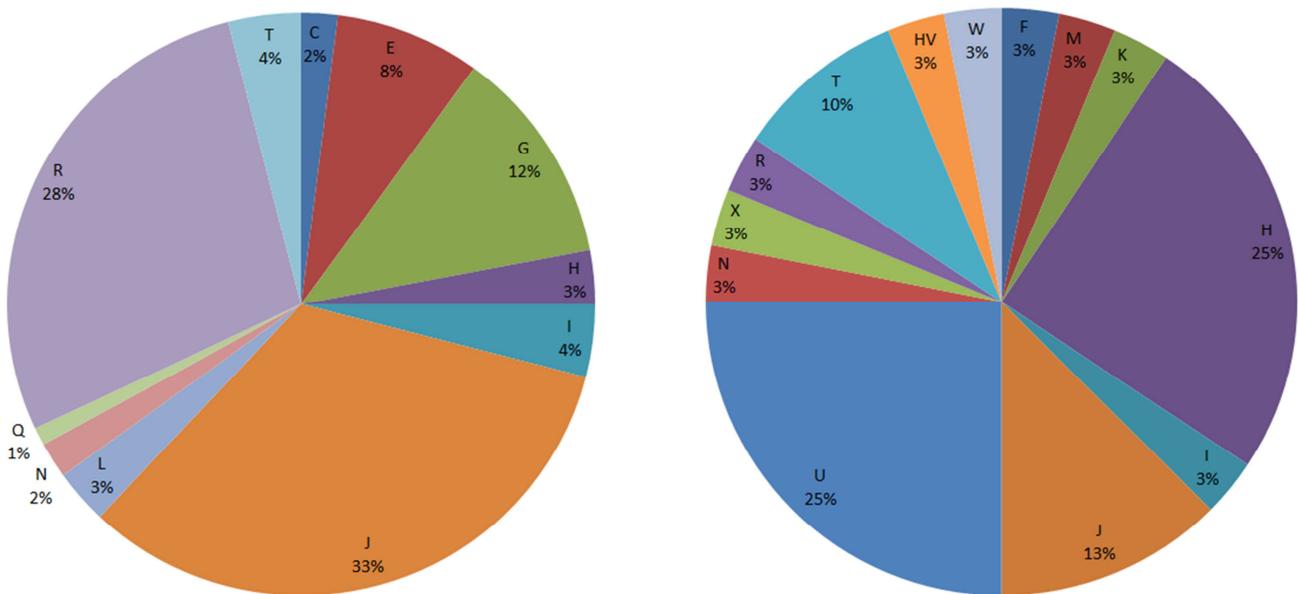


Figure 2. Distribution of Azerbaijanis by Y-DNA haplogroups (left) and by mt-DNA haplogroups (right), based on the FTDNA site data.

When examining the patrilineal part of the gene pool, we see that the results for the dominant haplogroups do not differ

so much, however, for the less common groups, the differences are quite noticeable. Especially, this affected

haplogroups Q, N and T (the differences are both in the direction of decreasing and increasing, more often by 2 times or more). In terms of the diversity of haplogroups in the matrilineal gene pool, the difference, as we can see, also affected the dominant haplogroups. Of course, the results on the diversity of Y-DNA and mt-DNA haplogroups, obtained both from literary sources and from the database of the Azerbaijani DNA project, cannot be accepted unconditionally as the starting materials for the genomic portrait of modern Azerbaijanis. This requires broader population studies covering the population of several regions, both in our country and in Iranian Azerbaijan. In our opinion, for reliable approximate conclusions, the total number of those examined should be at least 1000 people.

Nevertheless, in order to have a more reliable assessment of diversity, at least in terms of the distribution of patrilineal haplogroups among Azerbaijanis, we compared the corresponding results of some authors with data from the Azerbaijan DNA project. For comparison, the results of V.

Grugni et al. [17] and T. Zerjal et al. [13] were selected as the most appropriate. We were guided by the fact that these authors had a sufficiently deep range of study of haplogroups (the absence of "undertyped" variants), and also took into account the high diversification of their results in genome testing.

The comparison was carried out using the Kendall coefficient of concordance, which is also called the coefficient of concordance and is denoted by the letter W. After the calculation procedure, if the value of $W < 0.2-0.4$, then this indicates a weak consistency of opinions of observers (that is, their opinions differ greatly, which does not allow to speak about the presence of some tendency). When obtaining the results, when $W > 0.6-0.8$, we can talk about the agreement of the authors, which in our case means (at least) the similarity of the detected trends. The data that were used for Kendall's test are summarized by us in a separate table (table 3). The variables used were the numbers of specific subjects with a certain result of their DNA test.

Table 3. Comparison of data on the distribution of Y-DNA haplogroups among Azerbaijanis according to information from three independent sources.

A source	Number of examined	Y-DNA haplogroups										
		C	E	G	H	I	J	L	N	Q	R	T
T. Zerjal et al. [13]	19	0	1	1	0	0	11	1	1	0	4	0
V. Grugni et al. [17]	63	0	7	5	0	0	17	2	1	3	23	5
www.familytreedna.com [9]	100	2	8	12	3	4	33	3	4	1	28	2

The calculated Kendall coefficient of concordance W turned out to be 0.8 in our case (this indicates a high degree of agreement between the results of various authors). Evaluation of the significance of the coefficient of concordance using the Pearson matching criterion showed a high reliability of the results of this test at a significance level of $\alpha=0.01$. Thus, a comparison of different, independent from each other results of DNA testing of Azerbaijanis reveals significant similarities in the overall results of these studies. Based on these data, we considered it acceptable to combine these results. According to the statistical technique used, in this case, the probability of making a mistake with the whole (combined) picture of the results from all three considered sources does not exceed 1%. Thus, we got the basis for combining data obtained in the course of different studies. As a result, it became possible to assess the distribution of haplogroups in the patrilineal part of the gene pool based on the testing data of 182 people. Taking into account the above, we took the liberty of presenting a combined diagram of the diversity of Y-DNA haplogroups based on the results of testing Azerbaijanis, both from the Republic of Azerbaijan and Azerbaijanis from Iran (Figure 3).

Without any claims to originality, we believe that at this stage of studying the genetic diversity of Azerbaijanis, such a percentage distribution is the most adequate. The diagram was compiled taking into account the results of a sufficiently large number of those tested and for the first time unites groups from the north and south of historical Azerbaijan. In our opinion, this experience can be extended to the analysis of the results of tests for polymorphism of the HVR1 and HVR2 segments of mitochondrial DNA. However, there is

little work on this topic at the moment. In addition, we would like to know the opinion of colleagues regarding the correctness and validity of our methodology for combining the results of various studies into one format.

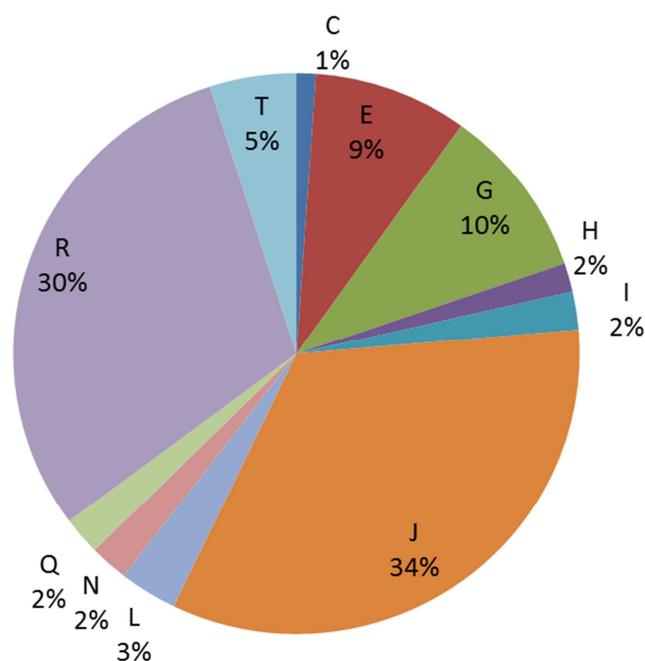


Figure 3. A combined diagram of the distribution of Y-DNA haplogroups in Azerbaijanis.

If we analyze the obtained picture of the distribution of subclades, we find that more than half (54.5%) are

haplogroups of Near Eastern origin - E-M35, G-P15, J-P209 and T-CTS6507, the appearance of which can be associated with tribes from the zone "Fertile Crescent", which inhabited Transcaucasia in the Meso- and Neolithic epochs [18]. The East European subclades of the haplogroup R (R-L584, R-PF7562, R-Y4364) cover about 20% of the investigated and indicate the links of the Transcaucasus with the ancient population of the East European Plain [3]. About 18% are haplogroups of Central Asian origin – C-M217, N-P43, O-F238, Q-M242, R-Z93 and R-M478, which can be compared with the emergence of the Turkic-Mongol populations [19]. 6% are haplogroups L-L595, H-M52 and R-F1092, indicating a connection with the South Asian region [20]. Western European influence (R-L2, I-P78) in total does not exceed 1-2% and is of a random nature.

The information we grouped on the genealogy of Azerbaijanis was also analyzed in the format of comparing them with the data on the anthropology of modern Azerbaijanis. There is no doubt that this kind of analysis within the framework of our study could not be thorough enough. However, we would like to draw your attention to some facts. It should be noted that, based on modern craniological and anthropological knowledge, 4 anthropological types are distinguished among Azerbaijanis: Caspian, Near East, Caucasian and Pontic [21]. These data are consistent with the fact that 4 haplogroups are distinguished in the patrilineal part of the gene pool, which are, as it were, the title haplogroups and characterize 80-85% of the diversity of Y-DNA haplogroups. Two of them are clearly dominant, and haplogroups G and E reach 9-10% of the level. We cannot yet explain the background of this phenomenon, if it exists. It is possible that some questions can be answered after DNA research from bone remains of more ancient historical periods (Neolithic, Mesolithic).

4. Conclusions

Using statistical and mathematical methods, based on the study of information from various independent sources, the distribution of the Y-DNA haplogroups of Azerbaijanis is shown. The relevant statistical analyzes made it possible to draw a conclusion about the possibility of combining similar results of some authors with data from the Azerbaijani DNA project into a single cluster. The description of such a combined sample today provides the maximum presentability for characterizing the genetic diversity of Azerbaijanis. According to the estimates obtained, approximately 55% of the gene pool is made up of haplogroups E-M35, G-P15, J-P209 and T-CTS6507, associated with the settlement of Transcaucasia in the Meso- and Neolithic epoch, about 20% cover the Eastern European subclades of haplogroup R and 18% these are haplogroups of Central Asian origin (C-M217, N-P43, O-F238, Q-M242, R-Z93 and R-M478). The rest falls on the haplogroups of the South Asian (L-L595, H-M52 and R-F1092) and Western European origin, the influence of which on the general picture of the gene pool is insignificant.

References

- [1] Shalyakho R. A., Pocheshkhova E. A., Teuchezh I. E. et al. Turks of the Caucasus: a comparative analysis of gene pools according to data on the Y-chromosome // Bulletin of Moscow University, series XXIII (anthropology), 2013, no. 34-48. юрки Кавказа: сравнительный анализ генофондов по данным о Y-хромосоме // Вестник Московского университета, серия XXIII (антропология), 2013, №2, с. 34-48.
- [2] Asgharzadeh M., Samadi H. K., Ranjbar A. et al. Molecular diversity of mitochondrial DNA in Iranian Azeri ethnicities vis-à-vis other Azeris in Asia // Iranian Journal of Biotechnology, 2011, Vol. 9, №2, p. 120-125.
- [3] Nasidze I., Sarkisian T., Kerimov A., Stoneking M.. Testing hypotheses of language replacement in the Caucasus: evidence from the Y-chromosome // Human Genetics, 2003, Vol. 112, p. 255–261.
- [4] Rootsi S., Myres N. M., Lin A. et al. Distinguishing the co-ancestries of haplogroup G Y-chromosomes in the populations of Europe and the Caucasus // European Journal of Human Genetics, 2012, Vol. 20 (12), p. 1275–1282.
- [5] Underhill P. A., Poznik G. D., Rootsi S. et al. The phylogenetic and geographic structure of Y-chromosome haplogroup R1a // European Journal of Human Genetics, 2015, Vol. 23 (1), p. 124–121.
- [6] Cruciani F., La Fratta R., Santolamazza P. et al. Phylogeographic Analysis of Haplogroup E3b (E-M215) Y Chromosomes Reveals Multiple Migratory Events Within and Out Of Africa // Am. J. Hum. Genet., 2004, Vol. 74 p. 1014–1022.
- [7] Malaspina P., Tsopanomichalou M., Duman T. et al. A multistep process for the dispersal of a Y chromosomal lineage in the Mediterranean area // Ann. Hum. Genet., 2001, Vol. 65, p. 339-349.
- [8] Andonian L., Rezaie S., Margaryan A. et al. Iranian Azeri's Y-Chromosomal Diversity in the Context of Turkish-Speaking Populations of the Middle East // Iranian J Publ Health, 2011, Vol. 40, №1, p. 119-123.
- [9] www.familytreedna.com/public/Azerbaijan/default.aspx.
- [10] Bahmanimehr A., Eskandari G, Nikmanesh F. The congruence between matrilineal genetic (mtDNA) and geographic diversity of Iranians and the territorial populations // Iranian Journal of Basic Medical Sciences, 2015, Vol. 18, №1, p. 30-37.
- [11] Weale M. E., Yepiskoposyan L., Jager R. F. et al. Armenian Y chromosome haplotypes reveal strong regional structure within a single ethno-national group // Human Genetics, 2001, Vol. 109, p. 659-67.
- [12] Wells R. S., Yuldasheva N., Ruzibakiev R. et al. The Eurasian Heartland: A continental perspective on Y-chromosome diversity // PNAS USA, 2001, Vol. 98 (18), p. 10244-10249.
- [13] Zerjal T., Wells R. S., Yuldasheva N. et al. A Genetic Landscape Reshaped by Recent Events: Y-Chromosomal Insights into Central Asia // Am. J. Hum. Genet., 2002, 71, p. 466–482.

- [14] International Society of Genetic Genealogy (2017). Y-DNA Haplogroup Tree 2017, Version 12.222, Date: 21 October 2017. <http://www.isogg.org/tree/index.html>; <https://yfull.com/tree>.
- [15] Glantz St. Primer of Biostatistics, 7th Edition. McGraw Hill Professional, 2011, 320 p.
- [16] Kendall M., Gibbons J. D. Rank Correlation Methods (5th Edition). London, Edward Arnold, 1990, 260 p.
- [17] Grugni V., Battaglia V., Kashani B. H. et al. Ancient Migratory Events in the Middle East: New Clues from the Y-Chromosome Variation of Modern Iranians // PLoS ONE, 2012, Vol. 7 (7) p. 1–14.
- [18] Nasidze, I., Ling, E. Y., Quinque, D., and Dupanloup, I., Mitochondrial DNA and Y-Chromosome Variation in the Caucasus, *Ann. Hum. Genet.*, 2004, vol. 68, pp. 205-221.
- [19] Khusnutdinova, E. K., Litvinov, S. S., Kutuev, I. A. et al. *Russ J Genet* (2012) 48: 640.
- [20] Watkins, W. S., Thara, R., Mowry, B. J., Zhang, Y., Witherspoon, D. J., Tolpinrud, W., Bamshad, M. J., Tirupati, S., Padmavati, R., Smith, H., Nancarrow, D., Filippich, C. and Jorde, L. B. (2008) Genetic variation in South Indian castes: evidence from Y-chromosome, mitochondrial, and autosomal polymorphisms. *BMC Genetics*, 9. doi: 10.1186/1471-2156-9-86.
- [21] Musaev Sh. M. Vozmozhnosti kranio-metricheskikh issledovaniy pri sudebno-meditsinskoy identifikatsii lichnosti zhiteley Kavkaza: avtoref. dis. d-ra med. nauk [Possibilities of craniometric researches at forensic medical identification of the Caucasians: author's abstract of dissertation to apply for the degree of the doctor of medical sciences]. Moscow, 1997, 40 p. [In Russian].