4-4-2017

Y chromosome haplogroups in the Bosnian-Herzegovinian population based on 23 Y-STR loci

Serkan Dogan
International Burch University, Department of Genetics and Bioengineering, serkan.dogan@ibu.edu.ba

Adna Ašić
International Burch University, Department of Genetics and Bioengineering, adna.asic@ibu.edu.ba

Gulsen Dogan
International Burch University, Department of Genetics and Bioengineering, gulsendoganbosna@gmail.com

Larisa Besic
International Burch University, Department of Genetics and Bioengineering, larisa.besic@ibu.edu.ba

Damir Marjanovic
International Burch University, Department of Genetics and Bioengineering, damir.marjanovic@ibu.edu.ba

Y chromosome haplogroups in the Bosnian-Herzegovinian population based on 23 Y-STR loci

Serkan Dogan[1], Adna Ašić[1], Gulsen Dogan[1], Larisa Besic[1], Damir Marjanovic[1]

[1]*International Burch University, Department of Genetics and Bioengineering*.

Corresponding author: Serkan Doğan, International Burch University, Department of Genetics and Bioengineering, Francuske revolucije bb, Ilidža, 71210, Sarajevo, Bosnia and Herzegovina. E-mail: serkan.dogan@ibu.edu.ba

Keywords: Y chromosome, Y-STRs, Y haplogroups, Whit Athey's haplogroup predictor, Bosnian-Herzegovinian population

Running header: Y chromosome haplogroups based on 23 Y-STR loci

Issue 88.3

Abstract

In a study of the Bosnian-Herzegovinian (B&H) population, Y chromosome marker frequencies for 100 individuals, generated using PowerPlex® Y23 kit, were used to perform Y chromosome haplogroup assignment via Whit Athey's Haplogroup Predictor. This Whit Athey's algorithm determines Y chromosome haplogroups from Y chromosome short tandem repeat (Y-STR) data using Bayesian probability-based approach. According to the results of the present study, the most frequent haplogroup appears to be I2a, with a prevalence 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%). These results confirm previously published preliminary B&H population data published over 10 years ago, especially the prediction about B&H population being a part of the Western Balkan area, which served as the Last Glacial Maximum refuge for the Paleolithic human European population. Furthermore, the results corroborate the hypothesis that this area was a significant stopping point on the "Middle East-Europe highway" during the Neolithic farmer migrations. Finally, since these results are 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 can be concluded that *in silico* analysis of Y-STRs is a reliable method for approximation of the Y chromosome haplogroup diversity of an examined population.

Introduction

The part of the human Y chromosome that is sex-specific is haploid and transmitted from father to son. It lacks recombination, which makes its DNA sequence distinctive in studying modern human evolution (Hughes and Rozen 2012; Larmuseau et al. 2014). Therefore, polymorphisms on the Y chromosome have been used in the investigations of migration patterns of male ancestors, ranging from recent history to the origins of modern humans (Petrejčikova et al. 2009; Wang et al. 2015).

There are two methods frequently used to determine Y chromosome haplogroups in a given population, namely the allele-frequency-goodness-of-fit and the Bayesian approach. The latter results in the probability of a Y-STR haplotype being found within a haplogroup. Whit Athey's Haplogroup Predictor is based on the application of Bayesian statistics in order to predict Y chromosome haplogroups from Y-STR data (Athey 2005; Athey 2006). This approach is less expensive and labor-intensive when compared to haplogroup determination based on Y-SNP analysis, which requires DNA typing and polymorphism detection using either capillary electrophoresis or DNA sequencing.

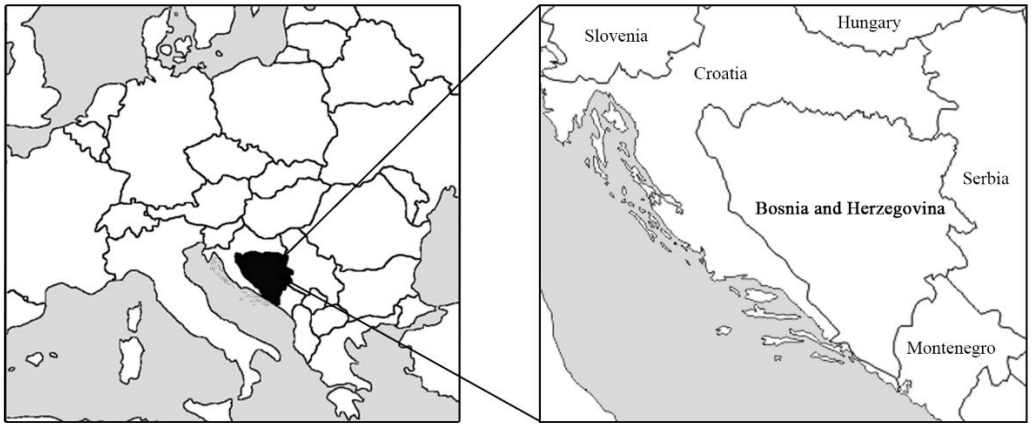

Figure 1. Position of Bosnia and Herzegovina in Europe. Left: Country territory is represented in black. Right: Enlarged image of Bosnia and Herzegovina and neighboring countries.

Bosnia and Herzegovina is a country located on the Balkan Peninsula between Croatia on the west, Serbia on the east and Montenegro on the south (Figure 1). It is a multinational and multireligious country with three main ethnic groups (Bosniaks, Croats and Serbs), and it has been a subject of massive migrations and numerous war activities in the previous century, all of which have had a huge impact on the structure of its population (Marjanović et al. 2005). A

previous population genetics study conducted by Marjanović et al. (2005) on 256 samples of a representative population from Bosnia and Herzegovina is so far the only available report about Y chromosome haplogroups in the B&H population, where a limited number of Y-SNPs were genotyped in order to determine the most prevalent haplogroups.

Due to the continuous migrations in this area, as well as the fact that the last data update was published more than 10 years ago, there is most certainly an imperative for a revision of Y chromosome haplogroup distribution in B&H population using a novel approach. The current study has been conducted in order to test the hypothesis that accurate haplogroup assignment from Y-STR data can be achieved via Bayesian statistics, as well as to update the image of Y chromosome haplogroup distribution in Bosnia and Herzegovina 10 years after the initial preliminary study (Marjanović et al. 2005). *In silico* analysis was used in order to determine whether a much less expensive and less labor-intensive method could be as reliable as conventional experimental methods.

Materials and Methods

In order to perform *in silico* assignment of Y chromosome haplogroups for the B&H population, data from 94 individuals based on 23 Y-STR loci contained in the PowerPlex® Y23 System (Promega Corporation, Madison, WI, USA) was used (Kovačević et al. 2013). Additional six samples from Tuzla Canton inhabitants were used for a total sample number of 100. The following Y-STR markers were analyzed in the present study: DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS385a/b, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS635, GATA-H4, DYS481, DYS533, DYS549, DYS570, DYS576, and DYS643 (Dogan et al. 2015). Assignment of Y chromosome haplogroups based on Y-STR values was done using Whit Athey's Haplogroup Predictor v.5, which uses Bayesian-allele-frequency approach (Athey 2006). The input data consisted of the twenty-three Y-STR markers, while the output of the algorithm was represented by the Bayesian probability for each haplogroup. The haplogroup nomenclature used in this study is in accordance with the recommendations of The Y Chromosome Consortium (The Y Chromosome Consortium 2002).

Results

Y chromosome haplogroup distribution was predicted for B&H population according to already existing PowerPlex® Y23 STR profiles (Table 1). A successful haplogroup assignment was established for all the 100 Y-STR profiles, and the prediction accuracy was estimated to be 100% in 97 cases. A total of nine different Y-chromosomal haplogroups were detected.

Table 1. Y chromosomal haplogroup assignments for each haplotype in Bosnian-Herzegovinian population using Whit Athey's algorithm.

	393	*390*	*19*	*391*	*385a*	*385b*	*439*	*389I*	*392*	*389II*	*458*	*437*	*448*	*H4*	*456*	*576*	*570*	*438*	*481*	*549*	*533*	*635*	*643*	*Assignment*	*Probability*
ID 1	13	24	16	11	14	15	12	13	11	31	18	15	19	11	15	18	19	10	30	11	12	23	10	I2a	100
ID 2	13	24	16	11	14	15	12	13	11	30	15	15	19	11	15	18	18	10	30	11	12	23	10	I2a	100
ID 3	12	24	15	11	12	15	12	13	13	29	18	14	19	11	16	18	17	12	23	14	12	23	10	R1b	100
ID 4	13	24	13	9	17	18	12	13	11	30	16	14	20	12	16	17	19	10	23	14	12	21	12	E1b1b	100
ID 5	13	24	15	11	14	15	13	12	11	30	18	15	19	11	15	19	18	10	30	11	12	21	10	I2a	100
ID 6	13	24	15	10	11	11	11	14	11	30	15	14	20	12	17	17	21	11	25	12	11	23	10	R1a	100
ID 7	13	24	16	11	14	15	13	13	11	30	18	15	19	11	15	18	18	10	30	11	13	23	10	I2a	100
ID 8	13	25	16	10	11	14	11	13	11	29	16	14	20	12	19	18	19	11	25	12	12	23	10	R1a	100
ID 9	13	24	16	10	14	15	13	13	11	31	21	15	20	11	15	21	17	10	30	11	13	24	10	I2a	100
ID 10	13	24	15	10	14	15	13	13	11	32	17	15	20	11	15	19	18	10	28	11	12	22	10	I2a	100
ID 11	13	22	15	10	12	14	11	12	11	29	14	16	20	11	15	16	21	10	24	12	11	20	12	I1	100
ID 12	13	24	15	11	14	15	13	14	11	32	16	15	19	11	15	18	18	10	30	11	12	24	11	I2a	100
ID 13	13	24	15	11	15	15	12	13	11	31	16	15	19	11	15	17	18	11	29	11	12	23	10	I2a	99.9
ID 14	13	24	16	11	14	15	14	13	11	31	18	15	19	11	15	19	19	10	31	11	12	23	10	I2a	100
ID 15	13	24	16	11	14	15	12	14	11	33	18	15	18	11	15	17	18	10	30	11	13	23	10	I2a	100
ID 16	13	25	13	10	16	18	12	14	11	31	15	14	20	12	16	17	19	10	22	12	11	22	12	E1b1b	100
ID 17	13	24	15	11	14	15	13	13	11	31	19	15	19	11	16	19	18	10	31	11	12	22	10	I2a	100
ID 18	13	24	13	9	13	14	10	14	11	30	18	14	20	12	16	19	22	10	27	11	11	21	12	E1b1b	100
ID 19	13	25	13	10	16	18	11	13	11	30	15	14	20	12	17	18	20	10	22	12	12	23	12	E1b1b	100
ID 20	13	24	16	11	14	15	13	13	11	31	17	15	19	12	15	18	18	10	31	11	13	23	10	I2a	100
ID 21	13	24	16	10	14	15	13	13	11	31	17	15	18	11	16	18	18	10	30	11	13	22	10	I2a	100
ID 22	13	24	16	11	14	15	13	13	11	32	17	15	19	11	15	19	19	10	31	12	13	22	10	I2a	100
ID 23	13	24	16	10	13	15	12	13	11	31	17	15	19	11	15	18	17	10	30	11	13	22	10	I2a	100
ID 24	13	24	17	11	14	15	13	13	11	31	17	15	19	11	15	20	19	10	31	11	12	23	10	I2a	100
ID 25	13	24	16	11	14	15	13	13	11	31	18	15	19	11	16	18	19	10	27	11	12	23	9	I2a	100
ID 26	13	24	16	10	14	16	12	12	11	30	17	15	19	11	15	18	19	10	31	11	12	23	10	I2a	100
ID 27	13	25	16	11	11	14	10	13	11	31	15	14	20	13	15	18	19	11	23	12	12	24	10	R1a	100
ID 28	13	25	16	11	15	16	12	13	11	30	18	15	19	11	15	20	18	10	28	11	12	23	10	I2a	100
ID 29	13	25	15	11	14	16	13	13	11	31	18	15	18	11	15	19	18	10	31	11	11	24	10	I2a	100
ID 30	13	25	13	10	16	18	11	13	11	30	15	14	20	12	17	17	19	10	22	12	12	23	12	E1b1b	100
ID 31	13	24	16	11	14	15	13	13	11	30	18	15	19	11	15	18	18	10	31	11	12	22	10	I2a	100
ID 32	13	24	15	11	14	16	12	13	11	31	17	15	19	11	15	19	18	10	31	11	12	22	10	I2a	100
ID 33	13	24	16	11	14	15	12	13	11	30	17	15	20	11	15	18	19	10	31	11	12	22	10	I2a	100
ID 34	13	24	14	11	11	11	12	13	13	29	15	15	19	12	15	17	18	12	23	14	12	23	10	R1b	100
ID 35	13	25	17	10	11	14	10	13	11	30	15	14	20	13	16	19	18	10	23	13	12	21	10	R1a	100
ID 36	13	24	13	10	16	18	12	13	11	30	15	14	20	11	16	17	19	10	23	13	12	22	12	E1b1b	100
ID 37	13	23	14	10	13	14	12	12	11	28	15	16	20	12	14	17	21	10	25	12	11	22	11	I1	100
ID 38	12	23	15	10	15	19	12	12	12	28	16	14	19	12	14	17	17	9	23	12	11	21	9	J2b	100
ID 39	13	24	16	11	13	15	12	13	11	31	17	15	19	11	14	19	18	10	31	11	13	22	11	I2a	100
ID 40	13	23	15	11	14	15	12	14	11	31	17	15	18	11	15	18	18	10	31	11	13	23	10	I2a	100
ID 41	13	24	16	11	14	14	13	13	11	32	17	15	21	11	15	17	19	10	29	11	12	23	10	I2a	100
ID 42	13	25	13	10	19	19	12	13	11	30	15	14	20	12	15	20	20	10	22	12	12	21	12	E1b1b	100
ID 43	13	24	13	10	16	18	9	13	11	30	17	14	19	12	16	18	19	10	22	12	12	21	12	E1b1b	100
ID 44	13	25	17	11	11	13	11	13	11	30	15	14	20	15	16	18	19	11	23	12	12	23	9	R1a	100
ID 45	13	25	15	11	14	15	12	13	11	30	18	15	19	11	15	17	20	10	33	11	13	23	10	I2a	100
ID 46	13	24	15	11	14	15	14	13	11	31	18	15	20	11	15	18	18	10	30	11	13	22	10	I2a	100
ID 47	13	24	14	11	14	15	13	13	11	30	17	15	19	12	15	19	18	10	30	11	13	22	10	I2a	100
ID 48	13	24	13	10	16	19	11	13	11	30	15	14	20	12	17	17	20	10	22	12	12	20	12	E1b1b	100
ID 49	13	25	15	10	12	14	11	14	11	30	18	15	20	12	17	18	20	11	26	12	12	23	10	R1a	100
ID 50	13	24	12	10	16	18	11	14	11	31	17	14	20	11	16	19	20	10	22	12	12	22	12	E1b1b	100
ID 51	13	24	13	11	14	15	14	13	11	31	17	15	19	11	15	17	18	10	29	11	12	23	10	I2a	100
ID 52	13	24	13	10	16	17	12	14	11	32	17	14	20	12	16	19	18	10	23	12	12	22	12	E1b1b	100
ID 53	12	24	15	10	14	17	14	12	11	29	16	16	19	11	13	16	20	9	23	13	10	21	9	J2b	100
ID 54	13	23	14	10	13	17	12	13	11	31	14	15	21	11	14	18	18	9	22	12	11	22	10	J2a	100
ID 55	13	24	16	11	11	15	10	12	11	28	15	14	20	13	15	17	21	11	23	12	12	23	10	R1a	100
ID 56	13	25	15	10	11	14	10	13	11	30	15	14	20	12	16	21	17	11	22	12	12	23	10	R1a	100
ID 57	13	24	14	11	11	11	12	13	13	29	16	15	19	12	15	18	17	12	22	13	12	23	10	R1b	100
ID 58	13	24	13	10	16	19	11	13	11	30	15	14	20	12	16	18	19	10	22	12	12	20	12	E1b1b	100
ID 59	12	22	16	10	13	15	12	13	11	32	15	14	21	12	15	17	17	9	22	12	11	20	8	J2a	100
ID 60	13	24	16	11	14	15	12	13	11	31	17	15	20	11	15	18	18	10	31	12	12	22	10	I2a	100
ID 61	13	23	14	10	13	14	11	12	11	28	15	15	20	11	14	16	19	10	27	12	11	22	13	I1	100
ID 62	13	24	13	10	16	18	12	13	11	31	16	14	21	12	18	16	21	10	22	12	12	22	13	E1b1b	100
ID 63	13	24	16	11	14	15	12	13	11	31	17	15	19	11	15	18	18	10	31	11	12	23	10	I2a	100
ID 64	13	22	15	10	14	14	12	12	11	29	16	16	21	11	15	17	18	10	21	11	9	20	12	G2a	100
ID 65	13	24	16	11	14	16	12	13	11	31	17	15	19	11	15	17	19	10	31	11	13	21	10	I2a	100

ID 66	12	24	15	9	13	16	12	13	11	29	14	14	21	12	16	16	19	9	23	12	13	21	10	J2a	100
ID 67	13	24	13	10	16	17	12	13	11	31	17	14	20	12	16	21	18	10	23	13	12	22	12	E1b1b	100
ID 68	13	24	16	11	14	15	13	13	11	32	16	15	20	11	15	20	18	10	30	11	11	22	10	I2a	100
ID 69	13	24	16	11	14	15	13	14	11	32	18	15	19	11	18	18	18	10	29	11	13	23	10	I2a	100
ID 70	13	25	16	10	11	14	10	13	11	30	16	14	20	12	17	20	17	11	21	12	13	23	10	R1a	100
ID 71	14	24	15	11	14	15	12	13	11	31	17	15	20	11	15	17	17	10	30	11	13	22	10	I2a	100
ID 72	13	25	16	11	11	14	10	13	11	29	13	14	20	12	16	18	19	11	24	12	12	23	10	R1a	100
ID 73	13	24	12	10	16	17	9	13	11	30	17	14	19	12	16	18	19	10	22	13	12	21	12	E1b1b	100
ID 74	13	25	16	10	11	14	11	13	11	29	16	14	20	12	17	17	19	11	25	12	11	23	10	R1a	100
ID 75	13	22	14	10	14	14	11	12	11	28	16	16	20	11	15	17	18	10	27	13	11	21	12	I1	100
ID 76	13	24	16	11	14	15	12	13	11	31	17	15	19	11	14	18	18	11	30	11	12	23	10	I2a	100
ID 77	13	24	15	11	14	15	13	13	11	30	17	15	19	11	15	17	18	10	31	11	12	23	10	I2a	100
ID 78	13	24	15	11	14	15	13	13	11	30	18	15	19	11	15	18	18	10	31	11	13	23	10	I2a	100
ID 79	13	24	16	11	14	15	12	13	11	31	17	15	19	12	15	18	18	10	31	11	13	22	10	I2a	100
ID 80	13	25	17	11	11	13	11	13	11	30	15	14	20	15	16	18	19	11	23	12	12	23	9	R1a	100
ID 81	14	25	17	11	11	13	11	13	11	30	15	14	20	15	16	18	19	11	23	12	12	23	9	R1a	100
ID 82	13	24	17	11	14	16	13	13	11	30	18	15	19	11	15	17	18	10	31	11	12	24	10	I2a	100
ID 83	13	24	13	10	16	18	11	14	11	31	15	14	20	12	16	18	19	10	22	12	12	23	12	E1b1b	100
ID 84	13	24	15	11	14	16	13	13	11	31	18	15	19	11	15	19	18	10	31	11	12	22	10	I2a	100
ID 85	12	22	14	10	15	17	11	14	11	29	18	14	19	12	15	19	17	9	23	14	12	20	9	J2a	98.1
ID 86	13	23	14	11	11	13	9	14	16	30	16	14	19	13	14	16	20	11	20	12	11	22	11	N	100
ID 87	12	24	15	11	12	16	12	13	13	29	18	15	19	11	17	19	18	12	22	14	12	23	10	R1b	100
ID 88	13	25	16	10	11	14	11	13	11	29	16	14	20	11	17	18	19	11	25	12	12	23	10	R1a	100
ID 89	13	25	16	11	14	16	12	13	11	31	17	14	19	11	15	18	18	10	30	11	13	23	10	I2a	100
ID 90	13	24	16	11	14	15	12	13	11	31	17	15	19	11	15	18	19	10	30	11	12	22	10	I2a	100
ID 91	14	25	16	10	11	13	10	14	11	31	15	14	20	12	17	20	18	11	21	12	12	23	10	R1a	100
ID 92	13	25	16	11	11	14	10	13	11	30	16	14	20	13	16	20	19	11	23	12	12	24	10	R1a	100
ID 93	12	23	14	9	13	15	10	13	12	30	15	15	20	11	17	18	18	9	24	11	11	22	10	J2a	100
ID 94	13	24	16	11	14	15	12	13	11	31	17	15	19	11	15	19	20	10	32	11	12	22	10	I2a	100
ID 95	13	24	16	11	14	15	13	13	11	30	17	15	20	11	15	18	18	10	29	11	12	24	10	I2a	100
ID 96	13	24	16	11	14	15	12	13	11	32	17	15	19	11	15	18	18	10	26	11	13	23	10	I2a	100
ID 97	13	24	16	11	11	14	11	13	11	31	15	14	20	13	16	17	18	11	23	12	12	23	11	R1a	100
ID 98	13	24	12	10	16	18	12	13	11	31	16	14	20	12	16	17	21	10	22	12	12	22	12	E1b1b	100
ID 99	14	25	14	10	11	16	12	14	14	29	19	14	20	12	14	17	18	10	24	12	12	21	11	I2a	86.7
ID 100	13	24	15	11	14	15	12	13	11	30	18	15	19	11	15	18	19	10	29	10	12	23	10	I2a	100

The most prevalent haplogroup appears to be I2a, which accounts for a total of 49%, while E1b1b and R1a represent the second most abundant haplogroups, each accounting for 17% of all Y chromosomes in the studied population. The remaining six haplogroups appear to be present to a smaller extent in the following manner (Table 2 and Figure 2): J2a (5%), I1 (4%), R1b (4%), J2b (2%), G2a (1%) and N (1%).

Table 2. Distribution of *in silico* assigned Y chromosome haplogroups in Bosnian-Herzegovinian population using Whit Athey's Haplogroup Predictor.

Haplogroup	*Number (n)*	*Percent (%)*
E1b1b	17	17
G2a	1	1
I1	4	4
I2a	49	49
J2a	5	5
J2b	2	2
N	1	1
R1a	17	17
R1b	4	4
Total	*100*	*100*

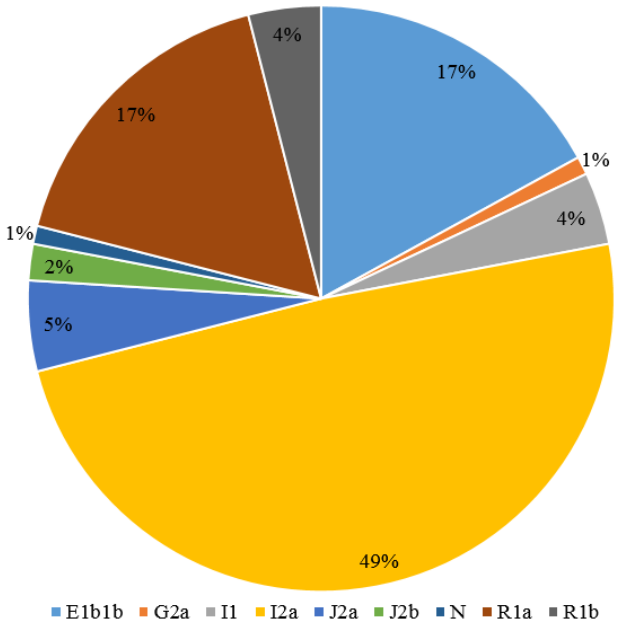

Figure 2: Frequencies of *in silico* assigned Y chromosome haplogroups in Bosnian-Herzegovinian population.

Discussion and Conclusions

In the present study, the Y chromosome haplogroups of 97 out of 100 samples were assigned with 100% accuracy. This can be explained by the fact that the B&H population is relatively small (around 4 million people), thus enabling more accurate software-based predictions. In general, Bayesian probability used by Whit Athey's Haplogroup Predictor has become widely used in the last few years due to its accurate, sensitive, as well as reproducible results. For example, Petrejčikova et al. (2009 and 2010) performed two separate studies in Slovak Romany and general Slovak populations, respectively. The haplogroup prediction accuracy of 12 Y-STR loci on 200 Slovak Romanies and 250 Slovaks appeared to be at least 60% in all 450 samples, with a prediction accuracy of above 95% in a great majority of samples (Petrejčikova et al. 2009; Petrejčikova et al. 2010). Considering the aforementioned two investigations, it is important to note that prediction accuracy increases if the prediction is based on more than 20 Y-STRs (Taylor et al. 2012), which is also the approach used in this study. Furthermore, in a recent report on research conducted on the Nicaraguan population (Nunez et al. 2012), Y chromosome haplogroups were predicted by Whit Athey's algorithm, as well as by Y-SNP genotyping experimentally. The results have shown that the algorithm made false predictions in eight out of 165 samples (4.8%) (Nunez et al.

2012). Also, Larmuseau et al. (2014) investigated novel Y chromosome haplogroups in Western Europe (Belgium and Netherlands), partially using the for preliminary haplogroup determination, and their results reported only a single mistake in the Y chromosome regarding haplogroup A, which is specific for African continent (Karmin et al. 2015; Larmuseau et al. 2014). Finally, a recent study has reported the application of Vadim Urasin's YPredictor, which is based on phylogenetic tree analysis of each haplogroup, and it uses differences in markers and information on their mutation rates in order to predict Y chromosome haplogroups from Y-STR data (Wang et al. 2015). It has therefore been concluded that the most reliable way to determine a haplogroup is to perform Y-SNP typing, with the convergence of STR profiles among different haplogroups not yet being completely flawless in haplogroup prediction (Wang et al. 2015).

Overall, nine haplogroups have been detected in B&H population, with I2a, E1b1b, and R1a being the most abundant haplogroups, each of them being present in more than 15% of the population. The most prevalent haplogroup in B&H population is by far the sublineage I2a of haplogroup I, with a frequency of 49% in the studied population. According to previous studies, haplogroup I has been thought to originate from the Balkan Peninsula since approximately 45,000 years ago, which corresponds to the time before the Last Glacial Maximum (LGM). This haplogroup most probably originated from the Middle East before coming to Europe through Anatolia (Battaglia et al. 2009; Primorac et al. 2011). The high frequency of haplogroup I2a observed in the studied population is in accordance with a previously published article on Y chromosome haplogroups in Bosnia and Herzegovina, in which haplogroup I was by far the most prominent haplogroup (Marjanović et al. 2005; Marjanović et al. 2006), making up more than 50% of all haplogroups detected in a representative B&H population. Furthermore, the results of this study are consistent with those published on the Croatian mainland population, in which haplogroup I2a accounted for 37.6%, as well as with the isolated populations of four Adriatic islands (Krk, Brač, Hvar and Korčula), in which it accounted for an average of 50.8% (Barać et al. 2003). The frequency of haplogroup I2a in Serbia is somewhat lower and accounts for 29.2% (Peričić et al. 2005), which shows similarities with Macedonia (28.8%), Albania (16.7%; Bosch et al. 2006), and Slovenia (20%; Battaglia et al. 2009; Table 3). On the Balkan Peninsula, average frequency of HgI is 36.3% (Battaglia et al. 2009). When compared to other European populations, it is evident that haplogroup I2a can indeed be considered as typical for south-east Europe, where it is most widely present and which is also suggested by newer scientific reports (Kushniarevich et al. 2015). For example, haplogroup I was not detected in the Lithuanian population (Kasperavičiute et al. 2004), accounted for only 19% of the population from UK (King and Jobling, 2009) and, despite relative geographic proximity, was present in only 0-2% of all Y chromosomes analyzed in North Italy (Battaglia

et al. 2009).

Table 3. A comparison of nine Y chromosomal haplogroups distribution observed in the present study with previously obtained SNP-based data for Bosnian-Herzegovinian population (Marjanović et al. 2005) and neighboring populations, namely Croatian mainland population and populations from four isolated Adriatic islands (Krk, Brač, Hvar and Korčula) (Barać et al. 2003), Serbian (Peričić et al. 2005), Macedonian (Bosch et al. 2006) and Slovenian (Battaglia et al. 2009) populations. For four isolated populations from Adriatic islands, arithmetic average of four values is calculated and presented here. N/A indicates that frequency of a certain haplogroup was not given in the reference study.

Haplogroup	*Present study*	*SNP-based Bosnian-Herzegovinian*	*Croatian (mainland)*	*Croatian (isolated islands)*	*Serbian*	*Macedonian*	*Slovenian*
I2a	49%	49.2%	37.6%	50.8%	29.2%	28.8%	20%
R1a	17%	13.7%	33.9%	23.3%	15.9%	13.5%	38.7%
E1b1b	17%	13.7%	5.5%	4.7%	18.6%	21.2%	2.7%
J2a	5%	N/A	N/A	N/A	N/A	N/A	1.3%
R1b	4%	3.9%	15.6%	6.2%	10.6%	13.5%	21.3%
I1	4%	2.3%	N/A	N/A	5.3%	N/A	6.7%
J2b	2%	2.7%	1.8%	4.1%	4.4%	11.5%	N/A
G2a	1%	2%	0.9%	4.4%	N/A	3.8%	1.3%
N	1%	N/A	N/A	N/A	N/A	N/A	N/A

The second most prevalent haplogroup R1a (17%) was brought to Bosnia and Herzegovina as a post-LGM event (less than 15,000 years ago), according to a commonplace theory that claims that it was most probably brought to Europe from Western Asia (Primorac et al. 2011). The result of this study regarding R1a compares well to the previous study of the B&H population, in which this haplogroup accounted for 13.7% (Marjanović et al. 2005). In comparison to the B&H population, R1a is more frequent in Croatia (33.9% in the mainland part of the country and an average of 23.3% in four Adriatic islands; Barać et al. 2003) and Slovenia (37%; Peričić et al. 2005). This haplogroup is however less frequent in Serbia (15.9%), Macedonia (15.2%; Table 3), and Albania (9.8%; Peričić et al. 2005), while being more frequent in northern parts of Europe, with a frequency of 44% in Ukraine (Kharkov et al. 2004), 21.6% in the Tatar population from Russia (Akhatova et al. 2013), 44.9% in Lithuania, 39.9% in Latvia, 46.7% in Russia, 54.5% in Poland (Kasperavičiute et al. 2004), and 36.4% of R1 in the Swedish population (Karlsson et al. 2006). HgR1 on the Balkans accounts for more than 30% of all investigated subjects on average (Battaglia et al. 2009). Contrary, subclade of HgR, R1b, was present in the study population as a minor fraction, accounting for 4% of all tested individuals. However, this sublineage can be

considered interesting for further investigation, as its origins in the region of South-Eastern Europe (SEE), more specifically on the Balkan Peninsula, have recently been reassessed. Older reports suggest that R1b-bearing males represent "Old Europeans", as this haplogroup came to Eurasia 25,000 years ago on the territory of the Iberian Peninsula and the Atlantic Coast, which served as one of four major refuges during the LGM (Wiik 2008). However, more recent investigations (Myres et al. 2011; Primorac et al. 2011) state that Hg R1b came to the Balkans as a post-LGM event (15,000-10,000 years ago) from West Asia, most probably through Lebanon. The latter opinion is supported by the fact that R1b males from this region miss R1b-specific M412 mutation. HgI- and HgR-related migrations are once again confirmed by Kovačević et al. (2014), who identified a bidirectional and, most probably, multiple gene flow between Europe and Middle East on the basis of both autosomal and Y-chromosomal markers (Kovačević et al. 2014).

Haplogroup E1b1b (also known as E3b1) appears to have the same frequency in the studied population as haplogroup R1a. It is believed to have originated from the African continent (Battaglia et al. 2009) and provides evidence of the last direct migration from Africa to Europe. It is a representative haplogroup of "Early Farmers" in Europe, since it has arrived to Europe during the Neolithic period, or after the LGM. The second opinion on haplogroup E3 is that it is a Balkan-specific lineage which has spread from the area of Ancient Greece to the northern part of the peninsula as far as 10,000-8,000 years ago (Primorac et al. 2011), meaning that it is another subclade of uncertain origin which is still under massive investigation. Previous studies of the representative B&H population by Marjanović et al. (2005 and 2006) have also shown that haplogroup E1b1b is the second most prevalent haplogroup in Bosnia and Herzegovina, accounting for around 13.7% of the population. This haplogroup is more prevalent in the southern rather than in the northern part of the Balkans since it accounts for 18.6% in Serbia, 21.2% in Macedonia, 23.3% in Albania, 17.1% in Greece (Peričić et al. 2005), with a total of 34.4% in the Albanian population living in FYR of Macedonia (Battaglia et al. 2009). On the other hand, within the inland Croatian population it accounts for merely 5.5%, with only 4.7% in the isolated population of Adriatic islands (Barać et al. 2003; Table 3). This haplogroup is typical for the Balkans and exhibits a strong decreasing south-to-north gradient, which supports a previously mentioned theory that HgE3 originates from the area of modern day Greece (Primorac et al. 2011). More precisely, E1b1b accounts for only 1% in Sweden (Karlsson et al. 2006) and in the UK (King and Jobling, 2009), while all sublineages of the E haplogroup together account for 4.2% in Ukraine (Kharkov et al. 2004).

These results represent an update of earlier preliminary B&H population data and confirm the prediction that Bosnia and Herzegovina can be thought of as the

part of the Western Balkan area that served as the Last Glacial Maximum refuge for the Paleolithic human European population. Furthermore, it has been confirmed that this area was a very important stopping point on the "Middle East-Europe highway" during the Neolithic farmer migrations. These migrations most probably occurred from Western Asia and Northern Africa towards Europe, which is confirmed by the presence of R and E haplogroups, respectively. These findings prove once again that the Y chromosome is of great importance for tracing human history and migration patterns, even in the case of mixed populations such as the one residing in Bosnia and Herzegovina, which is repeatedly proven in recent studies that compared NRY- and mtDNA-derived markers, concluding that Y chromosome is more convenient for genealogical studies and investigation of genetic history of populations, giving more precise results and improved resolution (Kovačević et al. 2014; Karmin et al. 2015; Kushniarevich et al. 2015).

As far as accuracy of the predictions made by this study is concerned, the results can be considered satisfying as they are in accordance with previous experimental investigations of the Y chromosome haplogroups in B&H population. The compliance with results published for neighboring populations makes this statement even stronger, since other recent investigations utilizing NGS-based approach for the analysis of autosomal, NRY and mtDNA markers did not identify significant differences within the population of South Slavs on the basis of any of these three sets of human DNA markers (Kovačević et al. 2014; Kushniarevich et al. 2015). Therefore, it can be concluded that the Bayesian probability is an excellent statistical tool for the assignment of Y chromosome haplogroups from Y-STRs. In this way, already published data can be applied in a completely new manner, as conducted within this study, providing important information about the genetic background of a population that can be easily and rapidly discovered. Even if Y-STR data are not published yet for a certain population and have to be genotyped *de novo*, such an approach is still less expensive, less labor-intensive and not prone to experimental mistakes when compared to Y-SNP analysis.

Acknowledgments The authors would like to thank Naida Babić for allowing the use of unpublished population data over 23 Y-STR loci for six individuals from Tuzla Canton, Bosnia and Herzegovina.

Received 4 May 2016; revision uploaded 10 July 2016.

Literature Cited

Akhatova, F. S., F. F. Rizvanova, and E. K. Khusnutdinova. 2013. Y-chromosome haplotypes in the populations of Tatar in Russia. *Middle East J. Sci. Res.* 17:507-509.

Athey, W. T. 2005. Haplogroup prediction from Y-STR values using an allele-frequency approach. *J. Genet. Geneal.* 1:1-7.

Athey, W. T. 2006. Haplogroup prediction from Y-STR values using a Bayesian-allele-frequency approach. *J. Genet. Geneal.* 2:34-39.

Barać, L., M. Peričić, I. Martinović-Klarić et al. 2003. Y chromosomal heritage of Croatian population and its island isolates. *Eur. J. Hum. Genet.* 11:535-542.

Battaglia, V., S. Fornarino, N. Al-Zahery et al. 2009. Y-chromosomal evidence of the cultural diffusion of agriculture in southeast Europe. *Eur. J. Hum. Genet.* 17:820-830.

Dogan, S., A. Ašić, S. Buljubašić et al. 2015. Overview of European population clustering based on 23 Y-STR loci. *Genetika (Belgrade)* 47:901-908.

Hughes, J. F., and S. Rozen. 2012. Genomics and genetics of human and primate Y chromosomes. *Annu. Rev. Genomics Hum. Genet.* 13:83-108.

Karlsson, A. O., T. Wallerstrom, A. Gotherstrom et al. 2006. Y-chromosome diversity in Sweden: A long-time perspective. *Eur. J. Hum. Genet.* 14:963-970.

Karmin, M., L. Saag, M. Vicente et al. 2015. A recent bottleneck of Y chromosome diversity coincides with a global change in culture. *Genome Res.* 25:459-466.

Kasperavičiute, D., V. Kučinskas, and M. Stoneking. 2004. Y chromosome and mtDNA variation in Lithuanians. *Ann. Hum. Genet.* 68:438-452.

Kharkov, V. N., V. A. Stepanov, S. A. Borinskaia et al. 2004. Structure of the gene pool of eastern Ukrainians from Y-chromosome haplogroups. *Genetika.* 40:415-421.

King, T. E., and M. A. Jobling. 2009. What's in a name? Y chromosomes, surnames and the genetic genealogy revolution. *Trends Genet.* 25:351-360.

Kovačević, L., K. Tambets, A. M. Ilumae et al. 2014. Standing at the gateway to Europe – The genetic structure of Western Balkan populations based on autosomal and haploid markers. *PLoS one* 9:e105090.

Kovačević, L., V. Fatur-Cerić, N. Hadžić et al. 2013. Haplotype data for 23 Y-chromosome markers in a reference sample from Bosnia and Herzegovina. *Croat. Med. J.* 54:286-290.

Kushniarevich, A., O. Utevska, M. Chuhryaeva et al. 2015. Genetic heritage of the Balto-Slavic speaking populations: A synthesis of autosomal,

mitochondrial and Y-chromosomal data. *PLoS one* 10:e0135820.
Larmuseau, M. H. D., N. Vanderheyden, A. Van Geystelen et al. 2014. Increasing phylogenetic resolution still informative for Y chromosomal studies on West-European populations. *Forensic Sci. Int. Genet.* 9:179-185.
Marjanović, D., S. Fornarino, S. Montagna et al. 2005. The peopling of modern Bosnia-Herzegovina: Y-chromosome haplogroups in the three main ethnic groups. *Ann. Hum. Genet.* 69:757-763.
Marjanović, D., N. Pojskić, B. Kalamujić et al. 2006. Most recent investigation of peopling of Bosnia and Herzegovina: DNA approach. *Documenta Praehistorica.* 33:21-28.
Myres, N. M., S. Rootsi, A. A. Lin et al. 2011. A major Y-chromosome haplogroup R1b Holocene era founder effect in Central and Western Europe. *Eur. J. Hum. Genet.* 19:95-101.
Nunez, C., M. Geppert, M. Baeta et al. 2012. Y chromosome haplogroup diversity in a Mestizo population of Nicaragua. *Forensic Sci. Int. Genet.* 6:192-195.
Peričić, M., L. Barać-Lauc, I. Martinović-Klarić et al. 2005. High-resolution phylogenetic analysis of Southeastern Europe traces major episodes of paternal gene flow among Slavic populations. *Mol. Biol. Evol.* 22:1964-1975.
Petrejčikova, E., M. Sotak, J. Bernasovska et al. 2009. Y-haplogroup frequencies in the Slovak Romany population. *Anthropol. Sci.* 117:89-94.
Petrejčikova, E., M. Sotak, J. Bernasovska et al. 2010. The genetic structure of the Slovak population revealed by Y-chromosome polymorphisms. *Anthropol. Sci.* 118:23-30.
Primorac, D., D. Marjanović, P. Rudan et al. 2011. Croatian genetic heritage: Y chromosome story. *Croat. Med. J.* 52:225-234.
The Y Chromosome Consortium. 2002. A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res.* 12:339-348.
Tylor, D., N. Nagle, K. N. Ballantyne et al. 2012. An investigation of admixture in an Australian Aboriginal Y-chromosome STR database. *Forensic Sci. Int. Genet.* 6:532-538.
Wang, C. C., L. X., Wang, R. Shrestha et al. 2015. Convergence of Y chromosome STR haplotypes from different SNP haplogroups compromises accuracy of haplogroup prediction. *J. Genet. Genomics* 42:403-407.
Wiik, K. 2008. Where did European men come from? *J. Genet. Geneal.* 4:35-85.