Overview of Human Lineage Genetic Marker Studies in Bosnia and Herzegovina: Y chromosome story

Aldin Pirić[1], Sabahudin Ćordić[1], Lejla Smajlović-Skenderagić[1], Serkan Dogan[1], Damir Marjanović[1,2]

[1]International Burch University, Sarajevo, Bosnia and Herzegovina

[2]Institute for Anthropological Research, Zagreb, Croatia

aldin.piric@stu.ibu.edu.ba

sabahudin.cordic@stu.ibu.edu.ba

l.smajlovic.skenderagic@ibu.edu.ba

serkan.dogan@ibu.edu.ba

damir.marjanovic@ibu.edu.ba

Abstract – **Modern Bosnia and Herzegovina is a state consisting of multiple ethnicities and regions located in the Western Balkan, with a very complex history. The earliest historical findings show that its area was inhabited since the Paleolithic. From that time, this part of Europe, especially the region of the Modern Bosnia and Herzegovina, could be recognized as the crossroad for the different human migration and the meeting point for different cultures, religions and gene pools. Mitochondrial DNA is being used for maternal lineage testing, while the Y chromosome is being used for paternal lineage testing. Therefore, these markers are being referred to as lineage markers. Lineage markers are often used for parental lineage monitoring in population genetics, human genetics, as well as in forensic genetics. The main intention of this paper is to construct a short overview of the Y chromosome studies performed in Bosnia and Herzegovina within the last two decades.**

Keywords - Bosnia and Herzegovina, lineage markers, molecular markers, population genetic studies, Y chromosome

1.   Introduction

Existent archeological artifacts are proving that territory of Bosnia and Herzegovina has been populated since Neolithic [1]. However, some of the archeological findings imply that the first inhabitants settled here in the Paleolithic era [2]. In the early Bronze Age, Indo-European tribes known as the Illyrians settled in the various region of the modern territory of Bosnia and Herzegovina. [3] the tribes were governed by the Romans for more than five centuries [4]. During that time, a lot of the residents of the Roman empire, including Roman soldiers settled down in the region [1].

After the fall of the Roman empire, this area remained a borderline between the Eastern and Western empires which encouraged. various tribes, such the Avars, the Slavs, and others, that massively invade this region. Additionally, two important events, along with several other historical episodes, significantly impacted the structure of B&H human population. The first of those are large migration waves from the North-East (which were extremely intensive during the 6th and 7th centuries) which moved different Gothic, Avar and Slavic clans into the area. The second one was the expansion of the Ottoman Empire into this part of the Balkans in the fifteenth century [5]. All these historical episodes left their imprint on the

population structure of modern B&H inhabitants and created fascinating genetic diversity within. Therefore, it is not surprising that modern B&H population is one of the most genetically studied regional populations, especially by the use of so-called lineage genetic markers.

Unlike autosomal markers, Y-linked and mitochondrial markers do not undergo each generation shuffling, but instead get passed down from one generation to the next, with the only differences being induced by mutations. For these reasons, these markers are often used for parental lineage monitoring in population genetics, human genetics, as well as in forensic genetics. Mitochondrial DNA is being used for maternal lineage testing, while the Y chromosome is being used for paternal lineage testing. Therefore, these markers are being referred to as lineage markers [6].

Previously published papers presented a short historical overview of earlier published human population studies in Bosnia and Herzegovina, conducted within the last three centuries [7,8]. However, usage of the lineage markers within those papers was just briefly noted. Expansion of human population studies based on these genetic markers, as well as the significance of the obtained results, initiated us to put more attention on this part of BH population genetics. Therefore, this paper will extensively elaborate usage of the Y chromosome DNA markers within analysis of the BH human genetic pool, including the most recent data published after previously mentioned papers.

2. Human Y Chromosome as Genetic Marker

Y chromosome has been given many different definitions, some of them being "nonrecombining desert" and "gene-poor chromosome". Compared to other chromosomes, the Y chromosome has low number of genes with half of its sequence consisting of repeated elements. Moreover, it lacks the recombination ability and is in continuous decay. The Y chromosome is inherited through the patrilineal inheritance pattern, i.e., from father to son, meaning that each male person from the same patrilineal lineage would have an identical profile. The relatively small degree of molecular diversity between markers located on this chromosome comes from the absence of gene recombination in 95% of its length and the mechanism of random mutations as the only possible source of polymorphisms [6].

Denver convention criteria classifies the human Y chromosome as G chromosome, that is, the category of the shortest chromosomes in the human set, consisting also of chromosomes 21 and 22. It contains about 50 million base pairs, which makes out around 1.8% of the total human genome. The Y chromosome contains important information used in determining the parental lineage of a specific male. This is possible because the Y chromosome contains highly polymorphic regions. The human Y chromosome is present in a sole copy in normal males, inherited from the father, and, as already mentioned, 95% of its complex does not undergo recombination. Only 5% of this chromosome has the potential ability to interact with the X chromosome, and the interacting region is called the pseudoautosomal region of the Y chromosome [6].

The Y chromosome has an important role in forensic analyses in cases of rape of women, in particular, those involving more than one man, especially in cases of mixed samples when there is an overwhelming amount of female DNA. Y-STR (Short Tandem Repeat) and Y-SNP (Single Nucleotide Polymorphism) markers are useful in cases of parenthood testing or further kinship through the male line, when the children

are male, and in the process of identification when only kin from the father's line is known. In addition, the Y chromosome is more and more being used in human migration studies due to its property of not undergoing recombination throughout the transfer of genetic material between generations [9].

Actually, since the first Y chromosome polymorphism was published [10], an entire decade has passed before the binary, and later STR markers, located on the NRY region found their wider application in phylogenetic studies monitoring human migration patterns, through the construction of phylogenetic trees [11]. The SNP patterns can be used to determine lineages which are referred to as haplogroups. Haplogroups can also be inferred from readily available Y-STR genotyping data. Vast amount of forensic Y-STR data is available for the use in population genetic studies [12].

3. Overview of the Y Chromosome Population Genetic Studies in Recent B&H Inhabitants

The analysis of STR and SNP variation, autosomal, and Y-chromosome markers were studied so that molecular genetic diversity of B&H could get incorporated into regional and European frames, but also to provide necessary reference for statistical calculations used in forensic genetics. In order to ensure the most relevant calculation, the data are still periodically updated.

Initial results were obtained by observing 28 Y-chromosome biallelic markers in the B&H population [13]. This study was constructed on the ground of regional data and designed to include 256 male individuals. The results showed extremely close genetic relationship between three populations (three main Bosnian and Herzegovinian ethnic groups) and their close relationship to other populations in the Balkans. Of course, further elaboration of this issue required additional studies with a multidisciplinary approach, application of additional molecular markers, expansion of the sample and structural investigation of each ethnic group, as well as the analysis of ancient genetic material from the archeological skeletal samples.

In the same year (2005) very first Y STR population data set for the BH human population was published [14]. Hundred tested males have been voluntary donors. The PowerPlex®Y System has been used in order to amplify 12 Y-STR loci by via PCR. These STR loci are: DYS19, DYS385a, DYS385b, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438 and DYS439. From eighty-one detected Y-STR haplotypes (from a total number of 100 obtained samples) 69 were unique, 7 appeared two times, 4 appeared three and only 1 five times. Statistical analysis incorporated: gene diversity, major allele frequency, the most frequent haplotypes, allele frequency distribution and observed haplotype diversity [3] for 12 PowerPlex®Y loci.

Four years later, with the intent to improve existing database and to obtain more specific results for local populations for a variety of DNA markers, group of authors decided to analyze additional individuals from Canton Sarajevo area. Estimation of genetic diversity at 12 Y-chromosomal STR loci included in the PowerPlex® Y System was used to extend the existing database, and create a more realistic view of the state of the genetic structure within regional Bosnian and Herzegovinian human population, in particular regarding the diversity among the isolated and non-isolated local populations. In addition, the intent of that study was to estimate genetic distinctiveness of the Canton Sarajevo population within the general B&H population as well as with populations of geographically neighboring countries.

Y-STR haplotypes were generated for a sample of 100 unrelated, healthy male individuals living in Canton Sarajevo (Bosnia and Herzegovina) using PowerPlex®Y System kit [15]. Within this pool, the totals of 81 different haplotypes were detected with 71 of them unique. Absolute frequencies of the remaining 10 haplotypes were two for six haplotypes, three for two haplotypes, five for one haplotype and six for one haplotype. Obtained results suggested that the local population of Canton Sarajevo, with respect to the detected haplotype and gene diversity, may be considered a projection of general B&H population. Since this population represents the largest regional population in Bosnia and Herzegovina with emphasized migration influx this is quite a logical outcome.

Four years later, in 2013, Y chromosome diversity of the B&H population was examined again, but with the increased number of STR loci. The sampling was performed using buccal swabs from unrelated, healthy men originating from all regions of Bosnia and Herzegovina. Total number of samples obtained was 100. DNA samples were typed for 23 Y STR loci, with 6 new loci: DYS481, DYS533, DYS576, DYS549, DYS643, and DYS570, which are included in the new PowerPlex® Y 23 amplification kit. The absolute frequency of generated haplotypes was calculated, and results showed that only two samples shared the identical Y 23 haplotype. DYS418 was identified as the most polymorphic locus, with 14 detected alleles and the minimum polymorphic loci were DYS437, DYS389I, DYS393, and DYS391. Decreasing the number of repeating haplotypes is very important in forensic DNA analysis, and this study showed that it can be achieved by increasing the number of highly polymorphic Y STR markers [16].

Whit Athey's Haplogroup Predictor was used to determine Y chromosome haplogroup frequencies via Y chromosome marker frequencies from the same 100 individuals [17]. According to those results, the most frequent haplogroup seems to be I2a, with a commonness of 49%, followed by R1a and E1b1b, each accounting for 17% of all haplogroups within the population. Remaining haplogroups encountered in this study are J2a (5%), I1 (4%), R1b (4%), J2b (2%), G2a (1%) and N (1%). Preliminary B&H population data published before 10 years was confirmed by these results. The prediction about B&H population as a part of the Western Balkan area, which served as the Last Maximum refuge for the Paleolithic human European population was also confirmed in this paper. Furthermore, these results corroborated the hypothesis that this region was an important stopping point on the "Middle East-Europe highway" during the Neolithic farmer migrations. Finally, since these results were almost completely in accordance with previously published data on B&H and neighboring populations that were generated by Y chromosome single nucleotide polymorphism (Y-SNP) analysis, it was concluded that in silico analysis of YSTRs is a reliable method for approximation of the Y chromosome haplogroup diversity of an examined population.

In the meantime, the same STR set of loci was employed to explore the distribution and polymorphisms of 23 short tandem repeat (STR) loci on the Y chromosome in the Turkish population recently settled in Sarajevo, Bosnia and Herzegovina and to investigate its genetic relationships with the homeland Turkish population and neighboring populations [18]. This study included 100 healthy unrelated male individuals from the Turkish population living in Sarajevo. Amplification was performed using PowerPlex Y 23 amplification kit. The studied population was compared to other populations using pairwise genetic distances, which were represented with a multi-dimensional scaling plot. Haplotype and allele frequencies of the sample population were calculated and the results showed that all 100 samples had unique haplotypes. The most polymorphic locus was DYS458, and the least polymorphic DYS391. The observed haplotype

diversity was 1.0000 ± 0.0014, with a discrimination capacity of 1.00 and the match probability of 0.01. Rst values showed that the observed population was closely related in both dimensions to the Lebanese and Iraqi populations, while it was more distant from Bosnian, Croatian, and Macedonian populations. At the end, the conclusion is that Turkish population living in Sarajevo can be observed as a representative Turkish population because results were the same as those published for the population from Turkey. This study showed that populations which are geographically close, were related genetically to each other.

The methods for haplogroup prediction were encountered in this study [19]. 23 loci from previously obtained Y-STR haplotypes from 100 unrelated healthy Turkish males, who had recently settled in Sarajevo, were utilized for the purpose of determining the haplogroups via Whit Athey's Haplogroup Predictor software. In total 90 studied haplotypes had the Bayesian probability greater than 92.2 % and had the range between 51.4% and 84.3% for the 10 haplotypes left. 17 differently distributed haplogroups were found, with Y-haplogroup J2a being the most prevalent one, with abundance percentage of 26% of all samples, while haplogroups R1b, G2a, and R1a were less prevalent, with the range from 10% to 15% of all the samples. These 4 haplogroups together contribute to 63% of all Y-chromosomes. in total 11 haplogroups (E1b1b, G1, I1, I2a, I2b, J1, J2b, L, Q, R2, and T) had a range from 2% to 5%, whereas other haplogroups, namely E1b1a and N were found in only 1% of all samples. Results have shown that a large percentage of the Turkish paternal line is linked with West Asia, Europe Caucasus, Western Europe, Northeast Europe, Middle East, Russia, Anatolia, and Black Sea Y chromosome lineages. Conclusion is that the analyzed Turkish population can serve as a representative sample for the Turkish population residing in Turkey, because results were consistent with those data published earlier in the literature for Turkish population in Turkey.

In years 2016 and 2017, similar studies were performed on the human population residing in Tuzla, Bosnia and Herzegovina. Namely, Tuzla Canton is one of the most populated regions in Bosnia and Herzegovina, thus its genetic analysis could serve as proof of past demographic events. The first study, which included in total 100 unrelated healthy adult males genotyped using 23-Y STR loci included within PowerPlex Y23 kit [20], employed statistical tests such as haplotype diversity, allele frequencies and Rst-based genetic distances calculated between new dataset and the one from Bosnia and Herzegovina and other places. The distances were afterwards visualized through multidimensional scaling plot and neighbor-joining phylogenetic tree analyses. Discrimination capacity of the PowerPlex Y23 kit appeared to be high, because all 100 individuals had the unique haplotypes, and newly incorporated loci seem very informative. However, no significant difference between the study population and the general population of Bosnia and Herzegovina, as well as between the population of Tuzla and neighboring populations. [20]

In the Second study, for the same 100 unrelated male individuals from Tuzla Canton, Bosnia and Herzegovina (B&H) in silico haplogroup assignments were made and it was based on 23-loci Y-STR data using the following four different algorithms [21]. Dominant haplogroups were I, R and E with their sublineages I2a, R1a, and E1b1b. It is in connection with the published Y-SNP data for the B&H population. In general, results which are represented in this study did not only constitute a concordance study on the four haplogroup assignment algorithms which are also most popular, but they also give a deep knowledge about differentiation that can be find within population of B&H based on Y haplogroups for the first time.

Those studies-initiated publication of the few more papers which were including Y STR data from B&H human population. The first one was published in 2015 and it was focused on the clustering of the European human population based on the Y-STR data [22]. Three overall clusters were formed as a result of autosomal STR loci analyses, namely the European, Asian and African. However, Y-STR analyses highlighted formations of new sub-clusters. This is confirmed since the European cluster was easily divided into four distinct groups represented as four branches of the phylogenetic tree, while the Asian population cluster consists of two sub-clusters. Given the aforementioned clustering trends evident in both phylogenetic trees, it was concluded that clusters were indeed formed as a consequence of geographical proximity that triggered a mixing of gene pools, which in turn resulted in the formation of neighboring populations that exhibit strong genetic similarities. Overall, this study effectively highlights that Y-STRs could be a more informative tool in structural population studies as they are more informative than autosomal STRs because they not only enable continental clustering but are also a great tool for additional regional studies as well. Formation of four sub-clusters of European populations is once again proving the great potential of Y-chromosomal markers in the wide spectrum of genetic analyses.

The second one was published in 2018 and it was focused on the analysis of the Balkan human population based on the Y-STR data [12]. This study aimed to provide insight into genetics relations in Balkan population using silico analysis of Y-STR haplotypes and predicting haplogroups as well as doing network analysis of the same haplotypes. The population dataset was obtained using 23, 17, 12, 9 and 7 Y-STR loci for 13 populations, including Bosnia and Herzegovina (B&H), Croatia, Slovenia, Greece, Macedonia, Romany (Hungary), Hungary, Serbia, Montenegro, Albania, Kosovo, Romania and Bulgaria. The overall dataset consists of 2179 samples with 1878 different haplotypes. Between thirteen analyzed Balkan populations, in four of them 12a was recognized as the major haplogroup. Each population with 12a as the major haplogroup (B&H, Croatia, Montenegro and Serbia) was from the former Yugoslavia republic. The last two major populations from Yugoslavia, Macedonia and Slovenia, had E1b1b and R1a haplogroups as the most prevalent. E1b1b haplogroup was the most prevalent in the population of Macedonia, Romania, as well as Albania and Kosovo. Comparing I2a haplogroup clusters to E1b1b and R1b haplogroup clusters, the former one is more compact, which indicates a larger degree of homogeneity within the haplotypes that belong to that haplogroup. This study indicates that an effective approach for utilization of publicly available Y-STR datasets may lie in combination of haplogroup prediction and network analysis.

4. Conclusion

Describing something that lasts for two decades as "a beginning" is quite unusual. However, that is the truth in the case of Y chromosome human population-genetic studies in Bosnia and Herzegovina. There are still many interesting features hidden within the existent diversity of local human populations in this small, but intriguing, country that are still waiting to be discovered and described. Several preliminary hypotheses were completely changed, such us origin of R1b haplogroup within this region, or significantly questioned, such us origin of notably high frequency of I2a haplogroup in Bosnia (as Balkan LGM refugium marker or

"Slavic migration marker" increased by founder effect) [23]. Those, and many other Y chromosome stories are just waiting to be told.

REFERENCES

[1]   Malcolm, N. (1996). Bosnia: A short history. NYU Press.

[2]   Hadžibegović, I., & Imamović, M. (1994). Bosna i Hercegovina od najstarijih vremena do kraja Drugog svjetskog rata.

[3]   Wilkes, J. (1995). The Illyrians. Wiley-Blackwell.

[4]   Klaić, V. (1990). Povijest Bosne do propasti kraljevstva. Svjetlost Sarajevo.

[5]   Marjanovic D., et al. Doc Praehistorica, 23 (2006) 21.-6.

[6]   Marjanović, D., et al. (2018). Forensic genetics: Theory and application.

[7]   Marjanović, D., Pojskić, N., Kapur, L., Haverić, S., Durmić-Pašić, A., Bajrović, K., & Hadžiselimović, R. (2008). Overview of human population-genetic studies in Bosnia and Herzegovina during the last three centuries: history and prospective. Collegium antropologicum, 32(3), 981-987.

[8]   Lasić, L. (2016). Historical Overview of the Human Population-Genetic Studies in Bosnia and Herzegovina: Small Country, Great Diversity. Collegium antropologicum, 40(2), 145-149.

[9]   Semino, O., Passarino, G., Oefner, P. J., Lin, A. A., Arbuzova, S., Beckman, L. E., ... & Marcikiæ, M. (2000). The genetic legacy of Paleolithic Homo sapiens sapiens in extant Europeans: AY chromosome perspective. Science, 290(5494), 1155-1159.

[10]   Casanova, M., Leroy, P., Boucekkine, C., Weissenbach, J., Bishop, C., Fellous, M., ... & Siniscalco, M. (1985). A human Y-linked DNA polymorphism and its potential for estimating genetic and evolutionary distance. Science, 230(4732), 1403-1406.

[11]   Underhill, P. A., Myres, N. M., Rootsi, S., Metspalu, M., Zhivotovsky, L. A., King, R. J., ... & Kutuev, I. (2010). Separating the post-Glacial coancestry of European and Asian Y chromosomes within haplogroup R1a. European Journal of Human Genetics, 18(4), 479.

[12]   Šehović, E., Zieger, M., Spahić, L., Marjanović, D., & Dogan, S. (2018). A glance of genetic relations in the Balkan populations utilizing network analysis based on in silico assigned Y-DNA haplogroups. AnthropologicAl review, 81(3), 252-268.

[13]   Marjanovic, D., Fornarino, S., Montagna, S., Primorac, D., Hadziselimovic, R., Vidovic, S., ... & Andjelinovic, S. (2005). The peopling of modern Bosnia-Herzegovina: Y-chromosome haplogroups in the three main ethnic groups. Annals of Human Genetics, 69(6), 757-763.

[14]   Marjanovic, D., Bakal, N., Pojskic, N., Kapur, L., Drobnic, K., Primorac, D., ... & Hadziselimovic, R. (2005). Population data for the twelve Y-chromosome short tandem repeat loci from the sample of multinational population in Bosnia and Herzegovina. Journal of Forensic Science, 50(1), JFS2004289-2.

[15]   Ćenanović, M., Pojskić, N., Kovačević, L., Džehverović, M., Čakar, J., Musemić, D., & Marjanović, D. (2010). Diversity of Y-short tandem repeats in the representative sample of the population of Canton Sarajevo residents, Bosnia and Herzegovina. Collegium antropologicum, 34(2), 545-550.

[16]   Kovačević, L., Fatur-Cerić, V., Hadžić, N., Čakar, J., Primorac, D., & Marjanović, D. (2013). Haplotype data for 23 Y-chromosome markers in a reference sample from Bosnia and Herzegovina. Croatian medical journal, 54(3), 286-290.

[17]   Doğan, S., Ašić, A., Doğan, G., Besic, L., & Marjanovic, D. (2016). Y-Chromosome Haplogroups in the Bosnian-Herzegovinian Population Based on 23 Y-STR Loci. Human biology, 88(3), 201-210.

[18]   Dogan, S., Primorac, D., & Marjanović, D. (2014). Genetic analysis of haplotype data for 23 Y-chromosome short tandem repeat loci in the Turkish population recently settled in Sarajevo, Bosnia and Herzegovina. Croatian medical journal, 55(5), 530.

[19]   Doğan, S., Doğan, G., Ašić, A., Bešić, L., Klimenta, B., Hukić, M., ... & Marjanović, D. (2016). Prediction of the Y-Chromosome Haplogroups within a recently settled Turkish Population in Sarajevo, Bosnia & Herzegovina. Collegium antropologicum, 40(1), 1-7.

[20]   Babić, N., Dogan, S., Čakar, J., Pilav, A., Marjanović, D., & Hadžiavdić, V. (2017). Molecular diversity of 23 Y-chromosome short tandem repeat loci in the population of Tuzla Canton, Bosnia and Herzegovina. Annals of human biology, 44(5), 419-426.

[21]   Dogan, S., Babic, N., Gurkan, C., Goksu, A., Marjanovic, D., & Hadziavdic, V. (2016). Y-chromosomal haplogroup distribution in the Tuzla Canton of Bosnia and Herzegovina: A concordance study using four different in silico assignment algorithms based on Y-STR data. Homo, 67(6), 471-483.

[22]   Dogan, S., Ašić, A., Buljubašić, S., Bešić, L., Avdić, M., Ferić, E., ... & Marjanović, D. (2015). Overview of European population clustering based on 23 Y-STR loci. Genetika, 47, 901-908.

[23]   Primorac, D., Marjanović, D., Rudan, P., Villems, R., & Underhill, P. A. (2011). Croatian genetic heritage: Y-chromosome story. Croatian medical journal, 52(3), 225-234.