Coll. Antropol. 40 (2016) 2: 99–103
Original scientific paper

Genetic Structure of an Isolated Rural Population in North Albania Evaluated Through Isonymic Method

Ilia Mikerezi and Muhamir Shyqeriu

University of Tirana, Faculty of Natural Sciences, Department of Biology, Tirana, Albania

ABSTRACT

The genetic structure of a rural isolated population living in Kukës District in northern part of Albania, was analyzed through the surnames distribution. The data suitable for this analysis were obtained from Electoral Register (2009) offered by Central Election Commission of Albania. In order to estimate the population diversity, the information from the populations of 10 administrative units (municipalities) for a total of 1768 surnames belonging to 39571 individuals was used. Indicators of genetic structure such as Fisher's α, an estimate of surname diversity and coefficient of consanguinity (F), were obtained. Different genetic distances between all possible pairs of 10 municipalities (populations) were calculated and the correlation with geographic distance was tested. Lasker's, Nei and Euclidean distances were positively correlated with geographic distance, indicating the presence of isolation by distance. In addition, the application of multivariate analysis such as Cluster and Principal Components to isonymic distance matrix revealed that the trend of genetic relationships among the investigated populations was according to their geographic locations. This is an important indication that geographic distance could be a determinant factor in the definition of the above population's genetic structure.

Key words: *isonymy, isolation by distance, population structure, surname distribution, coefficient of consanguinity, Kukës district*

Introduction

In most human populations surnames follow a patrilineal transmission in the same way as Y chromosome. Surnames, even being of cultural and not of biological origin, are transmitted according to rules followed by genes. In this context they can be used as neutral markers to study the populations' genetic structure or isolation and migratory processes (intensity and direction).

Most of investigations conducted in different countries have used data based on the whole country samples. Their results have confirmed the use of surnames as good estimators of populations' genetic structure indicating, in addition, that isonymic distances vary according to geographic ones.

In the last years some investigations have been conducted to analyze the genetic structure of the Albanian population by the use of some metric traits, blood groups frequency and even distribution of surnames[1–3]. In a previous work[4] the whole population has been analyzed through isonymic methods obtaining important indicators on the role of migration and drift in the definition of genetic structure of the population.

As the purpose of the present work we have choose to investigate through isonymic methods the genetic structure of Kukës population, a small and isolated region in North Albania. We were interested to examine the expectations of the isonymic distance variation with geography and to show the possible presence of isolation by distance.

Materials and Methods

Kukësi district

The district lies in the northeastern part of Albania (Figure 1) and is bordered by District of Prizren (Kosovo) in the east and northeast, by Has District in the north, by Pukë District and Mirditë District in the west and by

Received for publication June 21, 2016

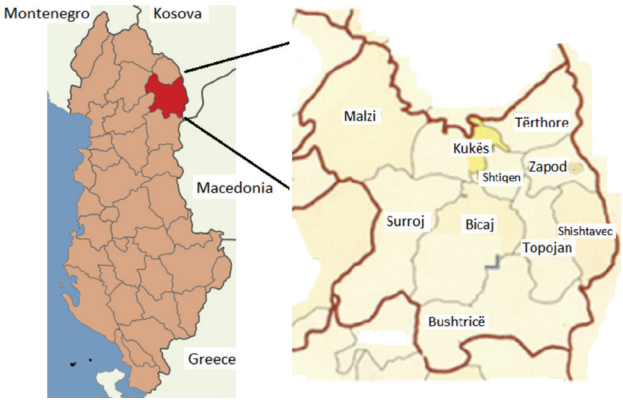

Fig. 1. Geographic position and the populations of Kukës district.

Dibër District in the south. Kukësi (42°04′36″N 20°25′18″E) is the center of the district with the same name. It is 933.86 km^2 in size and has a population of 47,985 (2011 census). It was recently reorganized according to the local government reform by merging the former municipalities Arrën, Bicaj, Bushtricë, Grykë-Çajë, Kalis, Kolsh, Kukës, Malzi, Shishtavec, Shtiqën, Surroj, Tërthore, Topojan, Ujmisht and Zapod.

Data source

All the data used in this paper were the voters' surnames of 2009 parliamentary elections. The data were offered by voting lists of Central Election Commission of the country. In Albania voting is allowed for all the people of age 18 years. From this point of view our sample under investigation represents all the reproductive population. All the individual surnames were registered according to their settlement in 10 municipalities (called populations in the present paper) of Kukës district. As a matter of fact the real number of municipalities (or populations) of Kukës district was 15 but for election reasons, according to Electoral Code (2009), the municipalities Kukës and Kolsh; Bicaj and Ujmisht; Bushtricë, Grykë-Çaj and Kalis; Surroj and Arrën have been merged. Our sample was made of 1768 surnames belonging to 39571 individuals from 10 municipalities (populations) as indicated in Table 1.

Isonymy parameters

Based on surname distribution, Fisher' α, an indicator of surname richness, was calculated according to[5]:

$$1/\alpha = \sum_k (p_{ik})^2 - 1/Ni$$

Random isonymy within each population was calculated as:

$$I_{ii} = 1/\alpha$$

From random isonymy, the coefficient of consanguinity was calculated according to the definition of [6] as

$$F = I_{ii} / 4$$

TABLE 1
SAMPLE SIZE AND SURNAME NUMBERS FOR ALL THE POPULATIONS

Population	N	S
1. Bicaj	5'586	133
2. Bushtricë	2'520	112
3. Kukës	16'067	829
4. Malzi	2'521	156
5. Shishtavec	3'564	156
6. Shtiqen	2'350	74
7. Surroj	1'416	69
8. Tërthore	2'070	73
9. Topojan	1'565	70
10. Zapod	1'912	96
Sum	**39'571**	**1'768**

On the other side, random isonymy between populations was estimated according to[7]as:

$$I_{ij} = \sum_k p_{ik} \times p_{jk}$$

where p_{ik} and p_{jk} are relative frequencies of the k-th surname in the i-th and j-th population respectively and the sum is over all surnames.

In order to assess the isolation by distance, the linear correlation between geographic distances and different surname distances (Nei's, Lasker's, Relethford's and Euclidean) of the populations under investigation were obtained. Nei's distance was estimated according to[8] as

$$N_d = -\log (I_{ij}/\sqrt{(I_{ii} \times I_{jj})}$$

Lasker's distance was defined according to[7] as

$$L = -\ln(I_{ij})$$

Euclidean distance[10] between populations I and J was estimated as

$$D = \sqrt{(1 - \cos\theta)} \quad \text{where} \quad \cos\theta = \sum_k \sqrt{p_{ik} \times p_{jk}}$$

Relethford distance[11] was calculated as

$$d^2 = I_{ii} - 2I_{ij}$$

where I_{ii} and I_{ij} were above defined.

In addition, distances between centers of populations (municipalities) were used as geographic ones that allowed us to construct the matrix of geographic distances. Cluster and Principal Components multivariate analysis were applied to distance matrices. Surname relationships between populations were graphically illustrated through dendrograms that were constructed from distance matrices using UPGMA algorithm of STATISTICA package (version 7).

Results and Discussion

Frequency distribution of surnames

The logarithmic transformation of surname distribution is shown in Figure 2. The graph shows the distribu-

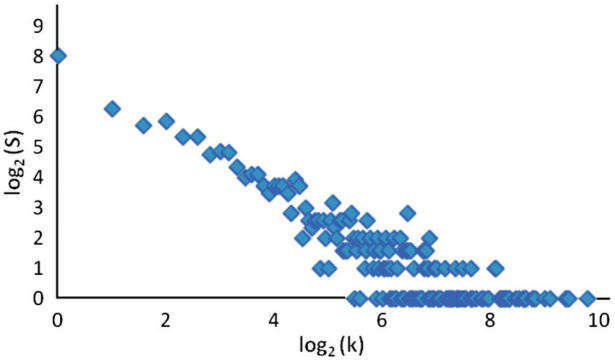

Fig. 2. Logarithmic distribution of surname number (S) occuring k times.

tion of logarithmic value of surname number (S) appearing k times.

The linearity of the distribution is evident and it is very similar to that obtained in some other investigations in different European countries[9,12,13]. It suggests the presence of the relatively late migration process in the populations under the study. This graph enabled us to better analyze the migration process. The increase of the graph's slope during the transition from occurence 2 to occurrence 1 indicates migration prevalence of single individuals over small groups.

In order to better describe especially the spatial structure of the populations under ivestigation some structure parameters were estimated based on distribution of surnames (Table 2).

TABLE 2
SAMPLE SIZE (N), NUMBER OF SURNAMES (S), NUMBER OF SURNAMES OF FREQUENCY 1 (S1), FISHER'S α AND COEFFICIENT OF CONSANGUINITY BY ISONYMY F

Populations	N	S	S1	α	F x (104)
1. Bicaj	5'586	133	43	38.51	64.5
2. Bushtricë	2'520	112	15	53.21	46
3. Kukës	16'067 829	184	139.49	17.8	
4. Malzi	2'521	156	32	50.74	48.3
5. Shishtavec	3'564	156	22	52.82	46.6
6. Shtiqen	2'350	74	21	12.32	201.8
7. Surroj	1'416	69	31	13.40	184.8
8. Tërthore	2'070	73	33	15.86	156.3
9. Topojan	1'565	70	18	7.26	342.4
10. Zapod	1'912	96	15	36.13	67.9
Total	39'571	1'768	414		

One of the most interesting parameters was Fisher's α, an indicator of surname richness and population diversity. A small value of Fisher's α indicates that the population could be closed to migration processes and consequently with large inbreeding and drift. Conversely, a large value of α would indicate an open population to migration and low inbreeding. In terms of population diversity, by taking into account the above parameters we observed that the populations with higher values of Fisher's α were Kukës (α = 139.49), Bushtricë (α = 53.21), Shishtavec (α = 52.82) and Malzi (α = 50.74), whereas those with lower values of Fisher's α were: Topojan (α 7.26), Shtiqen (α = 12.32) and Surroj (α = 13.04). Generally speaking, the populations with higher Fisher's α value would correspond to most densely populated municipalities but this was not always respected; Bicaj population, for example, had higher population size but presented an intermediate value of α.

Another important parameter was the coefficient of consangiunity by isonymy (F = Iii/4) evaluated according to[6]. Obviously, the population with higher F values wold be the most isolated and with less surname diversity. The populations of Topojan, Shtiqen, Surroj and Tërthore that presented tho lowest values of α were the most isolated and consequently with higher level of consanguinity. Distribution of the populations according to F value is shown graphically in Figure 3.

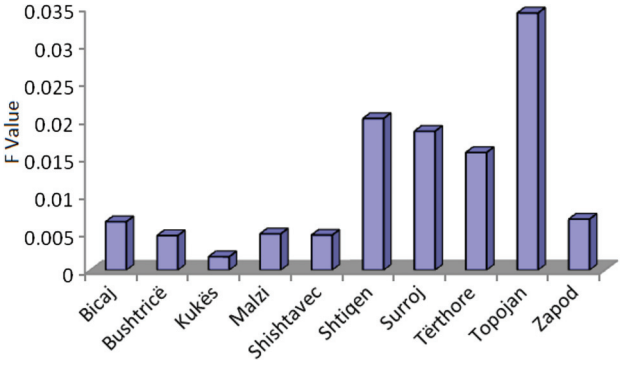

Fig. 3. F value and the distribution of the populations.

Isolation by distance

The presence of isolation by distance was investigated by the analysis of linear correlations between geographic distance and surname ones. The highest values were found with Euclidean (0.640), Lasker (0.582) and Nei (0.455) distances respectively. On the other side, we found high correlation values between Nei – Euclidean (r = 0.697) and Nei – Lasker (r = 0.951) surname distances that were similar to other investigations in some other countries[12,13].

As an example, in Figure 4 it is shown the variation of the geographic and Lasker's distances for the populations under investigation. The increase of geographic distance was followed by the increase of Lasker's surname distance that, otherwise, could be expressed as the decrease of the populations similarity. We think that in the present case, the mountainous relief of the whole district was another isolation factor, other than geographic distance.

Relationships between populations in Kukës district were analysed by applying multivariate Cluster and Prin-

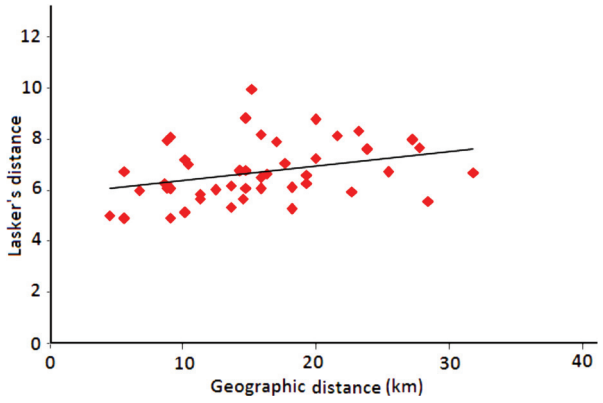

Fig. 4. Variation of Lasker's distance between the populations and the geographic distance. Relationships between populations.

cipal Components analysis to different surnames matrices. Here we are showing the results obtained with Lasker's matrix. The scatter of the populations according to the first two principal components that explained nearly 64% of the variation is shown in Figure 5.

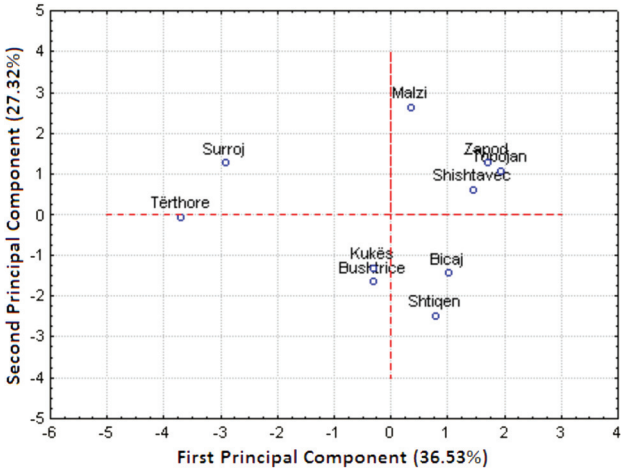

Fig. 5. Scatter of the populations in the space of two Principal Components.

It was observed the tendency of the formation of three main groups: the first one was composed of Kukës, Bushtricë, Bicaj and Malzi, the second included Shishtavec, Topojan, Zapod and Malzi and the third one was formed by Surroj and Tërthore populations.

Nearly the same tendency was observed even by applying Cluster analysis to the same matrix as is shown in the Figure 6.

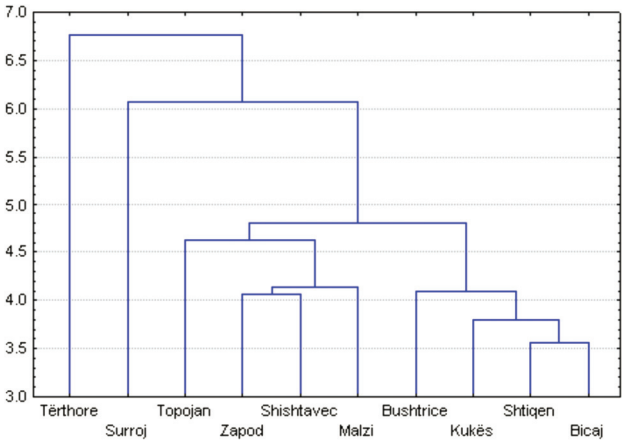

Fig. 6. Relationships between the populations according to Cluster analysis.

It is interesting to stress out that the populations of the same group had similar values of Fisher's α. For example the populations of Kukës, Bushtricë, Bicaj were among them of higher α values (with the exception of Shtiqen population). On the other side, the position of the two populations namely Tërthore and Surroj, that standed nearly at the margins of the scatter, could be explained by the fact thar they were among them of the lowest α value.

Another interesting result was that the three main observed clusters identified generally the geographic neighborhood correspondence for the populations of the same cluster as shown in the Figure 1. The only exception was the position of Malzi population that didn't coincide with that of the cluster.

In conclusion, we can confirm the validiy of isonymy method for the assesment of different population structure and processes even in rather small territories. According to this methodology we were able to analyse and consequently to identify specific charactersistics of the populations of Kukës district. In this context, it was possible to demonstrate the important role of isolation by distance and drift underlying them as special factors in the definition of district genetic structure.

Acknowledgements

We would like to thank the Central Election Commission of Albania for all the information that was made available through Election Code of Albania.

REFERENCES

1. DHIMA A, KUME K, BAJRAMI Z, MIKEREZI I, The Mankind Quarterly 36, Nr 3 & 6 (1996) 271. — 2. SUSANNE C, BAJRAMI Z, KUME K, MIKEREZI I, Gene Geography 10 (1996) 31. — 3. MIKEREZI I, PIZZETTI P, LUCCHETTI E, EKONOMI M, 27 (2003) 2 507. — 4. MIKEREZI I, XHINA E, SCAPOLI C, BARBUJANI G, MAMOLINI E, SANDRI M, CARRIERI A, RODRIGUEZ-LARRALDE A, BARRAI I, Ann. Hum. Genet. 77(3) (2013) 232. — 5. BARRAI I, FORMICA G, SCAPOLI C, BERETTA M, MAMOLINI E, VOLINIA S, BARALE R, AMBROSINO P, FONTANA F, Ann. hum. Biol. 19 (1992) 371. — 6. CROW JF, MANGE AP, Eugen. Quart, 12 (1965) 199. — 7. RODRIGUEZ-LARRALDE A, BARRAI I, NESTI C, MAMOLINI E, BARRAI I, Hum. Biol. 70 (1998b) 1041. — 8. NEI M, 1973 The theory and estimation of genetic distance. In: Morton N Editor. Genetic structure of populations. Honolulu: University Press of Hawaii. — 9. BARRAI I, SCAPOLI C, BERETTA M, NESTI C, MAMOLINI E, RODRIGUEZ-LARRALDE A, Ann. hum. Biol. 23 (1996) 431. — 10. CAVALLI-SFORZA L L, EDWARDS A W, Am J Hum Genet 19 (1967) 233. — 11. RELETHFORD J H, Hum. Biol. 6 (1988) 475. — 12. BARRAI I., RODRIGUEZ-LARRALDE A., MAMOLINI E., MANNI F., SCAPOLI C. Ann Hum Biol Vol 27, n 6, (2000) 607. — 13. RODRIGUEZ-LARRALDE A, DIPIERRI J, GOMEZ E, SCAPOLI C, MAMMOLINI E, SALVATORELLI G, DE LORENZI S, CARRIERI A, BARRAI I, American Journal of Physical Anthropology 144 (2011) 177.

I. Mikerezi

University of Tirana, Faculty of Natural Sciences, Department of Biology, Bulevardi Zog I, Tirane, Albania
e-mail: imikerezi@fshn.edu.al

GENETSKA STRUKTURA IZOLIRANOG RURALNOG STANOVNIŠTVA U SJEVERNOJ ALBANIJI EVALUIRANA UZ POMOĆ METODE ISONIMIJE

SAŽETAK

Genetska struktura seoske izolirane populacije koja živi u okrugu Kukes u sjevernom dijelu Albanije je analizirana kroz distribuciju prezimena. Podaci pogodni za ovu analizu su dobiveni od popisa birača (2009) od Središnjeg izbornog povjerenstva Albanije. Kako bi se procijenila raznolikost populacije, korišteni su podaci iz populacije od 10 upravnih jedinica (općina) za ukupno 1768 prezimena koja pripadaju 39571 pojedincu. Pokazatelji genetske strukture poput Fisherovog α, procjena raznolikosti prezimena i koeficijenta srodstva (F), su dobiveni iz mjerenja. Različite genetske udaljenosti između svih mogućih parova 10 općina (populacije) su izračunate i ispitana je korelacija s geografske udaljenosti. Lasker-a, Nei i euklidske udaljenosti su u pozitivnoj korelaciji s geografske udaljenosti, što ukazuje na prisutnost izolacije po udaljenosti. Osim toga, primjena multivarijatne analize, kao što su klaster i osnovne komponente za isonimijske matrice udaljenosti je pokazala da je trend genetskih odnosa među ispitivanim populacijama bio u skladu s njihovim geografskim lokacijama. Ovo je važan pokazatelj da geografska udaljenost može biti faktor u definiciji genetske strukture gore nevedenog stanovništva.