Anthropological Review • Vol. 82(3), 219–238 (2019)

sciendo

ANTHROPOLOGICAL REVIEW
Available online at: https://doi.org/10.2478/anre-2019-0016

Analysis of the mitochondrial *CYTB* gene sequence in human populations of northeastern Bosnia

Adisa Ahmic[1], Irma Mujkic[1], Anel Ismailovic[1], Emir Halilovic[1], Lejla Lasic[2], Belma Kalamujic[2], Naris Pojskic[2]

[1]Department of Biology, Faculty of Natural Sciences, University of Tuzla, Tuzla, Bosnia and Herzegovina
[2]Institute for Genetic Engineering and Biotechnology, University of Sarajevo, Sarajevo, Bosnia and Herzegovina

ABSTRACT: This study offers the first report on variation sequence of the mitochondrial cytochrome b (*MT-CYTB*) gene in populations from Bosnia (northeastern Bosnia). This study was designed on the analysis of the genetic diversity of two populations of different cultural-anthropological and genetic origin, Roma population and native/non-Roma population. The main aim of our study was to estimate the usefulness of the *CYTB* sequence in the analysis of genetic categorization of different populations and intergroup diversity, as well as to provide some additional information on haplogroup-associated polymorphisms within the *CYTB* region in defining haplogroup status. Estimation of the genetic diversity was done using intra and intergroup genetic indices. The population-specific polymorphisms have been found in both categories of the populations. The results of the analysis of genetic differentiation show significant pairwise Fst differences between the Romani and native populations. Also, registered significant genetic differentiation is illustrated on the level of genetic variation between subpopulations of the Roma and non-Roma origin. The important result in our study is the confirmation of the significance of the triad of polymorphisms T14783C-G15043A-G15301A, indicating the influence of Asian component of the maternal gene pool on the genetic structure of the studied population of the Roma. Our data show that the haplogroup polymorphisms exist in the *CYTB* region and can provide useful information on the haplogroups that were defined only by the control region of the mtDNA. The results of this study indicate the region of *CYTB* gene can be a benefit in providing some additional information in the analysis of genetic structure of human populations and can be additionally applied in population studies.

KEY WORDS: *CYTB* polymorphisms, genetic differentiation, haplogroups

Introduction

Sequence analyses of the hypervariable segments I and II (HVSI/II) of the non-coding region (control region, CR) of mtDNA are often used for genetic identification of individuals and categorization of different populations for fo-

Original Research Article Received: April 9, 2019; Revised: July 4, 2019; Accepted: July 17, 2019
DOI: 10.2478/anre-2019-0016

rensic purposes and population genetic study. Also, due to a high evolutionary rate of mutations and maternal inheritance, the HVS-s are an important tool for the estimation of genetic origin, gene flow, migration and population expansion (Richards et al. 2000; Simoni et al. 2000; Torroni et al. 2001; Malyarchuk et al. 2010). However, several features of the HVS-s of the mtDNA such as mutational hotspots, high mutation rate, variation in substitution rate between nucleotide sites, parallel mutations, different rates of substitution among HVSI/II, higher frequency of transitions than transversions and variability in bases composition can cause problems in the analysis of the no-coding region (Brown et al. 1979; Meyer et al. 1999; Tamura et al. 2000; Bandelt et al. 2001; Malyarchuk et al. 2002; 2013). In order to improve the information obtained from the mtDNA, research studies apply the complete mtDNA analysis or analysis which combines the data of the HVS-s and part of the coding region.

Few studies indicate that some parts of the coding region of the mtDNA, such as 10171–10659 (Yao et al. 2002), 14576–16047 (Kong et al. 2003) and 14747–15886 (Lee et al. 2002; Hwa et al., 2010; 2011; Ablimit et al. 2013; Amer et al. 2015; Farghadani and Babadi 2015) are very informative in the phylogenetic and phylogeographic analysis of mtDNA, forensic applications and population studies. Studies on the variations of the *CYTB* gene indicate that this gene can be used as a candidate with high discrimination power for individual maternal identification in forensic casework, estimation of intergroup diversity and defining of haplogroup status (Lee et al. 2002; Tsai et al. 2001; Kong et al. 2003; Hwa et al. 2010; 2011; Ablimit et al. 2013; Farghadani and Babadi 2015; Amer et al. 2015). The *CYTB* gene covers position 14747–15887 including 1140 np (nucleotide pairs) of the human mtDNA coding region (Anderson et al. 1981).

The main aim of our study was to analyse the variation of the *CYTB* gene among two populations of different cultural-anthropological and genetic origin (Roma population and native/no-Roma population) in order to evaluate the usefulness of the *CYTB* sequence in genetic identification of different populations, estimate intergroup diversity as as well as to provide additional information on haplogroup-associated polymorphisms within the *CYTB* region in defining haplogroup status. With the aim of determining whether the region of the *CYTB* is informative for the mentioned aims, the obtained parameters of the *CYTB* sequence diversity are compared with the data based on the very informative HVSI of the mtDNA.

Materials and methods

Population samples

A total of 74 buccal samples have been collected from unrelated individuals in the region of northeastern Bosnia. The collected samples included 39 Roma samples collected from two Romani communities (marked in the manuscript as Roma population 1 and Roma population 2) and 35 samples of the native population (no-Roma population) collected from the isolated villages of two mountain regions (Konjuh and Majevica) (marked in manuscript as native population 1 and native population 2). The collected samples of the Roma population were unspecified by individual ethnonyms and represent a mixed pop-

ulation (according to migration category and language dialect) of the Roma from northeastern Bosnia. Prior to sampling, all participants provided written consent for the collection of samples and subsequent analysis. Our study follows the principles of the Declaration of Helsinki and all subjects gave their consent to participate in this study.

DNA amplification and sequencing

Total genomic DNA was extracted from the dried swab samples of buccal mucosa, using the salting out method (Miller et al. 1998). The method was slightly modified in order to optimize the extraction of DNA from buccal swab. The set of primers L14724 and H15915 was used for PCR reaction (*Polymerase Chain Reaction*) of the *CYTB* sequence (Hwa et al. 2010; Ablimit et al. 2013). The PCR reactions were performed with GeneAmp® P*CR* System 9700 (*Applied Biosystem*). The success of the amplification reactions was checked by the process of separation of the obtained fragments on 1.5% agarose gel in 1xSB (Sodium Borate) buffer. Sizes of the obtained fragments were estimated using EDAS software (*Electrophoresis Documentation and Analysis System*. In the next step, sequencing of the *CYTB* sequences was performed using the primers the primers L14724 and 15283 (Ablimit et al., 2013). The HVSI has been sequenced in a total of 74 individuals (35 individuals of the native population from this study and 39 Roma samples from previous studies, Ahmić et al. (2018). The set of primers F15971/R16411 was used for PCR reaction of the HVSI region sequence. PCR products were sequenced within facilities of the Macrogen Korea Inc. as their regular capillary DNA sequence services.

Data analysis

The nucleotide sequences of the *CYTB* gene were aligned with the revised Cambridge reference sequence, rCRS (Anderson et al. 1981; Andrews et al. 1999) using the BioEdit software (Hall 1999), through ClustalW multiple alignment. The population genetic structure of the investigated populations was analyzed using the methods implemented in the Arlequin ver.3.11 software (Excofier et al. 2005). Intrapopulation genetic diversity parameters, such as the number of different sequences (DS), mean number of pairwise difference (pi), sequence diversity values (H) and nucleotide diversity (π) were calculated at the level of the *CYTB* sequence. *Pairwise* F_{ST} analysis (Weir and Cockerham 1984) was used for the assessment of intergroup genetic differentiation. Based on pairwise Fst values, multidimensional scaling (MDS) analysis was performed using SPSS Statistics 17.0 for Windows (SPSS, Chicago, IL,USA). The Median-joining network of the *CYTB* haplotype was constructed using a median-joining algorithm (Bandelt et al. 1999) as implemented in the Network 4.6.1.2. software. For the comparation of the obtained data based of the *CYTB* sequences, parameters of intra and interpopulation genetic diversity were analysed based on the level of the HVSI. For the identification of the haplogroups mitomaster software (www.mitomap.com), Haplogrep software (Kloos-Brandstätter et al. 2011; Weissensteiner et al. 2016) and the PhyloTree 17 (van Oven and Kayser 2009) were used. Haplogroup-associated polymorphisms of the *CYTB* gene were defined according to the scheme by Herrnstadt et al (2002).

Results

PCR fragments of 1190 np were amplified in all samples. The *CYTB* sequence comprised approximately 1140 np (region 14747–15887). The sequences of the *CYTB* gene of all samples were aligned with the standard reference sequence (rCRS, Andrews et al., 1999). Based on the obtained data, we determined: a) polymorph position the *CYTB* sequences and population-specific polymorphisms, as well as parameters of intra and inter-population of the genetic diversity of the studied populations; b) position of Bosnian Roma population in Asian population group and c) haplogroup-associated polymorphisms of the *CYTB* region.

Variations of the CYTB sequence in studied populations

Distribution of the observed polymorph positions of the *CYTB* sequence with regard to rCRS position in analysed populations in our study is showed in Table 1. A total of 61 nucleotide substitutions were found in all the studied samples. The majority of nucleotide substitutions (39) were transitions, and there were 22 transversions. There were no deletions and/or insertions observed. Six common polymorph positions (C14766T, T14798C, G14905A, A15326G, C15452A, A15607G) were noticed between the Roma population and the native population from north-eastern Bosnia (Table 1). Polymorphism A15326G, which is common to mtDNA sequences from Europe, Asia and Africa (Herrnstadt et al. 2002) was observed in more than 90% in both category populations. Second common nucleotide change C→T on nucleotide position 14766, which is present on African and West Euroasian mtDNA lineages (Finnilä et al. 2001; Herrnstadt et al. 2002) was observed with average frequency of 48.72% in the Roma population and 40.0% in the native population. Polymorphism T14798C which is specific for Caucasian populations (Hwa et al. 2011; Ablimit et al. 2013) was observed in the native populations of 11.43%, and in Roma population of 7.69%. Other common positions were G14905A (5.13%, the Roma; 2.85%, the native population), C15452A (7.69%, the Roma; 11.43%; the native population) and A15607G (7.69%, the Roma; 2.85%, the native population). The population-specific polymorph positions (Table 1) with higher frequencies such as T14783C (20.51%), G15043A (20.51%), 15287C (15.38%), G15301A (33.33%), T15310C, A15397G (10.26%) were identified only in the Roma population. The population-specific positions of the *CYTB* with higher frequencies identified only in the native group were G15077A (11.43%) and C14872T (8.33%). Other

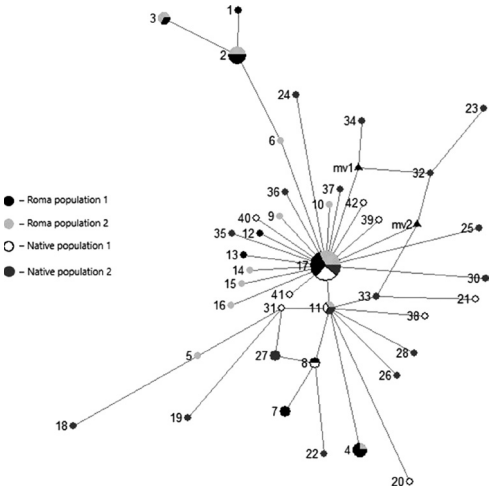

Fig. 1. Median-joining network of the *CYTB* haplotypes found in the studied populations. Circle areas are proportional to haplotype frequencies

Table 1. Distribution of common and population-specific polymorphisms of the *CYTB* sequences with regard to rCRS position in studied populations

rCRS position	Nucleotide change	Roma population 2 (N=20)	Roma population 1 (N=19)	Total (%)	Native population 1 (N=15)	Native population 2 (N=20)	Total (%)	Mutation type	Amino-acid substitution
14759	C→A					1	2.85	transversion	Arg→Ser
14766	C→T	9	10	48.72	5	9	40.0	transition	Thr→Ile
14783	T→C	5	3	20.51				transition	Leu→Leu
14793	A→G				1	1	5.71	transition	His→Arg
14798	T→C	1	2	7.69	1	3	11.43	transition	Phe→Leu
14809	C→T	1		2.56				transition	Leu→Leu
14814	C→T					1	2.85	transition	Thr→Ile
14831	G→A					2	5.71	transition	Ala→Thr
14841	A→C					1	2.85	transversion	Asn→Thr
14845	C→T				1		2.85	transition	Phe→Phe
14846	G→A				1		2.85	transition	Gly→Ser
14846	G→T					1	2.85	transversion	Gly→Cys
14859	G→A					1	2.85	transition	Gly→Asp
14872	C→T					3	8.57	transition	Ile→Ile
14905	G→A	1	1	5.13		1	2.85	transition	Met→Met
14947	C→A					1	2.85	transversion	Ala→Ala
14999	A→G					1	2.85	transition	Asn→Asp
15034	A→G	3		7.69				transition	Leu→Leu
15043	G→A	5	3	20.51				transition	Gly→Gly
15053	T→A					1	2.85	transversion	Tyr→Asn
15067	T→C					1	2.85	transition	Phe→Phe
15077	G→A				1	3	11.43	transition	Glu→Lys
15095	A→T					1	2.85	transversion	Ile→Phe
15131	A→C					1	2.85	transversion	Met→Leu
15133	A→G					1	2.85	transition	Met→Met
15179	G→T					1	2.85	transversion	Val→Leu
15218	A→G				1		2.85	transition	Thr→Ala
15247	C→T					1	2.85	transition	Gly→Gly
15257	G→A		2	5.13				transition	Asp→Asn
15266	A→ C					1	2.85	transversion	Thr→Pro

rCRS position	Nucleotide change	Roma population 2 (N=20)	Roma population 1 (N=19)	Total (%)	Native population 1 (N=15)	Native population 2 (N=20)	Total (%)	Mutation type	Amino-acid substitution
15287	T→A	3	3	15.38				transition	Phe→Leu
15289	T→C	2		5.13				transition	Phe→Phe
15295	C→G					1	2.85	transversion	Phe→Leu
15296	A→G					1	2.85	transition	Ile→Val
15301	G→A	7	6	33.33				transition	Leu→Leu
15310	T→C	1	3	10.26				transition	Ile→Ile
15326	A→G	20	16	92.31	12	20	91.43	transition	Thr→Ala
15368	C→A					1	2.85	transversion	Pro→Thr
15397	A→G	1	3	10.26				transition	Lys→Lys
15437	G→C					1	2.85	transversion	Gly→Arg
15452	C→ A	1	2	7.69	1	3	11.43	transversion	Leu→Ile
15474	C→A					1	2.85	transversion	Thr→Lys
15481	C→T		1	2.56				transition	Phe→Phe
15503	C→A					1	2.85	transversion	Pro→Thr
15513	A→C					1	2.85	transversion	Tyr→Ser
15524	A→G						2.85	transition	Asn→Asp
15527	C→A				1	1	5.71	transversion	Pro→Thr
15538	C→A					1	2.85	transversion	Thr→Thr
15546	A→T				1		2.85	transversion	Hys→Leu
15607	A→G	1	2	7.69		1	2.85	transition	Lys→Lys
15617	G→A	1		2.56				transition	Val→Ile
15677	A→G					1	2.85	transition	Lys→Glu
15699	G→C		1	2.56				transversion	Arg→Pro
15703	A→G					1	2.85	transition	Pro→Pro
15772	A→G					1	2.85	transition	Pro→Pro
15777	G→A					1	2.85	transition	Ser→Asn
15792	T→C	1		2.56				transition	Ile→Thr
15806	G→A		1	2.56				transition	Ala→Thr
15833	C→T	1		2.56				transition	Leu→Leu
15843	T→A		1	2.56				transversion	Met→Lys
15884	G→C				1		2.85	transversion	Ala→Pro

Table 2. Distribution of the *CYTB* haplotypes in studied populations

Haplotype/ sequence types	Polimorph positions	Roma population 2	Roma population 1	Native population 1	Native population 2	Total
1	C14766T, T14783C, G15043A, T15287C, G15301A, A15326G, G15806A		1			1
2	C14766T, T14783C, G15043A, T15287C, G15301A, A15326G	3	3			6
3	C14766T, T14783C, G15043A, T15289C, G15301A, A15326G	2	1			3
4	C14766T, A15034G, T15310C, A15326G, A15397G	1	3			4
5	C14766T, G14905A, A15326G, C15452A, A15607G	1				1
6	G15043A, T15287C, G15301A, A15326G	1				1
7	C14766T, T14798C, G15257A, A15326G		2			2
8	C14766T, T14798C, A15326G		1	1		2
9	A15326G, T15843A	1				1
10	C14809T, A15326G	1				1
11	C14766T, A15326G	1		1	1	3
12	A15326G, C15481T		1			1
13	A15326G, G15699C		1			1
14	A15326G, C15833T	1				1
15	A15326G, G15617A	1				1
16	A15326G, T15792C	1				1
17	A15326G	6	6	5	2	19
18	C14766T, T14798C, G14905A , A15326G, C15452A, C15527A, C15538A, A15607G				1	1
19	C14766T, C14947A, A14999G, T15067C, A15326G, C15452A				1	1
20	C14766T, A14793G, A14841C, G14846A, G14859A, A15326G			1		1
21	C14766T, A14793G, G15077A, A15218G, A15326G			1		1

Haplotype/ sequence types	Polimorph positions	Roma population 2	Roma population 1	Native population 1	Native population 2	Total
22	C14766T, T14798C, C15295G, A15296G, A15326G				1	1
23	C14872T, G15077A, A15095T, A15326G, C15368A				1	1
24	C14814T, T15053A, A15131C, A15326G, 15513C				1	1
25	C14759A, G15179T, A15266C, A15326G				1	1
26	C14766T, 14831A, A15133G, A15326G				1	1
27	C14766T, T14798C, A15326G, G15452A				2	2
28	C14766T, A15326G, A15772G, G15777A				1	1
30	A15326G, C15474A, C15503A A15677G				1	1
31	C14766T, A15326G, C15452A			1		1
32	C14872T, G15077A, A15326G				1	1
33	C14766T, G15077A, A15326G				1	1
34	G14831A, C14872T, A15326G				1	1
35	A15326G, A15524G, A15703G				1	1
36	C15247T, A15326G				1	1
37	A15326G, G15437C				1	1
38	C14766T, G15884C			1		1
39	A15326G, C15527A			1		1
40	C14845T, A15326G			1		1
41	A15326G, A15546T			1		1
42	G14846T, A15326G			1		1
Total number		12	9	11	18	42

population-unique polymorphisms were represented with a frequency of 2–8% in both category populations (Table 1).

The variation analysis of the *CYTB* sequence in the studied category populations enabled the identification of total 42 haplotype/sequence types, which differed from the rCRS sequence (Table 2). Among 39 sequences, 17 haplotypes were noted in the Roma population, and among 35 sequences in the native populations, 28 haplotypes were noted. The phylogenetic network of the *CYTB* haplotypes (Fig. 1) shows the existence of the major cluster with haplotype 17 (one nucleotide change A→G at np 15326) which is presents in all subpopulations and two clusters with sporadic presence in studied populations, the cluster with haplotype 11 and cluster with haplotype 8 (Table 2). Also, the network showed the existence of three common clusters in the Roma populations: the connected cluster with haplotypes 2 (polymorphisms C14766T, T14783C, G15043A, T15287C, G15301A, A15326G) and 3 (polymorphisms C14766T, T14783C, G15043A, T15289C, G15301A, A15326G, and the cluster with haplotype 4 (polymorphisms C14766T, A15034G, T15310C, A15326G, A15397G). In the native populations, excluding haplotype 17 and 11, there were no common hapotypes.

For the purpose of the clear information on the applicability of polymorphisms of the *CYTB* sequences in population studies, the obtained parameters of the intra and interpopulation genetic diversity based on the *CYTB* sequences were compared with the values of these parameters based on the level of the HVSI CR of the mtDNA in the studied category populations. Table 3 presents parameters of the intrapopulation diversity: number of different sequences (DS), the sequence diversity (H), the nucleotide diversity (π) and mean number of pairwise difference (pi) based on the *CYTB* and the HVSI sequences. The *CYTB* sequences showed a high level of the sequence diversity (H), which was found in both categories of the studied populations (0.8713–0.900 in the Roma populations; 0.9048–0.9947 in the native populations). In contrast, the nucleotide diversity (π) values were lower for each population. Also, similar relations of values of molecular diversity parameters

Table 3. Intrapopulation diversity parameters based on the *CYTB* and the HVSI sequences of studied populations

	N	DS	H±sd	π±sd	pi±sd
Based on CYTB/Populations					
Roma population 1	20	12	0.9000±0.0532	0.002547±0.0016	2.210526±1.2735
Roma population 2	19	9	0.8713±0.0547	0.004279±0.0025	3.555556±1.8904
Native population 1	15	11	0.9048±0.0719	0.005598±0.0034	1.561905±0.9860
Native population 2	20	19	0.9947±0.0178	0.003961±0.00297	1.168421±0.7845
Based on HVSI/Populations					
Roma population 2	20	13	0.9474±0.0323	0.00723±0.0052	1.315789±0.8550
Roma population 1	19	12	0.9357±0.0368	0.008001±0.0006	1.456140±0.9243
Native population 1	15	14	0.9905±0.0281	0.011866±0.0078	2.171429±1.2737
Native population 2	20	15	0.9684±0.0254	0.0144609±0.0090	2.636842±1.4689

N – number samples; DS – number of different sequences; H – sequence (haplotype) diversity; π – nucleotide diversity; pi – mean number of pairwise difference.

Table 4. Matrix of genetic differentiation (pFST) between comparative populations based on the *CYTB* sequences (below the diagonal) and the HVSI sequences (above the diagonal)

Populations	Native population 1	Native population 2	Roma population 2	Roma population 1
Native population 1	×	0.02077	0.03154	0.03744
Native population 2	0.04918	×	0.04210	0.04789
Roma population 2	0.09769	0.05263	×	0.05846
Roma population 1	0.11240	0.06669	0.11425	×

between the studied categories of populations were obtained on the level of the HVSI sequences (Table 3).

Table 4 presents the analysis of genetic differentiation between pairs of the observed populations (pF_{ST}) based on the *CYTB* and HVSI sequences. The results of the analysis of genetic differentiation based on the *CYTB* sequences show significant pairwise Fst differences ($p<0.05$) between the Romani and native populations. Also, significant pairwise Fst differences are obtained and on the level of the HVSI sequences.

Multi-dimensional scaling plot performed on pairwise Fst values based on *CYTB* sequences (Fig. 2) and HVSI sequnces (Fig. 3). Figure 2 shows clear genetic division between the Roma and their neighbouring populations. This analysis shows that the Romani subpopulations are clearly separated in two clusters, which indicates the fact of substructuring within the Romani population, which is consistent with the obtained high values of Fst (0.11425) on the subpopulation level (the Roma 1 and the Roma 2). Plot based on the HVSI sequences (Fig. 3) shows similar relations of population grouping. However, for a more precise approach more samples should be analyzed.

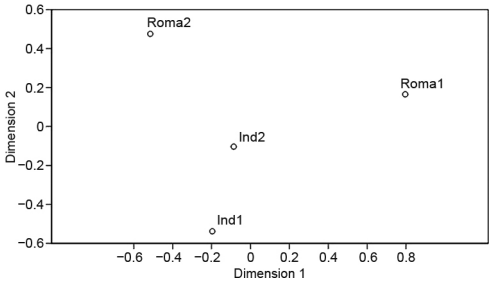

Fig. 2. Multi-dimensional scaling plot of populations according to pairwise FST values based on *CYTB* sequences

Roma 1 and 2 – Roma populations; Ind1 and 2 – native populations

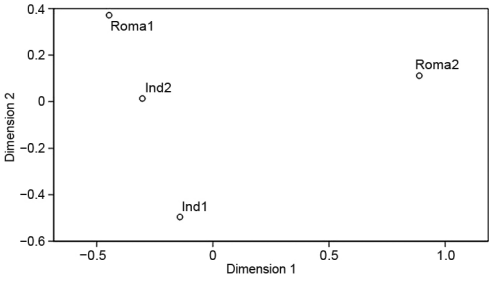

Fig. 3. Multi-dimensional scaling plot of populations according to pairwise FST based on HVSI sequences

Roma 1 and 2 – Roma populations; Ind1 and 2 – Native populations

Variation analysis of the CYTB sequences of the Bosnian Roma population and comparative Asian populations

In our study, the analysis of polymorphisms of the *CYTB* sequences indicates the presence of some nucleotide variations such as T14783C, G15043A and G15301A (with a higher frequency in comparison to other polymorphisms) only in the samples of the Romani group which are predominantly observed in

Table 5. Distribution of population-specific polymorphisms of the *CYTB* gene observed in comparative populations

rCRS position	Nucleotide change	Amino-acid substitution	Rom*	Uyg[1]	Han[2]	Chi[2]	Th[2]	Ph[2]	Vn[2]
14783	T→C	Leu→Leu	0.2051	0.3625	0.4366	0.5536	0.4500	0.3429	0.4545
15043	G→A	Gly→Gly	0.2051	0.3625	0.4366	0.5536	0.4500	0.3429	0.4545
15301	G→A	Leu→Leu	0.3333	0.3541	0.4366	0.5714	0.4500	0.4286	0.4545

*this study (Rom-Roma northeastern Bosnia); [1]Ablimit et al., 2013; (Uyg-Uyhgur population); [2]Hwa et al., 2010 (Han-Han Chinese; Chi-mainland Chinese; Th-Thai population; Ph-Filipino; Vn-Vietnam).

the Asian populations (Hwa et al. 2010; Ablimit et al. 2013) (Table 5).

Taking into consideration the mentioned fact, in the following parts of this research, the Roma population (Roma 1 and Roma 2) from our study and the previously studied Asian populations from a wider area of the Asia continent (Uyghur population from central Asia, Ablimit et al. 2013; the Chinese Han population, mainland Chinese, Thai popualtion, Vietnamese and Filipino Hwa et al. 2010) were additionally analysed by the parameters of interpopulation genetic diversity at the level of the *CYTB* sequences. Significant pairwise Fst differences ($p<0.05$) (Table 6) between the Roma population and different Asian populations (excluding the Uyghur population) were observed. The registered significant genetic differentiation based on the analysis of the *CYTB* sequences between the Roma population and different Asian populations illustrates the level of mtDNA genetic variation among populations based on ethnic isolation.

Some haplogroup-associated polymorphisms of the CYTB region and the HVSI of the mtDNA

In our study, additional intention was focused on the analysis of haplogroup-associated polymorphisms within the *CYTB* sequence (the region 14747–15887) in the context of providing additional information when the haplogroups are defined only by the HVSI or HVSI/II the CR of the mtDNA. Table 7 shows polymorphisms identified in the region of the *CYTB* sequences and the HVSI of the mtDNA with the corresponding haplogroups for 26 samples.

Table 6. Matrix of genetic differentiation (pFST) between comparative populations based on the *CYTB* sequences

	Rom*	Uyg[1]	Han[2]	Chi[2]	Th[2]	Ph[2]	Vn[2]
Rom*							
Uyg[1]	0.01101						
Han[2]	0.11011	0.00567					
Chi[2]	0.13355	0.00765	0.00358				
Th[2]	0.10916	−0.00072	−0.00291	0.00336			
Ph[2]	0.10654	0.00549	0.00635	0.01001	0.00525		
Vn[2]	0.12967	0.00405	0.00046	−0.00556	−0.00092	0.01069	

*this study (Rom-Roma northeastern Bosnia); [1]Ablimit et al., 2013; (Uyg-Uyhgur population); [2]Hwa et al., 2010 (Han-Han Chinese; Chi-mainland Chinese; Th-Thai population; Ph-Filipino; Vn-Vietnam)

Table 7. Haplogroup-associated polymorphisms of the *CYTB* sequences (region 14747–15887) and the HVSI CR mtDNA

Haplotypes of the region 14747–15887	Haplotypes of the HVSI (+16000)	RFLP polymorphisms	Haplogroup	Samples
C14766T, **T14783C, G15043A,** *T15289C,* G15301A, A15326G	129-223-230-233-304-344	10871, 10397	M35b2	7[1]
C14766T, **T14783C, G15043A,** *T15289C***,** G15301A, A15326G	129-223-230-233-304-344	10871, 10397	M35b2	28[1]
C14766T, **T14783C, G15043A,** *T15289C,* G15301A, A15326G	129-223-230-233-304-344	10871, 10397	M35b2	10[1]
C14766T, **T14783C, G15043A,** *T15287C,* G15301A, A15326G	129-223-291-298	10871, 10397	M5a1	15[1]
C14766T, **T14783C, G15043A**, *T15287C***,** G15301A, A15326G	129-144.1-223-291-298	10871, 10397	M5a1	16[1]
C14766T, **T14783C, G15043A,** *T15287C,* G15301A, A15326G	129-223-291-298	10871, 10397	M5a1	17[1]
C14766T, **T14783C, G15043A,** *T15287C***,** G15301A, A15326G	129-223-291-298	10871, 10397	M5a1	20[1]
C14766T, **T14783C, G15043A**, *T15287C,* G15301A, A15326G	129-223-291-298	10871, 10397	M5a1	31[1]
C14766T, **T14783C, G15043A,** *T15287C***,** G15301A, A15326G	129-223-291-298	10871, 10397	M5a1	4[1]
C14766T, **T14783C, G15043A,** *T15287C,* G15301A, A15326G, G15806A	129-223-291-298	10871, 10397	M5a1	43[1]
C14766T, **T14783C, G15043A,** *T15287C,* G15301A, A15326G	129-223-291-298	10871, 10397	M5a1	45[1]
C14766T, **T14783C, G15043A,** *T15287C,* G15301A, A15326G	129-223-291-298	10871, 10397	M5a1	48[1]
C14766T, **T14783C, G15043A,** *T15287C,* G15301A, A15326G	129-223-291-298	10871, 10397	M5a1	49[1]
C14766T, *A15034G,* **T15310C,** A15326G, *A15397G*	126-189-223-278	14766, 1715	X2e	37[1]
C14766T, *A15034G*, **T15310C,** A15326G, *A15397G*	126-189-223-278	14766, 1715	X2e	41[1]
C14766T, *A15034G,* **T15310C,** A15326G, *A15397G*	126-189-223-278	14766, 1715	X2e	42[1]
C14766T, *A15034G*, **T15310C,** A15326G, *A15397G*	126-189-223-278	14766, 1715	X2e	11[1]
C14766T, ***A14793G***, A14841C, G14846A, G14859A, A15326G	256-270	14766, 12308	U5a	n28
C14766T, ***A14793G*,** G15077A, A15218G, A15326G	256-270	14766, 12308	U5a1	n67
C14766T, **T14798C**, A15326G	16224-16311	14766, 12308	K/K	23[1]
C14766T, **T14798C**, A15326G	16093-16224-16311	14766, 12308	K	n10
C14766T, **T14798C,** *G15257A,* A15326G	16224-16311-16319	14766, 12308	K1b	33[1]

Haplotypes of the region 14747–15887	Haplotypes of the HVSI (+16000)	RFLP polymorphisms	Haplogroup	Samples
C14766T, **T14798C,** *G15257A,* A15326G	16224-16311-16319	14766, 12308	K1b	35[1]
C14766T, T14798C, **G14905A**, **C15452A**, C15527A, C15538A, **A15607G**	1609 16126-16294-16296	14766, 15606	T2 +16296	n43
C14766T, T14798C, A15326G, **C15452A**	T16063C,C16069T, T16126C, C16348T	14766, 13704	J1c (MF969048 (GenBank)	n19
C14766T, C14947A, A14999G, ***T15067C***, A15326G, **C15452A**	C16069T, T16126C, G16145A, T16172C, C16222T, C16261T	14766, 13704	J1b1	n42

*boldface – basal mutation of the haplogroup; italic – specific mutation of the subhaplogroup; [1] – samples with data of the polymorphisms of the HVSI region and RFLP polymorphisms took from study Ahmić et al.2018.

Other studied samples belong predominantly the haplogroup H and had insufficient number of the polymorphic positions within the *CYTB* region for discussion. In our study, 13 mtDNA sequences which belong to the haplogroup M have common polymorphisms such as C14766T, T14783C, G15043A, G15301A, A15326G. According to phylogenetic tree (van Oven and Kayser 2009), haplogroup M is primarily defined by the polymorphisms T14783C and G15043A in the region of the *CYTB* sequences. It was observed that the samples of the subhaplogroup M5a1, beside basal motiv 16129-16223-16291-16298 in the HVSI and haplogroup-defining polymorphism T14783C and G15043A, also have haplogroup-specific mutation T→C at site 15287 (samples 4, 15, 16, 17, 20, 31, 43, 45, 48 and 49). The members of the subhaplogroup M35b2 (samples 10, 7 and 28) have basal motiv 16129-16223-16230-16233-16304-16344 and share haplogroup-specific polymorphism T→C at site 15289. Furthermore, the polymorph position at T15310C in the region of the *CYTB* sequences is basal for the determination of the haplogroup X2e. Beside mutation T15310C, the samples of haplogroup X2e have two specific polymorphisms A15034G and A15397G (Table 7). Also, the results in our study confirm the data reported in the study by Kong et al (2003), Herrnstandt et al. (2002), Finnila (2001) that the region of the *CYTB* can provide additional information for defining some West Euro-Asian haplogroups. For instance, the haplogroup JT is characterized by transversion C→A on position 15452 (n43, n19, n42), the haplogroup T is additionally characterized by the polymorphisms G14905A and A15607G, and subhaplogroup J1b1 transition T15067C. The haplogroup K is characterized by polymorph position T14798C, and subhaplogroup the K1b transition G→A on 15257 (samles 33 and 35). Beside standard motiv 16256–16270 in the HVSI, polymorph position A14793G in the region of the *CYTB* gene is present in the subhaplogroup U5a. Phylogenetic tree of the haplogroups indentified based on data *CYTB* sequences, HVSI sequnces (Table 7) and data HVSII (previous studies, Ahmić et al. 2018) for 26 samples is showed in Fig. 4.

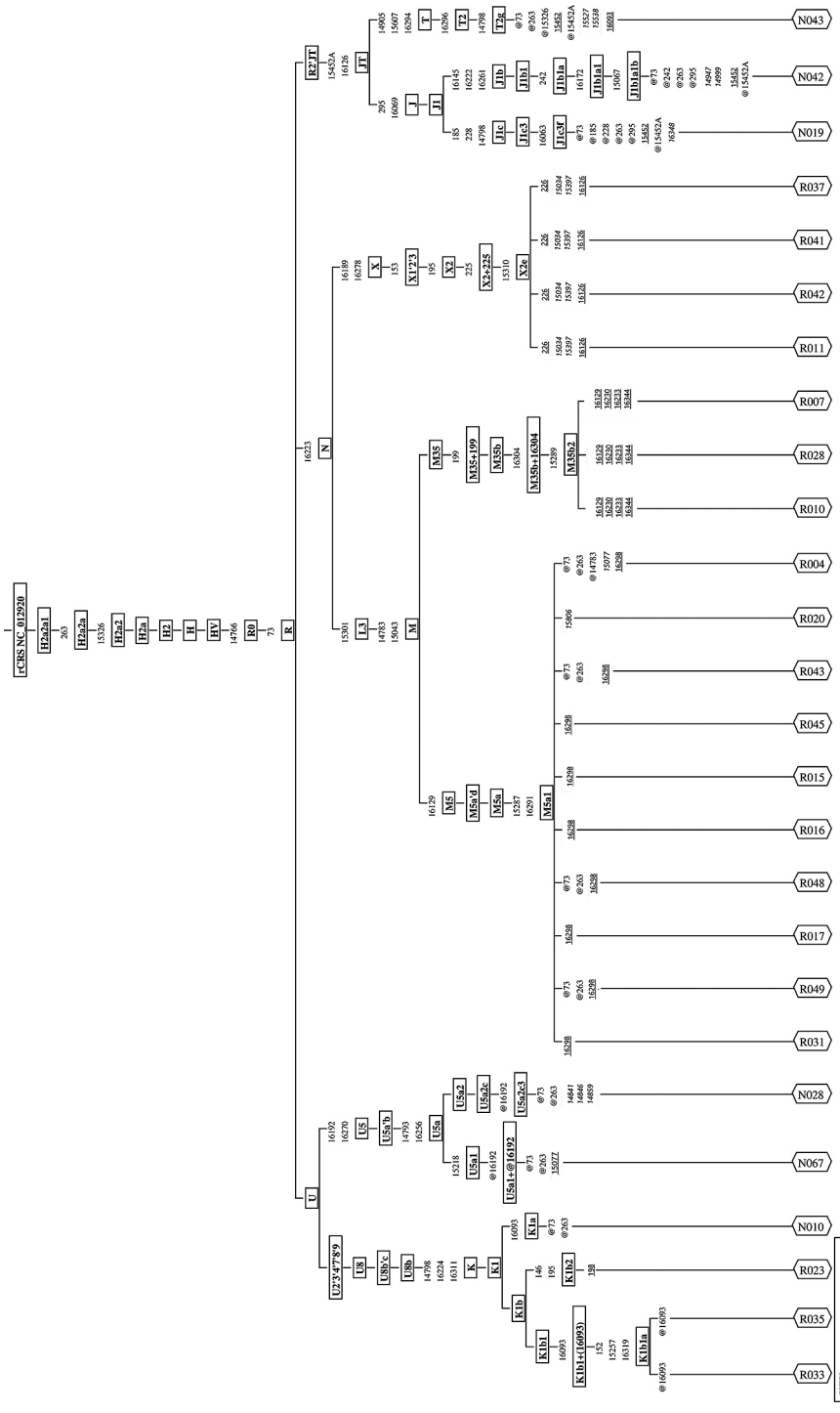

Fig. 4. Phylogenetic tree of the haplogroups identified based on data *CYTB* sequences (this study) and the HVSI/HVSII data (Ahmić et al. 2018) in 26 samples

Discussion

The results of our study indicate that the *CYTB* sequences are very polymorphic. The values of the molecular diversity parameters in our research (Table 3) confirm the results of previous studies (Hwa et al. 2010; 2011; Ablimit et al. 2013; Farghadani and Babadi 2015; Amer et al. 2015) indicating that *CYTB* gene is more polymorphic than expected for this relatively conservative segment of the coding region of the mtDNA. The majority of the sequence variations of the *CYTB* gene were non-synonymous with changes in amino acid products and viewed as polymorphisms without functional disturbances (Brown et al. 1992). The transition G15257A found in two samples in a group of the Roma has been reported in association with the primary mutations causing LHON (Leber's hereditary optic neuropathy) (Brown et al. 1992; Huoponen et al. 1993).

Our results indicate the presence of population-specific polymorphisms in the studied categories of the population. This is specially illustrative in the presence of polymorphisms with a relatively high frequency such as 14783T/C (20.51%), 15043G/A (20.51%) and 15301G/A (33.33%), which were observed only in the Roma population (Table 2). Interestingly, this triad of polymorphisms was most commonly observed in the populations of the Asian continent, including the Uyghurs population from central Asia (Ablimit et al. 2013), the Malaysian population (Farghadani and Babadi 2015), the Chinese Han population, mainland Chinese population and the populations from Taiwan (Hwa et al., 2010). Mutations at these three positions were seldom observed in the Caucasian population (Ablimit et al. 2013), and they were not observed in our native (non-Roma) Bosnian population. The results in our study reconfirm the significance of the triad of polymorphism T14783C-G15043A-G15301A as a potential ethnic classification signal of the Asian populations (Lee et al. 2002; Hwa et al. 2010; Ablimit et al. 2013), as illustrative in the presence of this triad of polymorphisms in the Roma population in our study, indicating the influence of Asian component of the gene pool on the genetic structure of the studied population of the Roma, which is consistent with the results of previous study based on the control region of the mtDNA sequences (Ahmić et al. 2018). These data can generally be associated with the genetic history and origin of the Roma from Indian subcontinent (Fraser, 1992; Greshem et al. 2001; Marushiakova and Popov 2001a, 2001b; Mendizabal et al. 2011, 21012; Go´mez-Carballa et al. 2013; Moorjani et al. 2013).

According to the analysis of the genetic structure of the studied populations, the results indicate that the *CYTB* sequences can be usable in terms of assessment of the genetic differentiation of the populations. Namely, similar relations of pairwise Fst differences ($p<0.05$) (Table 4) were observed on level of the *CYTB* sequences and on level of the HVSI sequences between the Roma populations and native populations. Observed significant pairwise Fst differences correspond to the level of mtDNA genetic variation based on different gene pool and genetic origin of the studied populations of the northeastern Bosnia. The data suggest that each individual subpopulation (under the categories of Roma and native populations) is characterized by their own local genetic identity, which is probably a consequence of the reproductive

isolation, a high level of endogamy and genetic drift present in the Roma populations, and geographic isolation of the native populations, which come from the isolated villages of mountainous region. Although the results of our study support the scientific knowledge that the level of genetic variations and subdivision among Roma populations is probably based on the effects of isolation, endogamy and genetic drift and also on the fact that geographic isolation can affect the gene pool of the isolated populations (native populations in our study), more samples should be analyzed for a more precise approach. Our suggestion for the future research studies based on the estimation of efficiency using the *CYTB* gene in population studies is to include a larger number of samples and populations.

One of the aims of our study was focused on the association of haplogroup-defining and haplogroup-specific polymorphisms of the *CYTB* sequence (region 14747–15887) and their combination with basal polymorphisms of the control region (HVSI) of the mtDNA, what can be a benefit for providing some additional information for a more precise determination of the haplogroups, especially in the cases when it is not able to use complete mtDNA. Beside specific motiv of the HVS-a of CR region of the mtDNA, a great number of highly informative coding-regions of polymorphisms for defining haplogroup exist (Finnilä et al. 2001; Herrnstadt et al. 2002; Yao et al 2002; Kong et al. 2003; Takana et al. 2004; Pereira et al. 2005). Our study also indicates that the *CYTB* gene region (14747–15887) contains some haplogroup-specific polymorph positions which can support branches of the Asian mtDNA phylogeny, such as polymorphism T15287C which defines subhaplogroup M5a1 and polymorphism T15289C which defines subhaplogroup M35b2. These results agree with the study by Kong et al. (2003) which indicate that the polymorphisms within the region 14576–16047 (including the *CYTB* gene region) are very informative for defining the East Asian haplogroups (such as M10, B5a, B4b, B4c, C, G1, G1a). Also, beside the polymorphism at T15310C within the *CYTB* gene which is basal for determination of the haplogroup X2e of Indian/European origin, two specific mutations A→G were observed on sites 15034 and 15397 which are present in all members of haplogroup X2e, but it was not previously reported in Phylotree (van Oven and Kayser 2009). The database of complete mtDNA, mtDNA Community Logan DNA Project (http://www.ianlogan.co.uk/sequences_by_group/haplogroupselect.htm) reported these polymorphisms in complete sequence of the mtDNA of one individual in the Hungarian population (accession number MG952802, Malarychuk et al. 2018). We suggest that polymorphisms A15034G and A15397G can probably be useful in mtDNA phylogeny of the haplogroup X2e, but the analysis of a larger number of mtDNA sequences of the haplogroup X2e is necessary for a better informativeness of these positions. Also, the data in our study indicate that certain nucleotide polymorphisms of the *CYTB* region can be useful in determination of the haplogroup status of the West Euroasian mtDNA phylogeny, as it is reported in the previous studies (Herrnstandt et al 2002; Finnila 2001; Kong et al 2003). Using the reported data of the studies based on identification of haplogroup coding-region polymorphisms of the mtDNA and phylogenetic tree of global human mitochondrial variation (van Oven and Kayser 2009), it has been observed that

the region of *CYTB* gene can be a benefit in providing some additional information in mtDNA phylogeny of the haplogroup.

This study offers first preliminary data on the diversity of human *CYTB* gene in the populations from northeastern Bosnia, serving as a supplement of the database on the evaluation of the usefulness of the *CYTB* sequence polymorphisms in geographical/ethnic classification of populations, estimation of the genetic diversity, forensic casework as well as haplogroup differentiation. Although mtDNA analysis is directed towards sequencing of the complete mtDNA molecule, the analysis of *CYTB* sequence in combination with HVSI or HVSI/II of the CR, can offer some useful additional information for population studies. The results of this study indicate the effectiveness of the *CYTB* gene analysis in identification of individuals and categorization of different populations, the analysis of the genetic structure of populations as well as the analysis of the mtDNA phylogeny of the haplogroups. However, for a more precise approach more samples should be analyzed.

Acknowledgements

We thank all who voluntarily gave samples of their buccal mucosa and to all institutions that participated in their collection.

This work was supported by the Ministry of Education, Science, Culture and Sport of Tuzla Canton, Bosnia and Herzegovina. This study was conducted within the scientific project titled "Application of the cytochrome b in the assessment of the diversity in Bosnian and Herzegovinian population" (number 01/2-3680-IV/16).

Authors' contributions

AA the population genetic analysis and discussion of the results; NP population genetic analysis and discussion of the results; BK and LL conducted DNA analysis of the samples; IM collected data and determination of polymorphisms; EH determination of polymorphisms; AI collected data, techical prepariation figures and tables.

Conflict of interest

The authors report no conflicts of interest.

Corresponding author

Adisa Ahmic, Department of Biology, Faculty of Natural Sciences, University of Tuzla, Univerzitetska 4, 75000 Tuzla, Bosnia and Herzegovina
e-mail: adisaahmic555@gmail.com

References

Ablimit A, Qin W, Shan W, Wu W, Ling F, Ling KH, et al. 2013. Genetic diversities of cytochrome B in Xinjiang Uyghur unveiled its origin and migration history. BMC Genetetics 1471-2156/14/100.

Ahmic A, Kalamujic B, Ismailovic A, Lasic L, Halilovic E, Mujkic I, et al 2018. Mitochondrial DNA diversity of the Roma from northeastern Bosnia, Bosnia and Herzegovina. HOMO-Journal of Comparative Human Biology 69:347–56.

Anderson S, Bankier AT, Barrell BG, DE Bruijn MH, Coulson AR, Droin J, et al. 1981. Sequence and organization of the human mitochondrial genome. Nature 290:457–65.

Andrews RM, Kubicka I, Chinnery PF, Lightowlers RN, Turnbull DM, et al. 1999. Reanalysis and revision of the Cambridge

reference sequence for human mitochondrial DNA. Nat Genet 23:147.

Amer SAM, AlHothali B, Tubaigy S. 2015. Possible Application of Cytochrome b Gene for Human Identification. AJFSFM 1:244–48.

Bandelt HJ, Forster P, Rohl A. 1999. Median–joining networks for inferring intraspecific phylogenies. Mol Biol Evol 16(1):37–48.

Bandelt HJ, Lahermo P, Richards M, Macaulay V. 2001. Detecting errors in mtDNA data by phylogenetic analysis. Int J Legal Med 115:64–9.

Brown MD, Voljavec AS, Lott MT, Torroni A, Yang CC, Wallace DC. 1992. Mitochondrial DNA complex I and III mutations associated with Leber's hereditary optic neuropathy. Genetics 130 (1):163–73.

Brown WM, George, MJr, Wilson AC. 1979. Rapid evolution of animal mitochondrial DNA. Proc Natl Acad Sci USA 76:1967–71.

Excoffier L, Laval G, Schneider S. 2005. Arlequin (version 3.0): An integrated software package for population genetics data analysis. Evol Bioinform Online 1:47–50.

Farghadani R and Babadi AA. 2015.Nucleotide variation of the mitochondrial cytochrome B gene in the Malay population. Rom J Leg Med 23(1):57–60.

Finnilä S, Lehtonen MS, Majamaa K. 2001. Phylogenetic network for European mtDNA. Am J Hum Genet 68:1475–84.

Fraser A. 1992. The peoples of Europe: The Gypsies. 1st edition. Oxford: Blackwell Publishers.

Gómez-Carballa A, Pardo-Seco J, Fachal L, Vega A, Cebey M, Martinón-Torres N, et al 2013. Indian Signatures in the Westernmost Edge of the European Romani Diaspora: New Insight from Mitogenomes. Plos One 8 (10), e75397.

Gresham D, Morar B, Underhill PA, Passarino G, Lin AA, Wise Ch, et al 2001. Origins and Divergence of the Roma (Gypsies). Am J Hum Genet 69:1314–31.

Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium 41:95–8.

Herrnstadt C, Elson JL, Fahy E, Preston G, Turnbull DM, Anderson Ch, et al. 2002. Reduced-Median-Network Analysis of Complete Mitochondrial DNA Coding-Region Sequences for the Major African, Asian, and European Haplogroups. Am J Hum Genet 70:1152–71.

Huoponen K, Lamminen T, Juvonen V, Aula P, Nikoskelainen E, Savontaus MR. 1993. The spectrum of mitochondrial DNA mutations in families with Leber hereditary optic neuroretinopathy. Hum Genet 92 (4):379–84.

Hwa HL, Ko TM, Chen YC, Chang YY, Tseng LH, Su YN, et al. 2010. Study of the cytochrome b gene sequence in populations of Taiwan. Journal Forensic Sciences 55 (1):167–70.

Hwa HL, Lin C-Y, Ko T-M, Yin H-Yi, Tseng Li-H T, Su Yi-N, et al 2011: Analysis of MTCOI and MTCYB Sequence Variations in Eight Population Groups Living in Taiwan. Rom J Leg Med 19:219–28.

Kloos-Brandstätter A, Pacher D, Schoenherr S, Weisensteiner H, Binna R, Specht G, et al 2011. HaploGrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. Hum Mutat 32:25–32.

Kong QP, Yao YG, Liu M, Shen SP, Chen C, Zhu CL, Palanichamy MG, Zhang YP. 2003. Mitochondrial DNA sequence polymorphisms of five ethnic populations from northern China. Hum Genet 113:391–405.

Lee SD, Lee YS, Lee JB. 2002. Polymorphism in the mitochondrial cytochrome B gene in Koreans. An additional marker for individual identification. Int J Legal Med 116 (2):74–8.

Malyarchuk BA 2002. The Effect of the Nucleotide Context on Induction of Mutations in Hypervariable Segment 1 of Human Mitochondrial DNA. Molecular Biology 36 (3):322–26.

Malyarchuk BA, Derenko M, Grzybowski T, Perkova M, Rogalla U, Vanecek T . 2010.

The peopling of Europe from the mitochondrial haplogroup U5 perspective. PloS One doi: 10.1371/journal.pone.0010285.

Malyarchuk BA 2013. Mutational Process in ProteinCoding Genes of Human Mitochondrial Genome in Context of Evolution of Homo Genus. Molecular Biology 47 (6):807–13.

Malyarchuk B, Derenko M, Denisova G, Litvinov A, Rogalla U, Skonieczna K, et al 2018. Whole mitochondrial genome diversity in two Hungarian populations. Mol Genet Genomics 293(5):1255–63.

Marushiakova E, Popov V. 2001a. Gypises in the Ottoman Empire. A contribution to the history of the Balkans. UK, Hatfield: University of Hertfordshire Press.

Marushiakova E, Popov V. 2001b. Historical and Ethnographic Backgraund. Gypsies, Roma, Sinti. In: W Guy, editor. Between Past and Future: the Roma of Central and Eastern Europe. UK, Hatfield: University of Hertfords Press.

Mendizabal I, Valente C, Gusmão A, Alves C, Gomes V, Goios A, et al. 2011. Reconstructing the Indian origin and dispersal of the European Roma: a maternal genetic perspective. PLOS.one 6: e15988.

Mendizabal I., Lao, O., Marigorta U.M., Wollstein A., Gusmão L., Ferak V., et al 2012. Reconstructing the Population History of European Romani from Genome-wide Data. Current Biology 22:2342–49.

Meyer S, Weiss G, von Haeseler A. 1999. Pattern of nucleotide substitution and rate heterogeneity in the hypervariable regions I and II of human mtDNA. Genetics 152:1103–10.

Miller SA, Dykes DD, Polesky HF. 1998. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids 16:1215.

Moorjani P, Patterson N, Po-Ru L, Lipson M, Kisfali P, Melegh BI, et al. 2013. Reconstructing Roma History from Genome-Wide Data. Plos One 8(3):e58633.

Pereira L, Richards M, Goios A, Alonso A, Albarrán C, Garcia O, et al. 2005. High-resolution mtDNA evidence for the late-glacial resettlement of Europe from an Iberian refugium. Genome Res 15:19–24.

Richards M, Macaulay V, Hickey E, Vega E, Sykes B, Guida V, et al 2000. Tracing European founder lineages in the Near Eastern mtDNA pool. Am J Hum Genet 67:1251–76.

Simoni L, Calafell F, Pettener D, Bertranpetit J, Barbujani G. 2000. Geographic patterns of mtDNA diversity in Europe. Am J Hum Genet 66:262–78.

Santos C, Lima M, Montiel R, Francalacci P. 2004. Determination of Human European Mitochondrial DNA Haplogroups by Means of a Hierarchical Approach. Hum Biol 76 (3):432–53.

Takana M, Cabrera V, González A, Larruga J, Takeyasu T, Fuku N, et al. 2004. Mitochondrial Genome Variation in Eastern Asia and the Peopling of Japan Complete Mitochondrial DNA Analysis of Eastern Eurasian Haplogroups Rarely Found in Populations of Northern Asia and Eastern Europe. Genome Res 14:1832–50.

Tamura K. 2000. On the estimation of the rate of nucleotide substitution for the control region of human mitochondrial DNA. Gene 23:259(1–2):189–97.

Torroni A, Bandelt HJ, Macaulay V, Richards M, Cruciani F, Rengo C, et al. 2001. A signal from human mtDNA of postglacial recononization in Europe. Am J Hum Genet 69:844–52.

Tsai L, Lin C, Lee J, Chang J, Linacre A, Goodwin W. 2001. Sequence polymorphism of mitochondrial D-loop DNA in the Taiwanese Han population. Forensic Sci Int 119(2):239–47.

Van Oven M, Kayser M. 2009. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Hum Mutat 30(2):386–94.

Weir BS, Cockerham CC. 1984. Estimating F-Statistics for the Analysis of Population Structure. Evolution 38:1358–137.

Weissensteiner H, Pacher D, Kloos-Brandstätter A, Forer L, Specht G, Bandelt HJ, Kronenberg F, et al 2016. HaploGrep 2: mitochondrial haplogroup classification in

the area of hight-throughput sequencing. Nucleic Acids Res 44 (W1),W58–W63.

Yao YG, Kong QP, Bandelt HJ, Kivisild T, Zhang YP. 2002. Phylogeographic differentiation of mitochondrial DNA in Han Chinese. Am J Hum Genet 70:635–51.