

**I CONGRESO IBEROAMERICANO DE ANTROPOLOGÍA: *ANTHROPOS 2007***  
**La Havana, Cuba 5-9 March 2007**

**LANGUAGE BOUNDARIES AND MICROEVOLUTIONARY PROCESSES IN SOUTH-EASTERN EUROPE<sup>1</sup>**

**Anita Sujoldzic**

**Institute for Anthropological Research, Zagreb, Croatia**

**e-mail: [anita@inantro.hr](mailto:anita@inantro.hr)**

**ABSTRACT**

A basic finding of human population-structure analysis is that cultural variables influence this process, while language is one of the most important of these cultural variables. A number of anthropological studies throughout the world carried out at every population level, from local and regional to global, have demonstrated consistently a direct relation between genetic diversity and language diversity. This evolutionary congruence is explained by the similarity between factors which operate to increase and consolidate genetic differences between and within populations, such as migration, geographic separation, or isolation, and the processes by which linguistic variation within a community can develop into differentiation of dialects or languages. Using a regional approach, the present study is aimed at depicting the history of human migrations in the Eastern Adriatic and Balkan Peninsula by comparing genetic, cultural and linguistic differences that reflect different aspects of its history by applying the Monmonier's maximum difference algorithm for the visual identification of linguistic and genetic boundaries. It is focused on the comparison of two different boundary maps: one based on the regional distribution of genetic markers obtained by typing the human Y chromosome that reflects genetic differences between populations linked to the populations' past demographic history, and the other based on the regional distribution of the recorded historical, Illyrian, pre-Slavic surnames. The results have major implications for regional population structure and indicate that genetic patterns of modern Balkan populations to a considerable extent still reflect the prehistoric ethnic and hence genetic diversity of these early European populations.

**INTRODUCTION**

---

<sup>1</sup> This work was supported by the Croatian Ministry of Science and Technology (grants 0196002, 0196003 and 0196005).

## **Balkan Geography**

The term «Balkan» refers to the southeasternmost peninsula of Europe and collectively, the countries located there. The Balkan Peninsula was named after a mountain range, the Stara Planina in northern Bulgaria whose Turkish name is Balkan (mountain). The relief of the region is shaped by mountain ranges: the Balkan Mountains lie east-west across Bulgaria, the Rhodope Mountains extend along the Greek-Bulgarian border, the Dinaric range extends down the Adriatic coast to the Pindus Mountains in Greece, while in the northwest there are the Julian Alps and the Carpathians. Although there is no sharp separation in the north between the peninsula and Central Europe, the line of the Sava and Danube rivers is commonly considered to be its northern limit. The geographical area covers about 629,000 square kilometers with a population of about 70 million people divided into many countries, including Albania, Bosnia-Herzegovina, Bulgaria, Macedonia, the mainland of Greece and the European part of Turkey. Geographically, it also includes parts of Croatia, Slovenia, Serbia and Montenegro (formerly Yugoslavia) and SE Romania, which are often considered Balkan countries as a whole because of their close ties with the region in history and politics.

Geographically, the area is composed of two parts: one southern and Mediterranean, the other northern and continental.

The southern part is open to the easily navigable seas with The Adriatic and Ionian seas on the west, the Sea of Marmora, the Dardanelle and the Aegean Sea on the south, and The Black Sea and the Bosphorus on the east. Surrounded on three sides by water, the peninsula has been linked through history with deep and lasting bonds to Anatolia, and Asia beyond, the Italian peninsula to the west, and to the eastern Mediterranean and Egypt to the south. It has been therefore always open to various influences and pressures in a permanent exchange of populations to and from all these destinations throughout the whole of its history. This communicative permeability and continuous intimate contacts of its peripheral parts with neighboring non-Balkan areas are contrasted with the central mountainous part characterized by isolation and perpetuation of a most conservative way of life.

Except for the northern plains, which are quite open towards Central Europe, the continental part is characterized by the mountains that have contributed to the continued fragmentation of human groups in the area. They have been physical obstacles, preventing efforts at regional economic and cultural integration dividing the region into small units, in which distinct ethnic groups have tended toward distinct national cultures and languages, local economies, and political autonomy. The mountains have also subdivided the whole area into vertical ecological zones, ranging from

lowland farming areas to wooded or rocky uplands acting as places of exile and refuge for defeated ethnic groups expelled from more desirable coast and valley lands.

Although deeply different, the two parts of the Balkan peninsula are connected by all sorts of historical, cultural and linguistic links as several rivers crossing the almost continuous mountain barrier provided for continental roads which, though difficult and dangerous, were always important as lines of communication and channels of exchange of goods and ideas.

### **Balkan History**

Since the beginning of its history Balkan has been a crossroads of migration and mixing of peoples for many centuries, so that ethnic diversity has come to be one of its most prominent features. Many different peoples have migrated into the area absorbing or being absorbed by already existing groups that left traces in the languages spoken there in different layers of language contact.

The whole area shows the same stratification of linguistic elements since the early prehistory, which reflects the successive settlement of the region by different populations until today.

Archaeological and historical data report that the Balkan Peninsula had been an inhabited area even before the immigration of the Indo-European (IE) tribes. This pre-IE substratum (or substrata) belonged to an old pre-Indo-European language complex of Mediterranean languages, far related to Indo-European (probably linked with the expansion of agrarian economy along the coast of Mediterranean around 5000 BC).

The gradual indo-europeanization of the Balkans was a rather long process occurring during the second millennium BC. It brought to the region tribal groups speaking several Indo-European languages termed collectively Palobalkan languages: Illyrian and Venetic to the northwest of the peninsula, Thracian and Ancient Macedonian to the northeast, Dacian in the north and Greek to the southeast. With the exception of Greek, very little is known about them from the fragments conserved in ancient Greek manuscripts consisting mainly of toponymic terms, personal names and some words, aside from the fact that they are all of Indo-European origin. They all seem to be related and had a set of common traits in syntax and morphology with Greek dialects and other languages of the region. The Greeks established colonies outside Greece during the last millennium BC while during the reign of the Macedonian Alexander the Great in the 4th century BC, much of the Balkan Peninsula came under Greek influence so that the Ancient Macedonian language assimilated entirely by Greek disappeared in the 3rd century BC. This process of

Hellenization of the Balkans caused the formation of several common features in all languages that were spoken here later in history.

The Romans first invaded the peninsula in the third century BC and by the first century AD it was entirely under their control. At the height of Roman power, the Balkan peoples were united under a common legal system, a single ultimate political and military power, and connected through several important commercial routes. From the third century BC to the time of Slav immigration in the sixth century AD, when the entire region belonged first to the Roman and later to the Byzantine Empire, radical romanization occurred that is reflected in a number of toponymic traces, while the Latin language of the Roman domination represents a major linguistic substratum of the subsequent Slavic languages.

Venetic was quickly assimilated by Latin under the Roman rule in the 1st century, Thracian language disappeared in the fifth or sixth century AD, as Thracians were first assimilated by Romans and later by Slavs and Bulgars. The Dacians were assimilated by Romans and are thought to be ancestors to the present Romanians. Some Illyrian tribes were either romanized or assimilated by later Slavic migrations, but others moved south into present-day Albania, where they managed to keep their identity, including Illyrian language that is said to be the ancestor of modern Albanian language. Some of the romanized Thracian and Illyrian populations survived by taking refuge in the isolated mountain areas where they lived like shepherds conserving several Balkan Romance varieties.

After the fall of the Roman Empire, this Latin was considerably modified in isolation, developing through absorption of different Paleobalkan substrata in what is today called the Vulgar Latin variants. The mixture of those tongues with later additions of Slavic and other elements created interesting structures and the vocabulary of several Romance varieties still spoken in the Balkans: Daco-Romanian, Aromanian, Istro-Romanian and Istro-Romance or Istriot, in addition to Dalmatian, an extinct language spoken along the Adriatic coast until the beginning of the 20th century.

### **Aim of the study**

Using a regional approach, the study continues previous research aimed at depicting the history of human migrations in the Eastern Adriatic and western Balkan by comparing genetic, cultural and linguistic differences that reflect different aspects of its history. The existing historical, archeological and linguistic data indicate a considerable regional variation in cultural development within this area since the Early Iron Age, including several cultural bases and

transition zones. The historic Illyrians inhabited large parts of the western Balkans during the first millennium BC. There is a continuing scholarly disagreement on which tribes should be included in the group of peoples called Illyrian, and the boundaries of the Illyrian lands have never been exactly defined. Classical sources mention more than 80 smaller or larger tribes, while a consolidation of ethnic groups took place in the course of the Iron Age when Illyrian “aggregate tribes” covering numerous smaller tribes began to emerge as better defined larger ethnic entities. Their geographic distribution in the area of the Eastern Adriatic based essentially on written resources is presented in Figure 1. The Illyrian territory covers the Adriatic coast and its hinterland from Albania in the south to Istria to the north, including Montenegro, Kosovo, Serbia, Herzegovina, Bosnia, Dalmatia and Northern Adriatic. Previously obtained results indicated quite a heterogeneous ethnic composition of the historic populations named Illyrians, and clearly demonstrated a clear north-western/southeastern gradient pointing a remarkable discrimination of the Northern Adriatic from other East Adriatic populations (Sujoldžić, 2000). The present study is therefore directed toward further understanding of the processes that caused the patterns of the colonization of this area by using Y chromosome sequence proved to be a valuable tool for the purposes of human history reconstruction. Support for a partially ethnohistorical determination of modern genetic patterns comes from previous studies linking linguistic patterns with genetic variation which show that the areal distribution of genetic markers largely corresponds to that of the world languages, while substandard dialect microareas also have close correspondences with the finer genetic differentiation (Contini et al. 1989, Sokal et al., 1993, etc.).

## **METHODS, SAMPLE AND DATA SOURCES**

The study focuses on 9 Iron Age Illyrian communities: *Noricum* in the north-western edge of the Balkans, *Veneti* in the north-eastern Italy, *Histri* in the Istrian Peninsula, *Liburni* in the northern Adriatic area between the rivers of Krka and Zrmanja, *Japodes* in Lika and eastern Bosnia behind the Velebit mountains, *Delmatae* in Middle Dalmatia and southern-central Bosnia, *Autariates* in Montenegro, Herzegovina, southeastern Serbia, *Dardanians* in Kosova and southcentral Serbia region and *Taulanti* in central Albania. The present analysis is based on the comparison of two different sets of data: the regional distribution of genetic markers obtained by typing the human Y chromosome that reflects genetic differences between present populations linked to the populations’ past demographic history, and the regional distribution of the recorded Illyrian surnames used as a proxy to Y-chromosome genetic variation and an indicator of

historical linguistic differences. The comparison of genetic and surname data was made by applying the Monmonier's maximum difference algorithm. The method provides a more detailed visualization of the geographic patterns of variation by identifying the “barriers” or genetic boundaries representing zones of abrupt change in the pattern of genetic variation and enables a better interpretation of microevolutionary processes through the identification of hidden boundaries resulting from secondary gene flow among previously isolated populations (Manni et al., 2004).

#### *Genetic data.*

Paternal genetic heritage was studied on DNA from 1,141 individuals analyzed for Y chromosome markers (data from the Institute for Anthropological Research, Zagreb, Croatia and from the published sources) (Barač et al., 2003; Rootsi et al., 2004; Peričić et al., 2005a, 2005b). The individuals were from the continental parts of the Balkan peninsula, the north-west Adriatic and the south-east Alpine areas as well as from the Eastern Adriatic islands, i.e. the areas where the different Illyrian tribes settled during the Iron Age period (Fig.1). The sample was analyzed for Y chromosomal haplogroup frequencies in 9 populations and classified as I1a, I1b\*-P37, R1a-SRY1532, R1b-M173, E3b1-M78 and J2e1 haplogroups. Detailed protocols for typing of Y-chromosomal polymorphisms and the selection of informative markers were reported by Barač et al. (2003) and Peričić et al. (2005a, 2005b). An *Fst* distance matrix between these populations was then computed.

*Linguistic data.* The limited remains of the Illyrian language consist of onomastic evidence including names of persons, peoples and places as recorded in Greek and Roman sources. The linguistic boundaries of these communities and their interrelationship were determined using the spatial distribution of 400 personal names compiled by Mayer (1957). Surnames can be considered one locus on the nonrecombining portion of the Y chromosome, and their analysis allows us to infer genetic structures of populations. As the main characteristic of this material is the exclusive occurrence of particular personal names in specific areas, the Euclidean distances were calculated between thus defined onomastic areas.

## **RESULTS AND DISCUSSION**

The results of the frequencies of selected haplogroups analyzed in present populations living in historically Illirian area indicate clear differences in their distribution (Table 1).

1) **Haplogroup I (I1b-P37)**: The highest frequency is found for **haplogroup I1b-P37** (Rootsi et al., 2004), in the historical continental areas of **Autariates** (63.80%) and **Delmatae** (48%), while

it decreases toward northern **Noricum** (20%) and toward south eastern areas of **Taulanti** (17%), **Dardanians** (13.50%) and **Liburni** (9.5%), with a very low value in **Veneti** (0.5%). This haplogroup is not present in the populations from the areas settled by **Japodes** and **Histri**.

2) **Haplogroup R** (R1b-M137): The highest frequency distribution of **R1b-M173** (Kivisild et al., 2003) in the area of **Veneti** (62%), decreasing in that of **Autariates** (27.27%) and **Japodes** (26.67%), **Noricum** (21%), **Histri** (18.18%), **Taulantii** (17.60%), **Liburni** (16.40%), **Dardanians** (14.55%) and **Delmatae**(7.06%).

3) **Haplogroup R1a**:. Its frequency peaks are in north-western areas of **Histri** (45.45%), **Liburni** (38.40%) and **Noricum** (37%), with decreasing frequency gradient slowly to the south-east in **Delmatae** (22.58%), **Japodes** (13.33%) and **Autariates** (12.10%), with the lowest values in **Taulanti** (9.80%) and **Dardanians** (8.82%). This haplogroup is absent in the area of **Veneti**.

4) **Haplogrupa E** (E3b1-M78): The frequencies in our sample are the following: **Dardanians** (36.80%), **Taulantii** (27%), **Japodes** (13.33%), **Veneti** (10.40%), **Histri** (9.09%), **Autariates** (8.50%) and **Delmatae** (4.66%). In the area of **Noricum** this haplogroup is absent.

5) **Haplogroup J** (**J2-M172**): The frequency peak is found in **Dardanians** (15.02%) and **Taulanti** (14.30%), decreasing in **Liburni** (10.90%), **Veneti** (9.60%) and **Japodes** (6.67%), reaching rather low values in **Delmatae** (2.09%) and **Autariates** (1.40%). It is absent in **Noricum** and the area of **Histri**.

The dendrogram of genetic distances (Figure 2) based on the *Fst* distance matrix shows three main clusters: one is composed of northern Histri, Liburni and Noricum, the second is formed by southeastern Taulanti and Dardanians, joined by Japodes, while a separate cluster is formed by Delmatae and Autariates. The clustering reflects the southeast-northwest division, with the continental mountain areas as a containment of the most stable and the most conservative element with roots in deep history.

The dendrogram of linguistic distances presented in Figure 3, based on the matrix of Euclidean surname distances (Table 2) clearly defines specific onomastic areas and the division between southeastern and northwestern tribes. The presence of two main clusters shows the clear separation of Noricum, Veneti, Histri and Liburni in the North Adriatic area from all other analyzed Illyrian tribes, probably as a result of a strong Venetic cultural influence mentioned in literature (Katičić, 1963). The cluster formed by the other tribes indicates also considerable language and/or dialectal differentiation. The distinct position of Japodes with Dalmatae within this cluster reflects various northern influences on their onomastics, including that of Celts. The

clustering of Autariates and Taulanti probably distinguishes them as “true Illyrians”, while the position of Dardanians is due to a considerable contribution of Thracian elements.

Figure 4 presents the genetic barriers detected in regional Y chromosome haplogroup distribution by Monmonier’s algorithm. The first and the strongest barrier (a), computed on the *Fst* distance matrix, divides the northern area of Veneti from all other investigated populations and mirrors the genetic differences between central-western and southeastern Europe. The second barrier (b) encloses the most southeastern areas of Taulanti and Dardanians. The third genetic barrier (c) divides the areas of Delmatae and Autariates from the remaining surrounding populations. The fourth barrier (d) simultaneously divides the remaining populations in the north-west into three separate genetic areas, i.e., Noricum, Japodes and the joined areas of Histri and Liburni. The last barrier (e) finally divides Dalmatae and Autariates into two separate areas.

Figure 5. shows the boundaries in genetic structure inferred from surnames. The strongest barrier (a) indicates that the main differentiation in the distribution of Illyrian surnames occurs along the northwestern-southeastern direction, separating the areas of Noricum, Veneti, Histri and Liburni from all other groups. The second differentiation zone (b) encloses the area of Dardanians, while the third boundary (c) appears between Dalmatae and Autariates. The fourth barrier (d) encloses Liburni within the northwestern area, while the fifth barrier (e) the differentiates between Delmatae and Japodes.

Figure 6 presents a geographic map with combined genetic and surname barriers. It shows a fairly good congruence of surname and genetic boundaries, as they overlap in almost all cases with the exception of the genetic barrier (a) that separates Veneti from all other groups, and the linguistic boundary between Taulanti and Dardanians (b) due to the Thracian linguistic influence on the latter.

## **CONCLUSIONS**

The obtained results demonstrate a clear northwestern/southeastern gradient pointing to a remarkable discrimination of the Northern Adriatic and Istria from other southeastern populations. The sub-differentiation of areas reflects quite clearly the effects of relative isolation of the continental mountain range with most conservative genetic traits as well as exposure of other analyzed areas to different neighbours, ecological niches, waves of intrusive immigrants introducing agriculture and later in Metal Ages, to waves of invading elites of different origin. The results also show that the linguistic and cultural differences are at least partially reflected by the present genetic differences and that in spite of subsequent historical demographic events in the region, the traces of genetic individuality of ancient peoples have been to a certain degree

preserved through the process of assimilation. Although these signals from the past detected in present genetic structure should be taken with caution and verified in further research, the results obtained affirm the usefulness of the method applied and provide a further historical and geographic insight into the role of the local migrations and cultural factors responsible for regional genetic diversity.

## REFERENCES

- BARAĆ L., PERIČIĆ M., MARTINOVIĆ KLARIĆ I., ROOTSI S., JANIĆIJEVIĆ B., KIVISILD T., PARIK J., RUDAN I., VILLEMS R., RUDAN P. 2003. Y chromosomal heritage of Croatian population and its island isolates. *European Journal Human Genetics* 11: 535-542.
- CAVALLI-SFORZA L.L., A.PIAZZA, P. MENOZZI, J.MOUNTAIN. 1988. Reconstruction of Human Evolution: bringing together Genetic, Archaeological and Linguistic Data, in *Proc. Natl. Acad. Sciences USA*, 85, pp. 6002-6006.
- CONTINI M., N.CAPPELLO, R.GRIFFO, S.RENDINE, A.PIAZZA. 1989. Géolinguistique et géogénétique: une démarche interdisciplinaire, in «Géolinguistique» 4, 129-197.
- KATIČIĆ, R. 1963. Das mitteldalmatinische namengebeiet, *Živa antika* 12.
- MAYER, A. Die Sprache der Alten Illyrier 1-2, Wien, 1957-9.
- KIVISILD T., ROOTSI S., METSPALU M., MASTANA S., KALDMA K., PARIK J., METSPALU E., ADOJAAN M., TOLK H. V., STEPANOV V., GOLGE M., USANGA E., PAPIHA S. S., CINNIUGLU C., KING R., CAVALLI-SFORZA L., UNDERHILL P. A., VILLEMS R. 2003. The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. *American Journal of Human Genetics* 72: 313-332.
- MANNI, F., E GUERARD, E. HEYER. 2004. Geographic Patterns of (Genetic, Morphologic, Linguistic) Variation: How Barriers Can Be Detected Using Monmonier's Algorithm. *Human Biology* 76: 173-190.
- PERIČIĆ M., BARAĆ LAUC L., MARTINOVIĆ KLARIĆ I., ROOTSI S., JANIĆIJEVIĆ B., RUDAN I., TERZIĆ R., ČOLAK I., KVESIĆ A., POPOVIĆ D., ŠIJAČKI A., BEHLULI I., ĐORĐEVIĆ D., EFREMOVSKA LJ., BAJEC Đ. D., STEFANOVIĆ B. D., VILLEMS R., RUDAN P. 2005a. Highresolution phylogenetic analysis of Southeastern Europe (SEE) traces major episodes of paternal gene flow among Slavic populations. *Molecular Biology and Evolution* 22(10): 1964- 1975.
- PERIČIĆ, M., L. BARAĆ LAUC, I. MARTINOVIĆ KLARIĆ, B. JANIĆIJEVIĆ, P. RUDAN. 2005b. Review of Croatian Genetic Heritage as Revealed by Mitochondrial DNA and Y Chromosomal Lineages. *Croatian Medical Journal* 46: 502-513.
- ROOTSI S., MAGRI C., KIVISILD T., BENUZZI G., HELP H., BERMISHEVA M., KUTUEV I., BARAĆ L., PERIČIĆ M., BALANOVSKY O., PSHENICHNOV A., DION D., GROBEI M., ZHIVOTOVSKY L. A., BATTAGLIA V., ACHILLI A., AL-ZAHERY N., PARIK J., KING R., CINNIUGLU C., KHUSNUTDINOVA E., RUDAN P., BALANOVSKA E., SCHEFFRAHN W., SIMONESCU M., BREHM A., GONCALVES R., ROSA A., MOISAN J-P., CHAVENTRE A., FERAČ V., FUREDI S., OEFNER P. J., SHEN P., BECKMAN L., MIKEREZI I., TERZIĆ R., PRIMORAC D., CAMBON-THOMSEN A., KRUMINA A., TORRONI A., UNDERHILL P. A., SANTACHIARABENERECETTI A. S., VILLEMS R., SEMINO O. 2004. Phylogeography of Y-

chromosome haplogroup I reveals distinct domains of prehistoric gene flow in Europe. *American Journal of Human Genetics* 75: 128-137.

SEMINO O., PASSARINO G., OEFNER P. J., ARBUZOVA S., BECKMAN L. E., DE BENEDICTIS G., FRANCALACCI P., KOUVATSI A., LIMBORSKA S., MARCIKIĆ M., MIKA A., MIKA B., PRIMORAC D., SANTACHIARA-BENERECETTI A. S., CAVALLI-SFORZA L. L., UNDERHILL P. A. 2000. The genetic legacy of Paleolithic *Homo sapiens sapiens* in extant Europeans: a Y chromosome perspective. *Science* 290: 1155-1159.

SOKAL, R. G. M. JACQUEZ, N.L. ODEN, D. GIOVANNI, A.B. FALSETTI, E. MC GEE, B. A. THOMSON. 1993. Genetic Relationships of European Populations reflect Their Ethnohistorical Affinities, *Am. J. Physical Anthropol.* 91:55-70.

STIPČEVIĆ, A. *Iliri*, Školska knjiga, Zagreb, 1988.

SUJOLDŽIĆ, A. 2000. Linguistic and cultural diversity of the ancient Illyrians reflected in modern genetic variances. 12th Congress of the European Anthropological Association *Millennial Perspectives: Past, Present and Future*. Mascie-Taylor, N (ed.). Cambridge: The Royal Anthropological Institute and *Annals of Human Biology*, 38.

WELLS R. S., YULDASHEVA N., RUZIBAKIEV R., UNDERHILL P. A., EVSEEVA I., BLUE-SMITH J., JIN L., SU B., PITCHAPPAN R., SHANMUGALAKSHMI S., BALAKRISHNAN K., READ M., PEARSON N. M., ZERJAL T., WEBSTER M. T., ZHOLOSHVILI I., JAMARJASHVILI E., GAMBAROV S., NIKBIN B., DOSTIEV A., AKNAZAROV O., ZALLOUA P., TSOY I., KITAEV M., MIRRAKHIMOV M., CHARIEV A., BODMER W. F. 2001. The Eurasian heartland: a continental perspective on Y-chromosome diversity. *Proceedings of National Academy of Science USA* 98: 10244-10249