OVERVIEW OF THE AUTOSOMAL STR CLUSTERING BETWEEN BALKAN POPULATIONS

Adna Ašić[1, *], Bea Bunjo[1], Serkan Doğan[1], Larisa Bešić[1], Imer Muhović[1], Damir Marjanović[2]

[1]International Burch University, Department of Genetics and Bioengineering, Sarajevo, Bosnia and Herzegovina,
[2]University of Sarajevo, Institute for Genetic Engineering and Biotechnology, Sarajevo, Bosnia and Herzegovina
adna.asic@gmail.com

ABSTRACT

Autosomal short tandem repeats (STRs) are the most widely used DNA markers in forensic investigation of the population history, human migration patterns, and genealogical research. In this study, the usefulness of 13 most widely used STR loci (D3S1358, TH01, D21S11, D18S51, D5S818, D13S317, D7S820, D16S539, CSF1PO, vWA, D8S1179, TPOX, and FGA) was examined along with the investigation of their application in the studies of the phylogeny of human populations. We compared allele frequencies of STR loci of the populations from the Balkan Peninsula to determine the similarities and differences among them and to determine how informative they are when it comes to the human identity testing. We made UPGMA phylogenetic tree using POPTREE2 software and Nei's table of genetic distances using MEGA5.21 software. Additionally, MDS (multidimensional scaling) plot was generated using SPSS 20.0 software. The results implied that both geographical proximity and shared history are determining the strong clustering of the populations on the Balkans. Another conclusion drawn from this overview is that the studied STR markers are highly polymorphic and thus, satisfyingly informative to be used for human identity testing and phylogenetic research.

Keywords: Balkan Peninsula, autosomal STRs, phylogenetic tree, genetic distance, clustering, population study

INTRODUCTION

Short tandem repeats (STRs), also called microsatellites or simple sequence repeats (SSRs), are the DNA sequences located in the non-coding region of the human genome and consist of 2-7 bp long repetitive units that are repeated 3-15 times, thus making STRs long up to approximately 400 bp (Gunn, 2006, Goodwin, Linacre, Hadi, 2011). They occur on all 22 pairs of autosomal chromosomes, and on X and Y sex chromosomes (Gunn, 2006). Since they are typically located between the genes, they can be of different size among the individuals without affecting the genetic health of the person (Butler, 2010). These differences can be the result of mutations, recombination, and independent chromosomal variation (Gunn, 2006), and it makes STRs effective for the human identification purposes (Butler, 2010). STRs are also used for studying the diversity among different populations, as well as for the determination of similarity between closely related populations (Doğan, Kovačević, Marjanović, 2013). 13 core STR loci that are most widely used are chosen to be the basis of the CODIS national DNA database and they are: CSF1PO, FGA, TH01, TPOX, VWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, and D21S11 (Butler, 2011).

Data in the form of allele frequencies are useful for the studies of the phylogenetic relationships between different species or populations. STRs are the markers of choice in the modern phylogenetic analyses. They are providing genetic distance measures used for the construction of the phylogenetic tree which shows similarities and differences between the populations. Two most widely used methods for the construction of the phylogenetic tree are: neighbor-joining (NJ) and un-weighted pair group method with arithmetic mean (UPGMA). The latter one is used for showing the relationships among closely related populations and usually produces a rooted tree (Takezaki, Nei, 1996).

In the present study, allele frequencies of 13 autosomal STR loci were compared for 14 populations originating from the area of the Balkan Peninsula and the phylogenetic tree was constructed showing the relationships among these populations. The aim of the study was to determine whether autosomal STR loci are informative enough to be used for human identification purposes and in the population studies. Additionally, we wanted to explore the similarities and differences among the populations on the Balkans.

MATERIALS AND METHODS

14 populations originating from the Balkan Peninsula were compared in this study: Bosnian-Herzegovinian, Croatian, Serbian, Montenegrin, Macedonian, Slovenian, Turkish (living in Turkey), Turkish (living in Bosnia and Herzegovina), Albanian (living in Kosovo, Serbia), Albanian (living in North-West Italy), Romanian, Hungarian, Greek, and the population from Vojvodina, Serbia (Figure 1). Table 1 lists all populations along with the references for the articles from which data were taken.
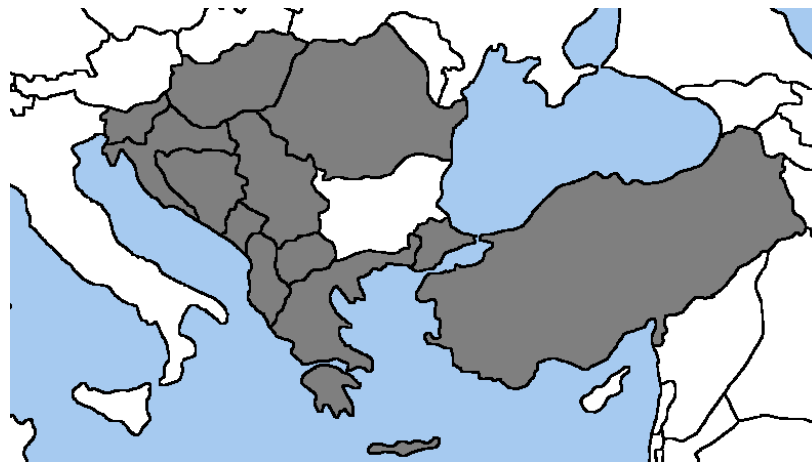
Figure 1: Geographical location of the countries of origin of the populations compared in the study (colored in grey)

Data on the allele frequencies was collected by searching through the journal databases. Only the articles giving the details about the experimental procedure and using the adequate sample size were included in the research. Additionally, studies which did not include all 13 STR loci of interest were excluded from this research.

Table 1: The list of the populations compared in the present study

Population	Reference
Bosnian-Herzegovinian	Marjanović *et al.* (2006)
Croatian	Projić *et al.* (2007)
Serbian	Keckarević *et al.* (2009)
Montenegrin	Veselinović *et al.* (2004)
Macedonian	Havaš *et al.* (2007)
Slovenian	Drobnič *et al.* (2005)
Turkish (Turkey)	Çakir *et al.* (2003)
Turkish (Bosnia and Herzegovina)	Dogan *et al.* (2013)
Albanian (Kosovo, Serbia)	Kubat *et al.* (2004)
Albanian (NW Italy)	Robino *et al.* (2001)
Romanian	Barbarii *et al.* (2004)
Hungarian	Rak *et al.* (2010)
Greek	Kovatsi *et al.* (2006)
Vojvodina, Serbia	Petrić *et al.* (2012)

Phylogenetic tree showing the relationships between the populations was constructed using POPTREE2 program (Takezaki, Nei, Tamura, 2010). The method of choice was UPGMA method since it is the best option for showing the distances between closely related species or populations and is giving the clearest results. The table of Nei's genetic distances was obtained using MEGA5.21 software (Tamura, Peterson, D., Peterson, N., Stecher, Nei, Kumar, 2011). Multidimensional scaling (MDS) analysis and plot generation were done using SPSS 20.0 software package (SPSS, Chicago, IL, USA).

RESULTS AND DISCUSSION

The results of this study are presented in the rectangular (Figure 2) and the circular (Figure 3) UPGMA phylogenetic trees and the Nei's table of genetic distances (Table 2). The results imply that the populations from the Balkan Peninsula are genetically close to each other since all of them clustered closely to each other in the phylogenetic trees. Also, large genetic distances between the populations were not observed in the genetic distance table. All populations from the Western Balkans, except Montenegro, are positioned close to each other which is explained by the fact that they share the common origin and historical background. Although it is on the opposite side of the phylogenetic tree when compared to the other populations from the Western Balkans, Montenegrin population still seems to be very close to these populations according to the Nei's table. General population from Serbia is positioned close to the population from Vojvodina province, which is also expected since these two populations inhabit the same country.
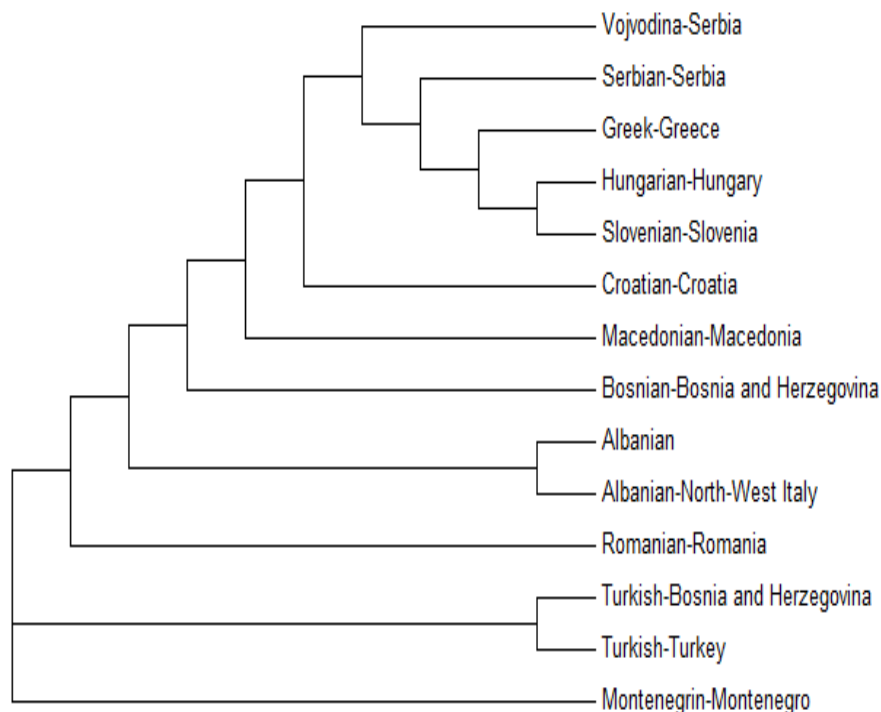

Figure 2: Rectangular UPGMA phylogenetic tree showing the relationships between the populations inhabiting the Balkan Peninsula

The similarity between the Balkan populations observed in this study is confirmed in a paper by Projić *et al.* (2007), where the similar results were obtained with the smaller number of populations from the Balkans. Marjanović *et al.* (2006) did not find any significant differences between Bosnian-Herzegovinian and Croatian populations. Havaš *et al.* (2007) did not find large genetic differences between Macedonians, on one side, and Serbians and Greeks, on the other. In that study, minor differences were found only between Macedonians and Romanians.

Figure 3: Circular UPGMA phylogenetic tree showing the relationships between the populations inhabiting the Balkan Peninsula

Turkish populations living in Turkey and in Bosnia and Herzegovina have clustered close to each other, as well as Albanian populations living in Kosovo, Serbia and in North-West Italy, which was expected since these are the same populations living in different countries. In this way, it is shown that the common origin plays an important role in determining the genetic similarity/difference between populations, apart from their geographical positions.

Table 2: Nei's table of genetic distances for the populations from the Balkan Peninsula

	TUR, B&H	B&H	CRO	SRB	MTN	MAC	ALB	GRE	ALB, IT	HUN	SLO	ROM	TUR
VOJ	0,018	0,013	0,011	0,011	0,016	0,012	0,016	0,010	0,016	0,007	0,007	0,014	0,016
TUR, B&H		0,022	0,016	0,014	0,020	0,014	0,018	0,015	0,018	0,011	0,017	0,019	0,012
B&H			0,013	0,012	0,020	0,017	0,018	0,012	0,018	0,010	0,011	0,018	0,020
CRO				0,010	0,018	0,013	0,015	0,010	0,015	0,008	0,010	0,017	0,015
SRB					0,013	0,011	0,015	0,008	0,015	0,005	0,009	0,015	0,013
MTN						0,017	0,018	0,017	0,018	0,013	0,014	0,023	0,018
MAC							0,014	0,012	0,014	0,009	0,012	0,017	0,013
ALB								0,015	0,000	0,011	0,013	0,017	0,015
GRE									0,015	0,005	0,007	0,014	0,014
ALB, IT										0,011	0,013	0,017	0,015
HUN											0,005	0,011	0,011
SLO												0,013	0,014
ROM													0,019

The results of this study were confirmed with the multidimensional scaling (MDS) plot which shows the distances between the compared populations in the two dimensions (Figure 4).
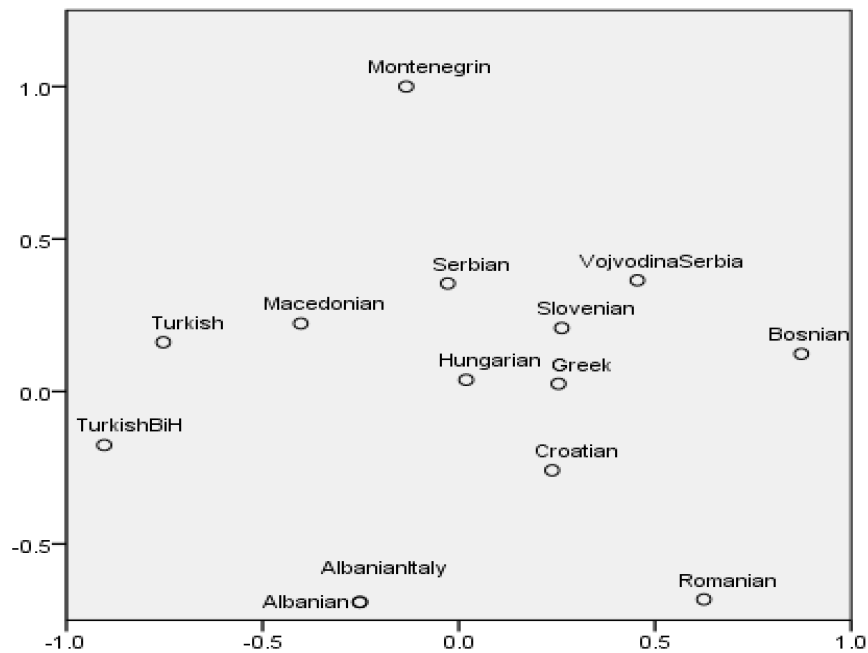

Figure 4: MDS plot which shows the relationships between the populations from the Balkan Peninsula

When it comes to the properties of the individual loci, two most informative loci in the populations from the Balkans are FGA with the average power of discrimination (PD) value of 0.964 and D18S51 with the average PD value of 0.963. On the other hand, the least informative loci are TPOX (average PD value of 0.809) and TH01 (average PD value of 0.921).

CONCLUSION

The general conclusion drawn from this research is that the populations inhabiting the Balkan Peninsula are genetically very close to each other since important genetic differences based on the allele frequencies of 13 autosomal STR loci between the populations were not observed. Additionally, it is concluded that the STR loci tested are informative enough to be used for the purposes of individualization and the population studies.

REFERENCES

Barbarii, L. E., Rolf, B., Constantinescu, C., Hohoff, C., Calistru, P., Dermengiu, D. (2004). Allele Frequencies of 13 Short Tandem Repeat (STR) Loci in the Romanian Population. Forensic Sci. Int; 141(2-3): 171-174.

Butler, J. M. (2010). Fundamentals of Forensic DNA Typing. Elsevier Inc: Burlington, MA, USA.

Butler, J. M. (2011). Advanced Topics in Forensic DNA Typing: Methodology. Elsevier: Waltham, MA, USA.

Doğan, S., Kovačević, L., Marjanović, D. (2013). Genetic Polymorphisms of 15 STR Loci within Turkish Student Population Living in Sarajevo, Bosnia and Herzegovina. Coll. Antropol; 37(4): 1313-1319.

Drobnič, K., Pojskić, N., Bakal, N., Marjanović, D. (2005). Allele Frequencies for the 15 Short Tandem Repeat Loci in Slovenian Population. J. Forensic Sci; 50(6): 1505-1507.

Goodwin, W., Linacre, A., Hadi, S. (2011). An Introduction to Forensic Genetics. John Wiley & Sons, Ltd: Chichester, West Sussex, England.

Gunn, A. (2006). Essential Forensic Biology. John Wiley & Sons, Ltd: Chichester, West Sussex, England.

Havaš, D., Jeran, N., Efremovska, Lj., Đorđević, D., Rudan, P. (2007). Population Genetics of 15 AmpFlSTR Identifiler Loci in Macedonians and Macedonian Romani (Gypsy). Forensic Sci. Int; 173(2007): 220-224.

Keckarević, D., Pavićević, D. S. (2009). Serbian Population Data. University of Belgrade, Faculty of Biology: Personal Communication.

Kovatsi, L., Parsons, T. J., Just, R. S., Irwin. J. A. (2006). Genetic Variation for 15 Autosomal STR Loci (PowerPlex 16) in a Population Sample from Northern Greece. Forensic Sci. Int; 159(1): 61-63.

Kubat, M., Skavić, J., Behluli, I., Nuraj, B., Bekteshi, T., Behluli, M., Klarić Martinović, I., Peričić, M. (2004). Population Genetics of the 15 AmpFlSTR Identifiler Loci in Kosovo Albanians. Int. J. Legal Med; 118(2): 115-118.

Marjanović, D., Bakal, N., Pojskić, N., Kapur, L., Drobnič, K., Primorac, D., Bajrović, K., Hadžiselimović, R. (2006). Allele Frequencies for 15 Short Tandem Repeat Loci in a representative Sample of Bosnians and Herzegovinians. Forensic Sci. Int; 156(2006): 79-81.

Petrić, G., Drašković, D., Bosić, D. Z., Budakov, B., Veselinović, I. (2012). Genetic Variation at 15 Autosomal STR Loci in the Hungarian Population of Vojvodina Province, Republic of Serbia. Forensic Sci. Int. Genetics; 6(2012): 163-165.

Projić, P., Škaro, V., Šamija, I., Pojskić, N., Durmić-Pašić, A., Kovačević, L., Bakal, N., Primorac, D., Marjanović, D. (2007). Allele Frequencies for 15 Short Tandem Repeat Loci in Representative Sample of Croatian Population. Croat. Med. J; 48(4): 473–477.

Rak, S. A., Zalan, A., Szabadosa, G., Pamjav, H. (2010). Population Genetic Data on 15 STR Loci in the Hungarian Population. Forensic Sci. Int. Genetics; 5(5): 543-544.

Robino, C., Gino, S., Torre, C. (2001). Allele Frequencies for the PowerPlex 16 STR Loci in an Albanian Population Sample from Northern Italy. J. Forensic Sci; 46(4): 998-999.

Takezaki, N., Nei, M. (1996). Genetic Distances and Reconstruction of Phylogenetic Trees from Microsatellite DNA. Genetics; 144(1): 389-399.

Takezaki, N., Nei, M., Tamura, K. (2010). POPTREE2: Software for Constructing Population Trees from Allele Frequency Data and Computing Other Population Statistics with Windows Interface. Mol. Biol. Evol; 27(4): 747-752.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S. (2011). MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Mol. Biol. Evol; 28(10): 2731-2739.

Veselinović, I., Kubat, M., Furač, I., Škavić, J., Klarić, I. M., Tasić, M. (2004). Allele Frequencies of the 15 AmpFlSTR Identifiler Loci in the Population of Vojvodina Province, Serbia and Montenegro. Int. J. Legal Med; 118: 184-186.

Çakir, A. H., Çelebioğlu, A., Altunbaş, S., Yardımcı, E. (2003). Allele Frequencies for 15 STR Loci in Van-Ağrı Districts of the Eastern Anatolia Region of Turkey. Forensic Sci. Int; 135: 60-63.