

# Isonymy and the Genetic Structure of Albanian Populations

Ilia Mikerezi<sup>1</sup>, Paola Pizzetti<sup>2</sup>, Enzo Lucchetti<sup>2</sup> and Milva Ekonomi<sup>3</sup>

<sup>1</sup> University of Tirana, Tirana, Albania

<sup>2</sup> University of Parma, Parma, Italy

<sup>3</sup> Institute of Statistics, Tirana, Albania

## ABSTRACT

*It is well known that in systems of surname transmission through the paternal line, surnames simulate neutral gene alleles belonging to the Y chromosome. This property of surnames was used to analyze the genetic structure of Albanian populations. Two large samples of surnames belonging to two different periods of time were analyzed. The analysis of indicators of population structure showed that geographical distance has an important effect on surname distribution. It seems that isolation by distance and genetic drift have been still important factors in the determination of the genetic structure of the Albanian population.*

**Key words:** isonymy, surname, genetic relationship, human populations

## Introduction

Investigations on the genetic relationships among Albanian human populations have not a long history. In this context, especially in the last ten years, the distribution of blood groups or some other metric traits have been studied<sup>1-3</sup>. Interestingly, these investigations have detected a trend of the isolation by distance. Of course, the geographic structure of Albania favors isolation between different areas.

It is well known that in the human populations surnames follow precise rules that regulate their transmission in a

similar way as genetic traits. When surnames are transmitted through the paternal line, they simulate neutral alleles of genetic markers located on the Y chromosome<sup>4-9</sup>.

From this point of view they present several advantages i.e. the records can be extended to the whole population and data can be obtained by a low cost. On the other side, the system is highly polymorphic and surnames memorize the origin of their carriers possessing, in the same time, an intrinsic linguistic value.

A general outline of the genetic structure of the Albanian populations can be inferred by the distribution of surnames. In this context, they present an interesting problem since Albania is a mountainous country with a high cultural variability. It is therefore possible to study in some details the effect of different factors on the distribution of surnames and, consequently in the genetic relationships among the local populations. In this investigation, the relationships among Albanian local populations are presented. They have been evaluated through the analysis of two samples belonging to two different periods of time.

**Materials and Methods**

The data source was made of two main samples. The first data set was a sample of 25,363 surnames obtained from a census of 1946 and found in the State Archives, whereas the second one was a sample of 47,339 surnames from a owners’ list belonging to the year 1998 kindly offered by the Institute of Statistics in Tirana. The surnames were not distinguished by sex.

The geographic distribution of local populations according to ten administrative regions of Albania is given in Figure 1. Composition of the data, and the composition of the samples for both time periods are given in the Tables 1, 2 and 3 respectively.



*Fig. 1. Geographic distribution of ten local Albanian populations according to ten administrative regions.*

The relationships among pairs of populations were estimated after<sup>3</sup> as the probability that two subjects drawn by chan-

**TABLE 1**  
COMPOSITION OF THE DATA UNDER THE INVESTIGATION

Total number of surnames in both 1946 and 1998 sample	62,594
Total number of surnames in 1946 sample	25,363
Total number of surnames in 1998 sample	47,339
Surnames present in both samples	10,108
Surnames presented in the sample of 1946 but not in 1998	15,255
Surnames presented in the sample of 1998 but not in 1946	37,231

ce from both populations have the same surname:

$$R_{ij} = \frac{\sum p_{ik} p_{jk}}{\sqrt{\sum p_{ik}^2 \sum p_{jk}^2}}$$

where  $p_{ik}$  and  $p_{jk}$  are the relative frequencies of surname  $k$  in the  $i$ -th and  $j$ -th populations respectively.

Non metric multidimensional scaling and cluster analysis were applied to the obtained matrices of distances in order to

obtain the graphical presentation of the scatter of the populations under the investigation. Tree presentation was obtained by *Mega* program.

**Results and Discussion**

*General description of the data*

The composition of the data belonging to both periods of time are shown in Table 1. In the sample of 1946 there were 25,363 surnames, from them 10,108 sur-

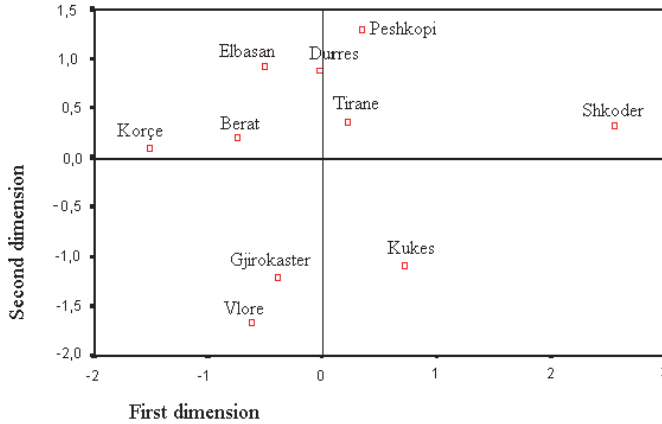


Fig. 2. Scatter of the populations according to the sample of 1946 after multidimensional scaling.

**TABLE 2**  
COMPOSITION OF DATA ACCORDING TO THE SAMPLE OF 1946.

Population	Subjects	Surnames	Subj/Surn	S/N
Berat	119070	4428	28.89	0.037
Durrës	71151	3750	18.97	0.053
Elbasan	100904	2824	35.73	0.028
Gjrokaster	113035	5238	21.58	0.046
Korçe	132796	5701	23.29	0.043
Kukes	45767	1464	31.26	0.032
Peshkopi	53614	1616	33.18	0.030
Shkoder	147673	3917	37.70	0.027
Tirane	128773	7122	18.08	0.055
Vlore	102053	4582	22.27	0.045

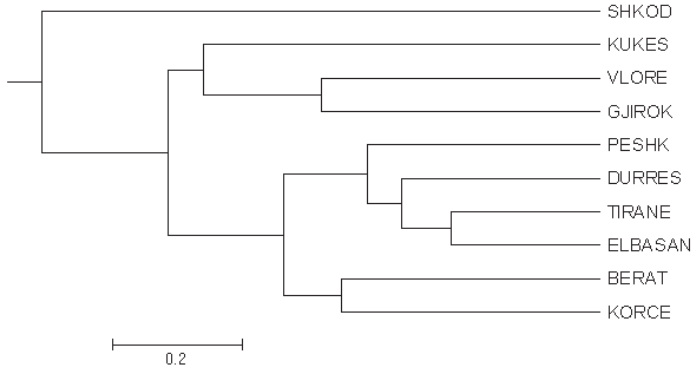


Fig. 3. Scatter of the populations according to the sample of 1946 after cluster analysis.

**TABLE 3**  
COMPOSITION OF DATA ACCORDING TO THE SAMPLE OF 1998

Population	Subjects	Surnames	Subj/Surn	S/N
Berat	58017	9401	6.17	0.162
Durres	47378	7487	6.33	0.158
Elbasan	50332	6607	7.62	0.131
Gjirokaster	32347	7195	4.50	0.222
Korçe	48033	9303	5.16	0.194
Kukes	21472	3806	5.64	0.177
Peshkopi	40519	6309	6.42	0.156
Shkoder	68580	10120	6.78	0.148
Tirane	27653	4562	6.06	0.165
Vlore	56108	7937	7.07	0.141

names are present in the sample of 1998 too. In the sample of 1998 there were 47,339 surnames. The variability of surnames was higher in the 1998 sample than in the sample of 1946. On the other side, it was shown that the S/N ratio was between 0.027 and 0.053 whereas in the first sample but nearly 0.15 in the second one. It means a higher isolation among the populations in the first period. In addition, a lot of surnames were introduced during the period of time that separates two samples. This difference is related

with the presence of the surnames that appear only once.

*Relationships among the populations according to the sample of 1946*

Figure 2 shows 10 populations belonging to the 1946 sample scattered in the space of two first dimensions after the method of non metric multidimensional scaling. It was shown that the first dimension recalls a North-South scatter: Shkoder, Kukes, Peshkopi, Tirane on the right side and all the others on the left.

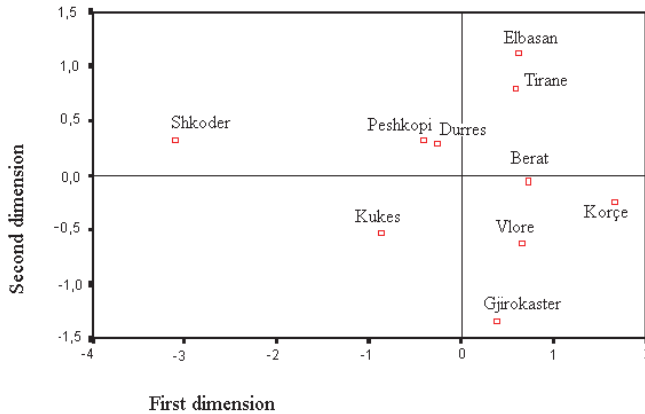


Fig. 4. Scatter of the populations according to the sample of 1998 after multidimensional scaling.

The second dimension could be interpreted as an indicator of the isolation situation of the populations: Kukes, Vlore and Gjirokaster were separated from the others as most isolated ones.

Interestingly, Kukes that is a North population, was included in the same group with Vlore and Gjirokaster most probably for reasons related to the isolation. On the other side, Tirane, Durres and Elbasan formed another group belonging to the central part, whereas Korçe and Berat formed the south-eastern group. In addition, Peshkopi was close enough to the central group most probably for the same reasons of isolation. However, generally speaking, the relationships among the populations were according to the geographic ones.

Figure 3 shows nearly the same scatter of the same populations after the cluster analysis. It was shown that Shkoder population was separated at the first ramification. The other populations were separated in two principal clusters. In the first one, that included the populations of Vlore and Gjirokaster, Kukes was closely related. The second big cluster grouped together the populations of the

Central part (Durres-Tirane-Elbasan-Peshkopi) and Southern part (Korce-Berat) respectively. Again, Peshkopi was very close to the central group for the same reason.

*Relationships among the populations according to the sample of 1998*

The results of the analysis of the 1998 sample were analogous to them of 1946 sample. The Multidimensional scaling analysis (Figure 4) showed clearly the separation of Shkoder and Korçe belonging to North and South parts of Albania respectively, according to the first dimension.

On the other side, the populations of the Central part Tirane and Elbasan appeared to be more close to each other especially on the first dimension that conserves the significance of North-South direction. Most probably, this could be caused by the facility of the communication or by the urbanisation phenomena during nearly 50 years that separate the two surname samples.

The tree presentation of the same data (Figure 5) confirmed the same results of Multidimensional scaling. However, Be-

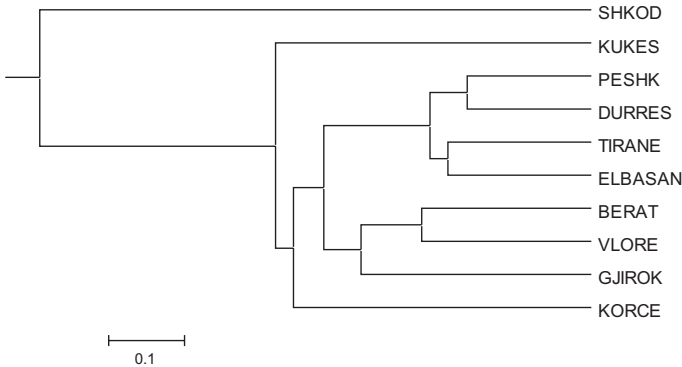


Fig. 5. Scatter of the populations according to the sample of 1998 after cluster analysis.

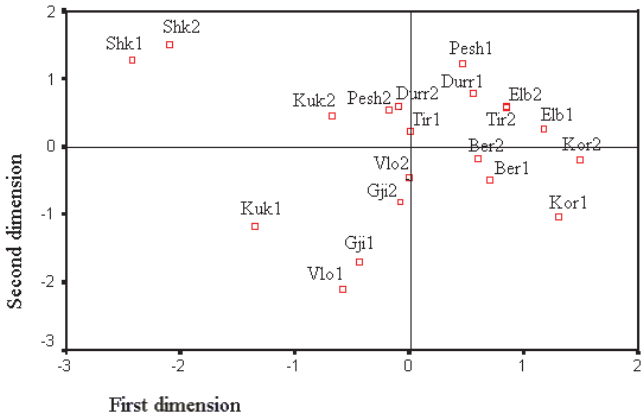


Fig. 6. Scatter of the populations according to aggregated data after multidimensional scaling.

rat that was included in the same group with Korçe in the tree of 1946, appears in the same group with Vlore and Gjiro-kaster. These new relationships could be caused, most probably, by the better ways of communication between these two regions in the last 50 years.

Generally speaking, the tree presentation confirmed the North-South trend of the populations' variability. In addition, the population of Kukës and Korçe were more differentiated from the central group of Durres-Tirane-Elbasan.

*Relationships among the populations according to the aggregated data.*

Figure 6 shows the scatter of the same populations according to the data of both sampling periods. It was shown that the same populations belonging to different periods of time were nearly overlapped to each other. The populations of Kukës, Vlore showed some differences among both periods of time. In addition, nearly all the local populations tended to be more grouped according to the period 2 (1998) when compared with the period 1 (1946).

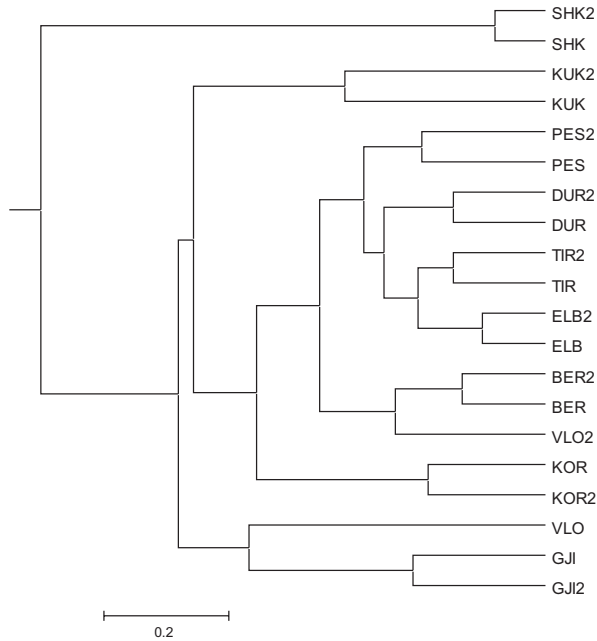


Fig. 7. Scatter of the populations according to the aggregated sample after cluster analysis.

This situation could be explained by a higher homogeneity as consequence of a mixture of the local populations or by a higher variability of the surnames observed in the second period. Nearly the same situation was shown by the tree presentation (Figure 7). It could be stressed the variation of Vlore that was associated with Gjirokaster in the scatter of 1946 and with Berat in the scatter of 1998.

All the above results obtained by different methods of analysis showed that geographical position of the local populations has an important effect on surname distribution. As other previous investigations have indicated, it seemed that isolation by distance and genetic drift have been important factors in the determination of the genetic structure of the Albanian populations.

The improvement of communication ways and the higher population mobility in the last years are manifested in a higher homogeneity as shown in Figure 6. In the same topological space, different local populations of the sample of 1998 are scattered in a more central position compared with them of 1946, conserving on the other side, their own peculiar features.

The relationships among Albanian human populations estimated according to the frequency of surnames are function of a North-South trend of differentiation. In addition, the presence of the isolated populations, because of the limited ways of communication and their geographical position, has been shown.

## REFERENCES

1. BELLEDI, M., S. E. POLONI, R. CASALOTTI, F. CONTERIO, I. MIKEREZI, J. TAGLIAVINI, L. EXCOFFIER, Eur. J. Hum. Genet., 8 (2000) 480. — 2. DHIMA, A., K. KUME, Z. BAJRAMI, I. MIKEREZI, The Mankind Quarterly, 36 (1996) 271. — 3. SUSANNE, C., Z. BAJRAMI, K. KUME, I. MIKEREZI, Gene Geog., 10 (1996) 31. — 4. CROW, J. F., Hum. Biol., 52 (1980) 1. — 5. CHEN, K., L. L. CAVALLI-SFORZA, Hum. Biol., 55 (1983) 367. — 6. LUCCHETTI, E., E. SIRI, L. SOLIANI, Riv. di Antrop., 74 (1996) 171. — 7. PIAZZA, A., S. RENDINE, G. ZEI, A. MORONI, L. L. CAVALLI-SFORZA, Nature, 329 (1987) 714. — 8. RODRIGUEZ-LARRALDE, A., G. FORMICA, C. SCAPOLI, M. BERETTA, E. MAMOLINI, I. BARRAI, Ann. Hum. Biol., 20 (1993) 261. — 9. BARRAI, I., C. SCAPOLI, M. BERETTA, C. NESTI, A. RODRIGUEZ-LARRALDE, Ann. Hum. Biol., 23 (1996) 431.

*I. Mikerezi*

*University of Tirana, Tirana, Albania*

## IZONIMIJA I GENETSKA STRUKTURA STANOVNIŠTVA ALBANIJE

### SAŽETAK

Dobro je poznato da u sustavu prijenosa prezimena putem paternalnih linija, prezimena simuliraju alele neutralnih gena koji pripadaju Y-kromosomu. Ova karakteristika prezimena korištena je kako bi se analizirala genetska struktura stanovništva Albanije. Analizirana su dva velika uzorka prezimena koji pripadaju dvama različitim razdobljima. Analiza indikatora populacijske strukture pokazala je kako zemljopisna udaljenost ima značajan učinak na raspodjelu prezimena. Izgleda da izolacija s udaljenošću i genetički otklon su bili značajni čimbenici u određivanju genetske strukture stanovništva Albanije.