Divergent Patrilineal Signals in Three Roma Populations

Maria Regueiro,[1] Aleksandar Stanojevic,[2] Shilpa Chennakrishnaiah,[1] Luis Rivera,[1] Tatjana Varljen,[2] Djordje Alempijevic,[2] Oliver Stojkovic,[2] Tanya Simms,[1,3] Tenzin Gayden,[1,3] and Rene J. Herrera[1]*

[1]*Department of Molecular and Human Genetics, College of Medicine, Florida International University, Miami, FL*
[2]*Institute of Forensic Medicine, School of Medicine, University of Belgrade, Belgrade, Serbia*
[3]*Department of Biological Sciences, Florida International University, Miami, FL*

KEY WORDS Y-chromosome; Romani; Indian; phylogenetic analyses; Y-SNP; Y-STR

ABSTRACT Previous studies have revealed that the European Roma share close genetic, linguistic and cultural similarities with Indian populations despite their disparate geographical locations and divergent demographic histories. In this study, we report for the first time Y-chromosome distributions in three Roma collections residing in Belgrade, Vojvodina and Kosovo. Eighty-eight Y-chromosomes were typed for 14 SNPs and 17 STRs. The data were subsequently utilized for phylogenetic comparisons to pertinent reference collections available from the literature. Our results illustrate that the most notable difference among the three Roma populations is in their opposing distributions of haplogroups H and E. Although the Kosovo and Belgrade samples exhibit elevated levels of the Indian-specific haplogroup H-M69, the Vojvodina collection is characterized almost exclusively by haplogroup E-M35 derivatives, most likely the result of subsequent admixture events with surrounding European populations. Overall, the available data from Romani groups points to different levels of gene flow from local populations. Am J Phys Anthropol 000:000–000, 2010. © 2010 Wiley-Liss, Inc.

The ancestry of the Romani people has been the subject of much debate since their initial entry into Europe during the late 13th century. Although originally believed to be of Egyptian descent (hence their more common name, "Gypsies"), similarities observed between the Romani language (of Indo-Aryan stock) and those present in the Indian subcontinent (in particular, Hindi) have suggested a more likely medieval Central Indian ancestry for this group (Fraser, 1992; Liégeois, 1994), possibly originating in present-day Rajasthan, India (Hancock, 2002). Today, several million Romanis live in Central Eastern and Western Europe as well as in the Americas. Throughout their diaspora from Northern India, Romanis have been perceived as aliens and have endured discrimination, persecution and slavery.

In addition to language, this group also displays other cultural influences characteristic of Indian populations: the social status attained by Romanis (as entertainers) throughout Europe is reminiscent of the Shudra caste of India (Hancock, 2002), while the endogamous professional-group organization exhibited by Romani groups most closely resembles that of the Indian *jatis* (Fraser, 1992; Marushiakova and Popov, 1997).

After their initial dispersal from the Indian subcontinent (500–1000 AD), Romanis are believed to have migrated through portions of Central Asia (Afghanistan and Persia) to Armenia in the north and the Near East in the south (Fraser, 1992). From the Near East, this group moved in two directions; either toward North Africa with subsequent dispersion through the Strait of Gibraltar for entry into the Iberian Peninsula, or across Anatolia (via the Strait of Bosporus) and into the Balkans (between the 11th and 12th centuries) with a later expansion into Western Europe (completed by the 15th century) (Fraser, 1992).

The Roma are classified as a subgroup of the Romani people that live primarily in Central and Eastern Europe as well as in the Balkan Peninsula and Western Anatolia (Hancock, 2002). Of these regions, the greatest genetic diversity is reported in the Balkans where numerous Romani populations with well-defined social boundaries reside (Marushiakova and Popov, 1997). According to the 2002 Census (Statistical Office of the Republic of Serbia, 2003), the number of Roma individuals living in the regions of Belgrade and Vojvodina is 19,191 and 29,057, respectively, whereas 23,512 individuals have been reported in Kosovo in the 2006 Census (Statistical Office of Kosovo, 2008). Most of the Roma in these three locales are Orthodox or Catholics with a small minority being Muslim. In Belgrade and Vojvodina, Romas speak the Gurbet (Ackerley, 1941) and Gurbet-Rabeste (Boretzky, 1986) dialects, respectively (both of which belong to the Southern Vlax branch of the Romani language) whereas in Kosovo, their dialect is Prizren (Balkan branch) (Boretzky, 1999).

Genetic studies delineating the ancestry of Romani populations provide strong evidence in favor of the common descent of all Romanis regardless of their group identity, country of residence, and/or rules of endogamy (Kalaydjieva et al., 2005). In addition, the Romanis have been characterized as a conglomerate of genetically isolated founder groups (Kalaydjieva et al., 2001). To date, Y-chromosome and mtDNA markers have been used to

Additional Supporting Information may be found in the online version of this article.

Maria Regueiro and Aleksandar Stanojevic contributed equally to the study.

*Correspondence to: Rene J. Herrera, Department of Molecular and Human Genetics, College of Medicine, Florida International University, University Park, OE 304, Miami, FL 33199.
E-mail: herrerar@fiu.edu

Received 6 April 2010; accepted 1 June 2010

DOI 10.1002/ajpa.21372
Published online in Wiley Online Library (wileyonlinelibrary.com).

examine the origins and diversification of paternal and maternal lineages, respectively, in Romani populations. These reports indicate that all Romani communities examined share Indian-specific lineages, namely Y-haplogroup H (defined by the M69 mutation) (Gresham et al., 2001; Kalaydjieva et al., 2001; Gusmão et al., 2008) and mtDNA haplogroup M5 (Gresham et al., 2001; Malyarchuk et al., 2006, 2008).

Along with haplogroup H, which accounts for almost 50% of their male genetic component, Romanis also possess G-M201, I-M258, J-M304, R-M207 and E-M35 Y-chromosomes, reflecting genetic influences from the surrounding European populations (Gresham et al., 2001; Kalaydjieva et al., 2001, 2005; Gusmão et al., 2008; Klarić et al., 2009). Haplogroup H and its sub-clades, H1 and H2, are almost exclusive to the Indian subcontinent (26.4% of all Y-chromosomes) (Sengupta et al., 2006) including Sri Lanka (10.2%) (Kivisild et al., 2003) and Pakistan (2.5%) (Firasat et al., 2007). European haplogroups present in the Romani gene pool include those which have originated in Europe such as I and E-V13, as well as those believed to have arisen in Asia, namely G, J, and R (Semino et al., 2000, 2004; Kivisild et al., 2003; Battaglia et al., 2009). Haplogroups I and E-V13 are particularly abundant in the Balkans where they represent 36.3% (Battaglia et al., 2009) and 19.6% (Cruciani et al., 2007) of the total Y-chromosomes, respectively. Haplogroup R lineages are common throughout Europe and the Indian subcontinent comprising about 50% of the European (Semino et al., 2000) and 26% of the Indian (Sengupta et al., 2006) patrilineages. The J-M304 clade has been detected at high frequencies in the Middle East, North Africa, Europe, Central Asia, Pakistan and India, whereas haplogroup G is found predominantly in the Middle East, the Mediterranean, and the Caucasus Mountains (Karafet et al., 2008).

In addition to the Y-chromosome and mtDNA data, genetic ties between Romanis and subcontinental Indians have also been detected through analysis of autosomal components of the genome. For instance, the 1267delG mutation in the Cholinergic Receptor Nicotinic Epsilon (*CHRNE*) gene (which causes congenital myasthenia), originally identified in patients from India/Pakistan (Croxen et al., 1999), has also been observed in Romanis (Abicht et al., 1999). Similarly, Romani populations were also found to possess mutations in the *GJB2* gene (which codes for connexin26, a gap junction protein) (Minárik et al., 2003; Álvarez et al., 2005; Bouwer et al., 2007) known to cause autosomal recessive deafness in Indian families (Brown et al., 1996; Kelsell et al., 1997; Scott et al., 1998; Maheshwari et al., 2003). Also, Ali et al. (2009) have linked the presence of null mutations in the Latent Transforming growth factor Beta Binding Protein 2 (*LTBP2*) gene to Primary Congenital Glaucoma in four consanguineous families from Pakistan and in patients of Romani ancestry. The sharing of these genetic disorders among Romanis across Europe also suggests a common origin and subsequent founder effects by modest numbers of founder groups (Gresham et al., 2001; Kalaydjieva et al., 2001). This contention is compatible with a limited number of migrants partitioning from a specific caste or tribal group (Gresham et al., 2001). It has been estimated that the Romani as a group had its origin 32 to 40 generations ago, with subsequent secondary and tertiary founder episodes taking place ~16 to 25 generations ago (Morar et al., 2004).

Several sources describe Romanis as a mosaic of numerous geographically dispersed and genetically isolated founder populations (Fraser, 1992; Marushiakova and Popov, 1997; Kalaydjieva and Morar, 2003). This conclusion is consistent with the limited haplotype variance observed in Y-haplogroup H1-M82 and the reduced sequence diversity in mtDNA M lineages (Gresham et al., 2001). Furthermore, analysis of the rapidly mutating minisatellite MSY1 also supports genetic homogeneity among Romanis, as diversity values attained with this marker system are even lower than those of known geographically isolated populations, i.e., the Finns, Basques and Cook Islanders (Kalaydjieva et al., 2001). As such, this study was undertaken to examine the patrilineal influences present in three Roma populations residing in the Balkans (Belgrade and Vojvodina from Serbia, and Kosovo) and, subsequently, compare their genetic profiles to those present in other Romani collections.

MATERIALS AND METHODS

Sample collection

DNA samples from 88 healthy, unrelated Roma male donors, collected from three locations in the Balkans [Belgrade (n = 20), Vojvodina (n = 26), and Kosovo (n = 42)], were analyzed in the present study. Genealogical information from each donor was recorded for at least two generations to establish regional ancestry. All samples were procured with informed consent following the ethical guidelines stipulated by the research institutions involved in this project. Table 1 lists sample sizes, geographic locations, and linguistic affiliations for the aforementioned Roma populations as well as the reference collections utilized for phylogenetic comparisons.

Y-chromosome haplotyping

Fourteen bi-allelic markers were genotyped, following the hierarchy of the human Y-chromosome phylogeny, according to previously published protocols: YAP (Hammer and Horai, 1995) by PCR/AFLPs; M78, M34, M52, M82 (Underhill et al., 2001) and V13 (Cruciani et al., 2006) by PCR/RFLP; and M9 (Underhill et al., 1997), M35, M69, M96, M201, M207 (Underhill et al., 2001), M258 (Rootsi et al., 2004), and M304 (Cinnioğlu et al., 2004) by allele-specific PCR (Gayden et al., 2008). The markers investigated in this study were previously reported informative for Romani, European and Indian populations. Haplogroup designation is in accordance with Karafet et al. (2008) and the recent updates by Chiaroni et al. (2009).

DNA samples (N = 86) were also typed for 17 Y-STR loci (DYS19, DYS385a,b, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS635, GATA H4) using the AmpF*l*STR® Yfiler kit (Applied Biosystems, Foster City, CA). PCR was performed as described by the manufacturer (Applied Biosystems, 2006), and the amplicons generated were separated in an ABI Prism 310 Genetic Analyzer. The GeneMapper® software v3.2 was used to determine the fragment sizes and alleles were designated by comparison to an allelic ladder supplied by the manufacturer. Allelic nomenclature is in accordance with the recommendations of the International Society for Forensic Genetics (ISFG) (Gusmão et al., 2006).

TABLE 1. Populations analyzed

Population	Abbr.	*N*	Linguistic affiliation	Reference
NON-ROMANI				
South Central Asia (SCA)				
India	IND	728	Indo-European, Indo-Iranian	Sengupta et al., 2006
Pakistan	PAK	176	Indo-European, Indo-Iranian	Sengupta et al., 2006
Gujarat	GUJ	29	Indo-European, Indo-Iranian	Kivisild et al., 2003
Chenchu (Andhra Pradesh)	CHE	41	Dravidian	Kivisild et al., 2003
Koya (Andhra Pradesh)	KOY	41	Dravidian	Kivisild et al., 2003
Lambadi (Andhra Pradesh)	LAM	35	Indo-European, Indo-Iranian	Kivisild et al., 2003
Sinhalese (Sri Lanka)	SIN	39	Indo-European, Indo-Iranian	Kivisild et al., 2003
Pardhan (Andhra Pradesh)	PAR	128	Dravidian	Thanseem et al., 2006
Naikpod (Andhra Pradesh)	NAI	68	Dravidian	Thanseem et al., 2006
Andh (Andhra Pradesh)	AND	54	Unclassified	Thanseem et al., 2006
Southeastern Europe Slavic speaking (SEES)				
Croatians	CRO	108	Indo-European, Slavic	Barać et al., 2003
Herzegovinians	HER	141	Indo-European, Slavic	Peričić et al., 2005
Serbians	SER	113	Indo-European, Slavic	Peričić et al., 2005
Macedonians	MAC	79	Indo-European, Slavic	Peričić et al., 2005
Bosniacs	BOS	85	Indo-European, Slavic	Marjanovic et al., 2005
Slovenians	SLO	75	Indo-European, Slavic	Battaglia et al., 2009
Southeastern Europe non-Slavic speaking (SEENS)				
Kosovar Albanians	KOA	114	Indo-European, Albanian	Peričić et al., 2005
Albanians	ALB	30	Indo-European, Albanian	Bosch et al., 2006
Romanians	ROM	67	Indo-European, Italic	Bosch et al., 2006
Greece	GRE	76	Indo-European, Greek	Semino et al., 2000
Albanians from Former Yugoslavia Republic of Macedonia	AYR	64	Indo-European, Albanian	Battaglia et al., 2009
ROMANI				
Southeastern Europe				
Croatian Bayash (Baranja)	BAR	96	Indo-European, Indo-Iranian	Klarić et al., 2009
Croatian Bayash (Medjimurje)	MED	55	Indo-European, Indo-Iranian	Klarić et al., 2009
Macedonia Roma	MAR	57	Indo-European, Indo-Iranian	Peričić et al., 2005
Kosovo Roma	KOS	42	Indo-European, Indo-Iranian	Present study
Belgrade Roma	BEL	20	Indo-European, Indo-Iranian	Present study
Vojvodina Roma	VOJ	26	Indo-European, Indo-Iranian	Present study
Residing in Bulgaria				
Kalaidjii North	KAL	20	Indo-European, Indo-Iranian	Gresham et al., 2001
Feredjelli	FER	21	Altaic, Turkish	Gresham et al., 2001
Turgovzi	TUR	36	Altaic, Turkish	Gresham et al., 2001
Lom	LOM	39	Indo-European, Indo-Iranian	Gresham et al., 2001
Monteni	MON	17	Indo-European, Indo-Iranian	Gresham et al., 2001
Lingurari (North)	LIN	16	Indo-European, Indo-Iranian	Gresham et al., 2001
Intreni	INT	17	Indo-European, Indo-Iranian	Gresham et al., 2001
Northern Europe				
Lithuanian Romani	LIT	20	Indo-European, Indo-Iranian	Gresham et al., 2001
Western Europe				
Spanish Romani	SPA	27	Indo-European, Indo-Iranian	Gresham et al., 2001
Portuguese Romani	POR	126	Indo-European, Indo-Iranian	Gusmão et al., 2008

Data analyses

Haplogroup frequencies for the Belgrade, Kosovo, and Vojvodina collections, determined by the direct counting method (Li, 1976), were used to generate genetic (*Fst*) distances and associated *P*-values (based on 1000 permutations) with the Arlequin software package, ver. 3.11 (Excoffier et al., 2005). The same program was also employed to perform two sets of Analyses of Molecular Variance (AMOVA), to explore significant correlations between genetic diversity and either language or geography. Significance was assessed at $\alpha = 0.05$.

Multidimensional Scaling (MDS) analysis, which utilizes a pairwise *Fst* matrix, was performed with the statistical software package SPSS ver. 16.0 (SPSS Inc., Chicago, IL, USA). All *Fst* distances were estimated based on haplogroup frequencies at the resolution of the major branches (A-T) of the Y chromosome phylogram. In addition, to assess the Indian and European genetic contributions to the paternal gene pool of the three Roma collections, admixture analysis was conducted with SPSS utilizing the Weighted Least Squares method proposed by Long et al. (1991). In this analysis, groups of populations were used as parental sources in an attempt to establish regional connections as well as possible gene flow or shared ancestry among the three hybrid collections (Belgrade, Kosovo and Vojvodina).

Median-joining (MJ) networks ($\varepsilon = 0$) (Bandelt et al., 1999) were constructed using the NETWORK 4.5.1.6 software package available at http://www.fluxus-engineering.com. The network generated for the three Roma populations reported in this study is based on 15 Y-STR loci typed on the background of haplogroup H. The analysis was repeated using only seven of the 15 Y-STR markers (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392 and DYS393) to allow inclusion of the reference collections (Table 1) for comparative purposes. MJ networks were weighted according to their relative

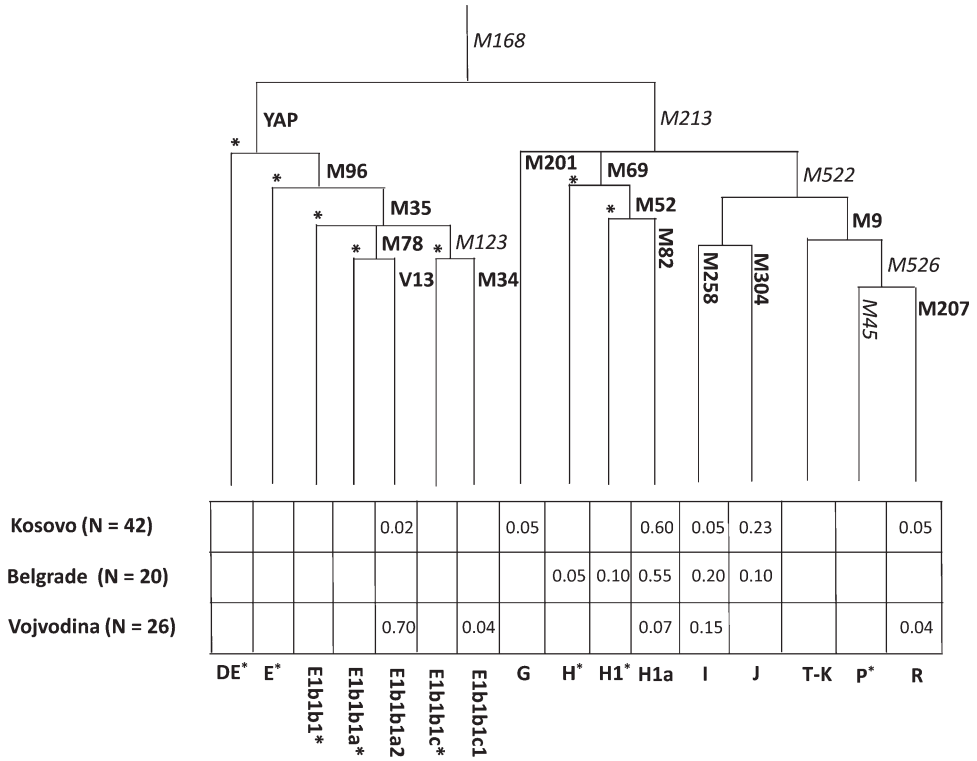

Fig. 1. The hierarchical phylogenetic relationships and frequencies of the major haplogroups observed in the three Roma populations reported in this study. N indicates the number of individuals examined. Markers not genotyped are shown in italics.

Y-STR variation within Y-haplogroup lineages (Martinez et al., 2007).

The coalescence times for the H haplogroup were estimated from microsatellite variation within the haplogroup as the average squared difference in the number of repeats between all the chromosomes and the founder haplotype, averaged over the number of microsatellite loci and divided by the mutation rate, with the SD calculated over the number of loci (Zhivotovsky et al., 2004; Zhivotovsky and Underhill 2005). We computed the median value of repeat scores to generate a median haplotype, which was then taken as a founder (Sengupta et al., 2006). In all cases, except for LOM and LIT, the median haplotype coincides with the Romani Modal Haplotype. Time estimates for all populations were calculated with the Romani Modal haplotype utilizing both the genealogical mutation rate of 2.5×10^{-3} per 35 years (Goedbloed et al., 2009) and the evolutionary effective mutation rate of 6.9×10^{-4} per 25 years (Zhivotovsky et al., 2004). For the Roma populations reported in this study, the expansion times for haplogroup H were calculated across the 15 Y-STR loci. The analyses were also repeated across the 7 Y-STR loci common to all collections included in this study to allow for comparison of divergence times among them. Time estimates were performed using the Network subprogram to estimate the age of the Romani Modal Haplotype (DYS19*15, DYS389I*14, DYS389II*16, DYS390*22, DYS391*10, DYS392*11, DYS393*12) (Gusmão et al., 2008) in all populations examined.

Microsatellite variances for haplogroup H were calculated using the vp equation of Kayser et al. (2001). Haplotype diversity indices were generated according to Nei (1987) using Arlequin ver. 3.11 and were based on 7 Y-STR loci (DYS19, DYS389 I/II, DYS390, DYS391, DYS392, and DYS393). The two DYS385 loci were not considered in the MJ networks, time estimations as well as haplotype diversity calculations given the duplicity of the sequence.

RESULTS

Y-chromosome haplogroup distribution

A total of six haplogroups were identified among the three Roma populations presented in this report, namely E, G, H, I, J, and R. Hierarchical phylogenetic relationships among Y-chromosome haplogroups along with their frequencies are displayed in Figure 1. Interestingly, only two haplogroups (H1a and I) are shared across the three Roma communities.

Haplogroup H-M69 constitutes 70%, 60%, and 7% of the Belgrade, Kosovo and Vojvodina samples, respectively. Specifically, the H1a sub-clade, defined by the M82 mutation, predominates over all other Y-lineages in the Kosovo (60%) and Belgrade (55%) collections and is detected at levels comparable to those exhibited by other Roma populations residing in the Balkan Peninsula, including the Macedonian Roma (59.6%) (Peričić et al., 2005) and the Bayash Roma from Croatia (50.3%) (Klarić et al., 2009).

A contrasting distribution pattern for haplogroup E1b1b1-M35 (the only branch of haplogroup E-M96 detected in this survey) was observed. Of the three collections examined, the Vojvodina population, which exhibits the lowest frequency of haplogroup H, is characterized almost exclusively by E-M35 Y-chromosomes (E1b1b1a-M78 = 70% and E1b1b1c-M123 = 4%), while few to no individuals with this haplogroup were observed in the Kosovo

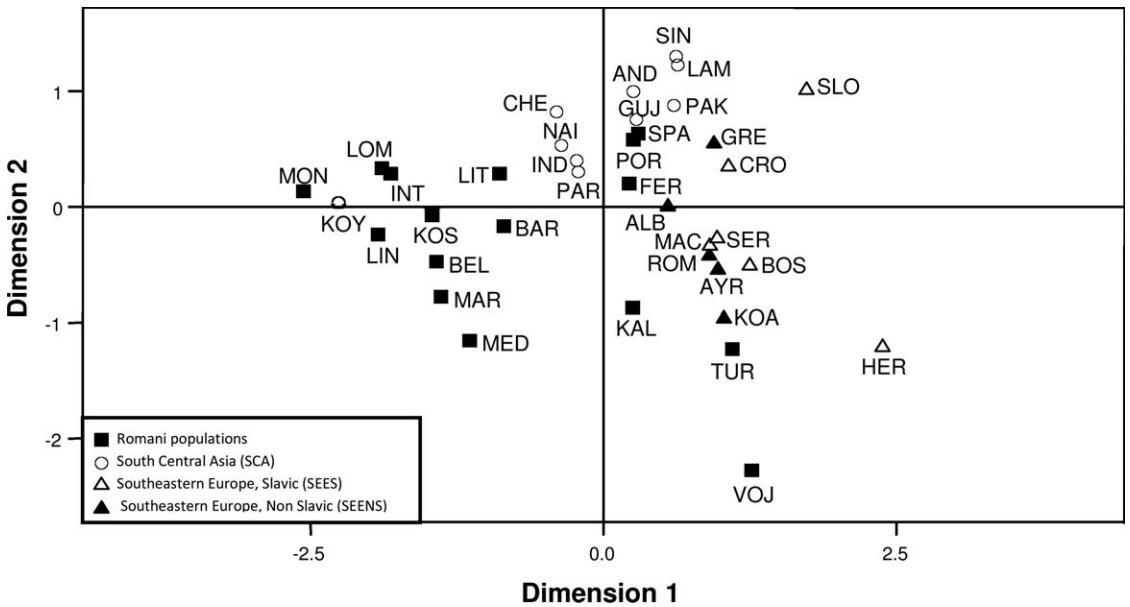

Fig. 2. MDS plot based on the Y-haplogroup frequency data of 37 populations. Stress value = 0.14481. References, number of samples, and name codes for all populations are listed in Table 1. Black boxes represent Romani populations, white circles South Central Asia populations, black triangles Non-Slavic South Eastern Europe populations and white triangles Slavic South Eastern Europe populations.

(2%) and Belgrade (0%) populations. In the Vojvodina Roma, E-M78 is completely represented by the E1b1b1a2-V13 derivative, the only branch of this sub-clade that reaches its highest levels outside of continental Africa.

I-M258 is the only haplogroup besides H-M69 that is shared across all three Roma populations. In this study, the Belgrade (20%) and Vojvodina (15%) samples display the highest proportions of I-M258 derived Y-chromosomes, whereas the lowest levels are present in the Roma residing in Kosovo (5%). Because of amplification failure of the I-M258 derived samples for the downstream mutations, Whit Athey's haplogroup predictor (Athey, 2006) (http://www.hprg.com/hapest5/index.html) was used to assign them to their respective I sub-clades based on their Y-STR profiles. Using this approach, 50% percent of the Belgrade and Kosovo I-M258 chromosomes were characterized as representatives of the I2a (defined by P37.2) subhaplogroup, while the remaining 50%, as well as all the M258 derived samples in the Vojvodina collection, were predicted as members of the I1 (defined by M253) sub-clade. The Bayesian probabilities of the predicted haplogroups yielded confidence estimates ranging from 97.7 to 100.0%.

Of the remaining haplogroups (G-M201, J-M304, and R-M207), J-M304 comprises 23% of the paternal gene pool of the Kosovo Roma, the second most abundant marker present in this group. All J-M304 chromosomes but one (J2b-M-12) were found to be J2a4b-M67 by Whit Athey's haplogroup predictor (Bayesian probabilities ranging from 94.9 to 100.0%). J-M304 Y-chromosomes were also observed at a reduced level (10%, and all predicted as J2a4b-M67 with Bayesian probabilities ranging from 70.1 to 77.6%) in the Roma from Belgrade but are completely absent in the Vojvodina samples. Although not as abundant, haplogroups R-M207 and G-M201 were also detected in our samples. The former was observed at equivalent frequencies in the Kosovo (5%) and Vojvodina (4%) collections, whereas the latter, which is prominent in the Caucasus region, is present only in the Kosovo community (5%).

Genetic structure

In the MDS plot (Fig. 2), eight populations define a Romani cluster in the left half of the graph, whereas the right portion is occupied by an amalgamation of Southeast European Slavic (SEES) and Non Slavic (SEEN) speaking collections. The Koya, a Dravidian group from Southern India, segregates within the well-defined Romani cluster while the remainder of the Romani populations are scattered around the periphery of the SEE, SEEN and South Central Asian (SCA) aggregates, reflecting different levels of admixture/relatedness. In the center of the upper section of the plot, the SCA assemblage occupies an intermediary position between the Romani and SEE groups of populations. Noteworthy is the segregation of the Lithuanian (Northern European) and the Baranja (Balkan Peninsula) Romani populations away from the Romani cluster and in the direction of the SCA collections. Given the variation in the distribution of haplogroup H-M69 among the three Roma collections under investigation, it is not surprising that Kosovo and Belgrade are part of the assemblage of eight Romani collections since both groups are characterized by high frequencies (>50%) of M69 derived Y-chromosomes. In contrast, the Vojvodina population, with only 7% Y M69 chromosomes, partitions as an outlier and maps at a considerable distance from the other Romani collections (with the exception of TUR and KAL) displaying greater affinities with the Southeast Europeans than with the South Central Asian populations.

Phylogenetic associations apparent in the MDS graph are also detected in the admixture analyses. The admixture proportions (Table 2) generated for the three hybrid populations reveal contrasting genetic profiles; while the Kosovo and Belgrade Roma receive their largest paternal component from SCA (98.7% and 74.9%, respectively), genetic contributions made to the Vojvodina gene pool are exclusively from the SEEN populations.

Significant correlation between genetics and geography but a low correlation between genetics and language

TABLE 2. Admixture estimates

	Hybrid populations		
Parental groups[a]	Kosovo Roma	Belgrade Roma	Vojdovina Roma
SCA	0.987 ± 0.295	0.749 ± 0.360	0.000 ± 0.070
SEES	0.000 ± 0.179	0.251 ± 0.256	0.000 ± 0.221
SEENS	0.013 ± 0.257	0.000 ± 0.244	1.000 ± 0.271

[a] Abbreviations for the parental groups are provided in Table 1.

has been reported for European populations indicating the primary importance of geography in the shaping of patterns of variation (Rosser et al., 2000). Therefore, to investigate the importance of linguistic and geographic partitioning to the genetic variation of the European Romani populations, AMOVA was performed utilizing the Y-SNP data (Table 3). The highest fraction of the variability, as expected, resides within the populations (81%). When grouped based on geography, little differentiation is detected among regions ($\Phi_{ct} = 0.08$) but greater variation is observed among populations residing within the same geographical area ($\Phi_{sc} = 0.11$). A similar pattern is evident when the populations are grouped according to language family ($\Phi_{ct} = 0.08$ and $\Phi_{sc} = 0.12$, respectively).

TABLE 3. Analysis of molecular variance (AMOVA) using Y-SNPs

	Total variation	Fixation indices		
Classification criterion	(%)	Φ_{ST}	Φ_{SC}	Φ_{CT}
Language (8 groups)[a]				
Among groups	8.28			0.08**
Among populations within groups	10.79		0.12*	
Within population	80.93	0.19*		
Geography (4 groups)[b]				
Among groups	8.09			0.08*
Among populations within groups	10.82		0.11*	
Within population	81.09	0.18*		

* *P*-value < 0.00001.** *P*-value < 0.001.[a] Linguistic partitioning (8 groups): 1-Indo-Iranian (PAK, GUJ, LAM, SIN, IND, KAL, BAR, MED, MAR, KOS, BEL, VOJ, LOM, MON, LIN, INT, POR, LIT, SPA); 2-Dravidian (PAR, NAI, KOY, CHE), 3-Slavic (CRO, HER, SER, MAC, BOS, SLO); 4-Albanian (KOA, ALB, AYR); 5-Italic (ROM); 6-Greek (GRE); 7-Turkish (FER, TUR) 8-Unclassified (AND).[b] Geographical partitioning (4 groups): 1-South Central Asia (PAK, GUJ, LAM, SIN, IND, PAR, NAI, AND, KOY, CHE); 2-Southeastern Europe (CRO, HER, SER, MAC, BOS, SLO, KOA, ALB, GRE, ROM, AYR, KAL, FER, TUR, BAR, MED, MAR, KOS, BEL, VOJ, LOM, MON, LIN, INT); 3-Western Europe (POR, SPA); 4-Northern Europe (LIT).

Y-STR haplotypes and time estimates

Haplotype diversity for the Kosovo, Belgrade, and Vojvodina Roma across the 15 Y-STR loci were 0.9630 ± 0.0149, 0.9815 ± 0.0147, and 0.9508 ± 0.0271, respectively. Analysis of the profiles obtained for the H-M69 derived samples revealed 18 different haplotypes of which the most frequently observed haplotype, 15-15,17-14-30(16)-22-10-11-12-14-9-11-19-15-18-20-12, is shared by 12 of the 86 (14%) individuals examined, seven from the Kosovo (18%) and five from the Belgrade (20%) collections (Supporting Information Table 1). Calculation of haplotype diversities within haplogroup H at a lower (7 of the 15 loci) resolution revealed the highest diversity values in Pakistan and India (0.95 ± 0.04 and 0.94 ± 0.01, respectively), whereas Belgrade and Kosovo exhibit much lower values, 0.55 ± 0.13 and 0.51 ± 0.08, respectively (Table 4). Also noteworthy is the complete lack of diversity within haplogroup H in the Romani from Spain and Portugal (a total of 26 individuals) at the western fringes of the Romani range.

For the majority of the populations, time estimates based on Zhivotovsky et al., (2004) and NETWORK using the evolutionary mutation rate are comparable. On the other hand, time estimates using the genealogical mutation rate (Goedbloed et al., 2009) seem to fit better with historical data of the Romani diaspora. When the Network software was used, the oldest age estimates for haplogroup H were attained for India (15.782 ± 3.465 kya) and Pakistan (8.799 ± 3.148 kya), whereas relatively younger ages were estimated for the Kosovo (1.882 ± 1.663 kya) and Belgrade (1.725 ± 0.909 kya) Roma. Interestingly, age estimates for H-M69 in these two Roma populations did not change substantially (2.195 ± 0.918 kya in Kosovo and 2.146 ± 0.782 kya in Belgrade) when the 15 Y-STR loci were utilized. Similarly, when the Zhivotovsky et al., (2004) method was used, India and Pakistan also exhibited the oldest ages for haplogroup H (17.705 ± 7.850 and 4.545 ± 1.033 kya, respectively). Also, as was detected within the Network analysis, the age estimates for the Kosovo (0.727 ± 0.727 kya) and Belgrade (0.631 ± 0.250 kya) Roma are considerably younger (Table 4) than in the Indian subcontinent. Examination of the 12 reference Romani populations revealed that only five of these groups exhibit sufficient haplotype diversity to allow calculation of time estimates with the Network program, which requires a minimum of three different haplotypes for time estimation.

Figure 3 illustrates star-shaped MJ networks constructed from the Y-STR data of H-M69 derived individuals from Belgrade, Vojvodina, and Kosovo based on 15 loci (3A) as well as the Eurasian data utilizing the 7 loci in common (3B). The most frequently detected lineage, the Roma modal haplotype (DYS19*15, DYS389I*14, DYS389II*16, DYS390*22, DYS391*10, DYS392*11, DYS393*12) (Gusmão et al., 2008), is shared by all Romani populations, most likely accounting for the high proportion (77%) of H chromosomes present in these groups. The remaining haplotypes differ from the modal profile, in most instances, by only one microsatellite molecular step.

DISCUSSION

In this study, the paternal gene pools of three Balkan Roma populations (Belgrade, Vojvodina and Kosovo) were examined and compared phylogenetically to targeted reference collections to ascertain the genetic ancestry of these groups. In our dataset, all the three Roma populations surveyed exhibit haplogroup H-M69 Y-chromosomes (known to be of Indian origin) but also display varying frequencies of five additional haplogroups that characterize paternal lineages of Southeastern Europe (i.e., E-M35, G-M201, I-M258, J-M304, and R-M207). The J-M304 chromosomes have been detected at high frequencies in the Middle East, Europe, and India, whereas haplogroup G is found predominantly in the Middle East and Europe, and is minimally observed in

TABLE 4. Time estimates, mean variance, and haplotype diversity for haplogroups H and E

	N	Time estimates using NETWORK (kya ± sd)[e]	Time estimates by the method of Zhivotovsky et al. (2004)[e]		Mean variance	Haplotype diversity ± sd	Reference
			Evolutionary mutation rate[c]	Genealogical mutation rate[d]			
HG H							
KOS[a]	22	2.195 ± 0.918	4.391 ± 3.001	1.696 ± 1.159	0.113	0.839 ± 0.056	This study
BEL[a]	18	2.146 ± 0.782	2.415 ± 0.986	0.933 ± 0.381	0.065	0.902 ± 0.049	This study
KOS[b]	22	1.882 ± 1.663	1.882 ± 1.882	0.727 ± 0.727	0.043	0.515 ± 0.081	This study
BEL[b]	18	1.725 ± 0.909	1.634 ± 0.648	0.631 ± 0.250	0.046	0.555 ± 0.130	This study
BAR	49	1.509 ± 0.628	1.690 ± 0.751	0.653 ± 0.290	0.047	0.301 ± 0.085	Klarić et al., 2009
MED	27	<3 different HT	0.575 ± 0.575	0.222 ± 0.222	0.015	0.262 ± 0.097	Klarić et al., 2009
MON	14	0.739 ± 0.522	0.739 ± 0.739	0.285 ± 0.285	0.022	0.274 ± 0.148	Gresham et al., 2001
LOM	26	3.3583 ± 3.197	3.583 ± 3.143	1.384 ± 1.214	0.041	0.566 ± 0.067	Gresham et al., 2001
FER	4	<3 different HT	1 HT	1 HT	1 HT	1 HT	Gresham et al., 2001
TUR	4	<3 different HT	1 HT	1 HT	1 HT	1 HT	Gresham et al., 2001
KAL	5	<3 different HT	1.035 ± 1.035	0.400 ± 0.400	0.029	0.400 ± 0.237	Gresham et al., 2001
LIN	12	<3 different HT	1 HT	1 HT	1 HT	1 HT	Gresham et al., 2001
INT	12	1.294 ± 0.964	1.293 ± 0.897	0.500 ± 0.346	0.034	0.439 ± 0.158	Gresham et al., 2001
LIT	10	4.658 ± 3.768	4.658 ± 3.598	1.800 ± 1.390	0.059	0.688 ± 0.103	Gresham et al., 2001
SPA	5	<3 different HT	1 HT	1 HT	1 HT	1 HT	Gresham et al., 2001
POR	21	<3 different HT	1 HT	1 HT	1 HT	1 HT	Gusmão et al., 2008
HG E							
VOJ[a]	18	1.671 ± 1.246	4.562 ± 1.672	1.762 ± 0.646	0.118[a] – 0.047[b]	0.817 ± 0.069	This study
BAR[a]	25	2.070 ± 1.270	7.536 ± 2.392	2.912 ± 0.924	0.178[a] – 0.104[b]	0.856 ± 0.042	Klarić et al., 2009
MED[a]	12	<3 different HT	1 HT	1 HT	1 HT	1 HT	Klarić et al., 2009

[a] 15 Y-STRs.[b] 7 Y-STRs.[c] Mutation rate equal to 6.9×10^{-4} per 25 years (Zhivotovsky et al., 2004).[d] Mutation rate equal to 2.5×10^{-3} per 35 years (Goedbloed et al., 2009).[e] Time estimates based on the Romani Modal Haplotype (DYS19*15, DYS389I*14, DYS389II*16, DYS390*22, DYS391*10, DYS392*11, DYS393*12) (Gusmão et al., 2008). Because of the small sample sizes, the estimates of the expansion time should be considered indicative.
HT: haplotype.

India. R chromosomes, in contrast, are common throughout Europe and the Indian subcontinent but their distribution in the Middle East is limited. Haplogroups I and E-V13 are particularly abundant in Europe and almost absent in India and the Middle East (Karafet et al., 2008).

Although haplogroup H derivatives account for greater than 50% of the Y-chromosomes present in the Kosovo and Belgrade samples, only 7% of the individuals from Vojvodina belong to this haplogroup. Instead, 93% of this collection's Y-chromosomes are from Southeastern Europe. The conclusions based on these values should be tempered in light of the low number of individuals analyzed.

Although the actual distribution of haplogroup H is almost entirely restricted to the Indian subcontinent including Pakistan and Sri Lanka (Wells et al., 2001; Karafet et al. 2005; Sahoo et al., 2006; Sengupta et al., 2006; Thanseem et al., 2006), several studies report its presence in Romani groups living in Europe (Gresham et al., 2001; Kalaydjieva et al., 2001; Peričić et al., 2005; Gusmão et al., 2008; Klarić et al., 2009) at frequencies ranging from 11% in the Turgovzi to 87% in the Monteni, both of which reside in Bulgaria (Gresham et al., 2001). The limited diversity within this lineage in Romani populations supports the possible Indian ancestry of this group (Gresham et al., 2001; Kalaydjieva et al., 2001 Gusmão et al., 2008; Klarić et al., 2009). Specifically, the H1-M52 sub-clade, which is present at elevated levels in the Koya (61%) and the Chenchu (36%) (Kivisild et al., 2003) populations of Southern India, typically predominates over all other Y-lineages in Romani groups. Thus, it is not surprising that this haplogroup and its downstream mutations account for the majority of Y-chromosomes in both the Kosovo (60%) and Belgrade (70%) collections. The relatively high frequency of H1-M52 derived samples in both groups is comparable to those observed in other Roma populations from the Balkans, including the Vlax from Bulgaria (73% of haplogroup H1-M52) (Kalaydjieva et al., 2001), the Roma from Macedonia (59.6% of haplogroup H1a-M82) (Peričić et al., 2005), and the Bayash from Croatia (50.3% of haplogroup H1a-M82) (Klarić et al., 2009). The lower level of H1a samples in the Vojvodina collection, when compared to the Belgrade and Kosovo Roma, may be indicative of different ancestral population(s), gene flow from local populations, genetic drift and/or the small number of individuals sampled from this region. In addition, it is possible that converted Romani may help to explain the genetic uniqueness of the Vojvodina collection. The low number of Vojvodina individuals carrying H chromosomes does not imply that they are not Romani.

It is notable that the Romani Modal Haplotype consisting of 7 Y-STR loci (DYS19*15, DYS389I*14, DYS389II*16, DYS390*22, DYS391*10, DYS392*11, DYS393*12) (Gusmão et al., 2008) is detected across all Romani populations, including the three Roma collections examined in the present study, highlighting the fact that they all share a common ancestor (Fig. 3B). In addition, 5 out of the 12 reference European Romani populations exhibit only this modal haplotype, i.e., they possess no haplotype diversity within haplogroup H (Table 4), which is indicative of close genealogical relationships among them. Interestingly, four Dravidian (Koya Dora, Vanniyar, Vellalar and Muria) and four Indo-European (Chamar, Rajput, West Bengal Brahmin and Nav

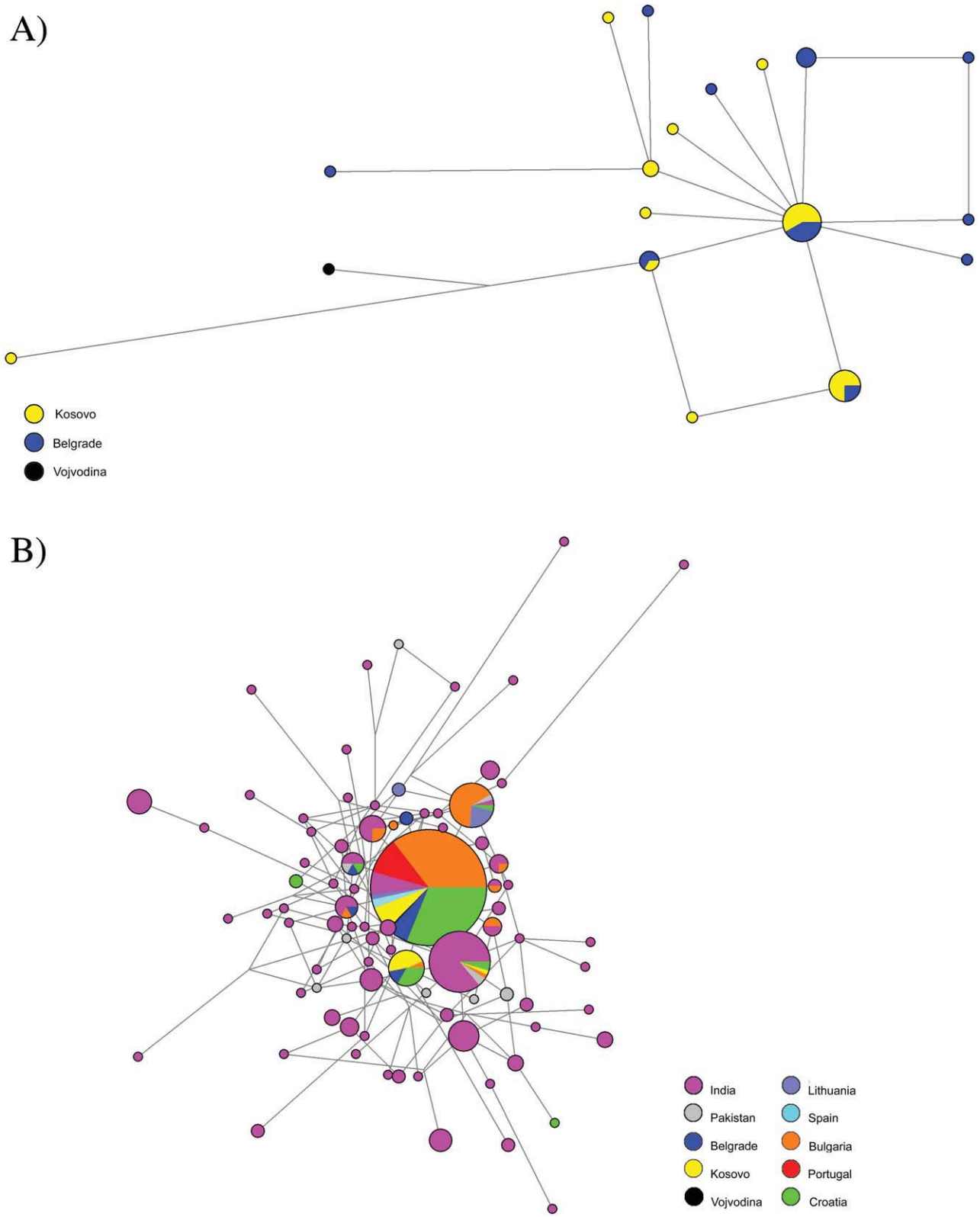

Fig. 3. Median-joining networks of Y-STR haplotypes within haplogroup H. **A**: MJ network of the three Roma populations reported in this study based on 15 STRs. **B**: MJ network of European Romani and Indian populations based on 7 STRs. The area of the pies is proportional to the frequency of that haplotype within the haplogroup, the smallest area is equivalent to one individual. Branch length is proportional to the number of mutations between haplotypes.

Buddha) populations from India (Sengupta et al., 2006) also display the Romani Modal Haplotype making it difficult to identify the Indian source population of the European proto-Romani groups. The reduced haplotype diversity combined with limited mean variance and recent time estimate of the founding Romani node (Table 4) suggest a founder effect and/or a bottleneck event in the proto-Romani group after their initial migration out of India, consistent with previous studies (Gresham et al., 2001; Gusmão et al., 2008).

In contrast to the Belgrade and Kosovo collections, the Vojvodina population exhibits a substantially higher proportion of E-M78 Y-chromosomes (70%) compared to those observed in other Romani groups; 30% in the Roma from Macedonia (Peričić et al., 2005) and 42% in the Bayash from Međimurje in northwestern Croatia (Klarić et al., 2009). Of the E-M78 lineages, E-V13 is the only branch of this sub-clade that reaches high frequencies outside of continental Africa, representing about 85% of the European M78 derived chromosomes (Cruciani et al., 2007). In Balkan populations, the highest proportion of haplogroup E1b1b1a2 (E-V13) is detected in Kosovar Albanians (45.6%) (Peričić et al., 2005) as well as in Albanians native to the former Yugoslavian Republic of Macedonia (34.4%) (Battaglia et al., 2009). An elevated level of this marker in the gene pool of the Vojvodina Roma is likely the result of recent gene flow from Southeastern European populations with subsequent isolation and/or genetic drift. This contention is supported in the present study by both the MDS and admixture analyses. In the MDS plot (Fig. 2), the Vojvodina collection is an outlier, plotting equidistant from the KOA (SEENS) and HER (SEES) populations from Southeast Europe and according to the results of the admixture tests (Table 2), this group receives its entire paternal component from SEENS. Moreover, the expansion time for the E-V13 marker in the Roma from Vojvodina (1.67 $\pm$ 1.2 kya) is more recent than those reported for non-Romani groups in Europe (4.0–4.7 kya) (Cruciani et al., 2007) or the Balkan Peninsula (5.9 $\pm$ 1.7 kya) (Battaglia et al., 2009). Yet, it is also possible that the Vojvodina collection does not share the same ancestors in the Indian subcontinent but are derived from local populations. In this scenario, the presence of haplogroup H may be the result of limited gene flow from authentic Roma population(s). All together, the available Y-chromosome data indicate that different Romani groups have experienced various levels of gene flow from local populations.

Haplogroup I, which comprises approximately 18% of the European Y-chromosomes (Rootsi et al., 2004), has been detected in the Balkan Peninsula (Barać et al., 2003; Rootsi et al., 2004; Pericić et al., 2005; Battaglia et al., 2009) at frequencies ranging from 9.8% in Greeks to 73.3% in Croatians from Bosnia (Battaglia et al., 2009). In this study, haplogroup I was detected in all three Roma communities (20%, 15% and 5% in Belgrade, Vojvodina, and Kosovo, respectively) signaling differential gene flow from SEE. This inference is further supported by the presence of the I2a-P37.2 sub-haplogroup (predominantly found in the Balkans) in the Roma from Belgrade and Kosovo (10% and 2.5%, respectively) (Battaglia et al., 2009; Mirabal et al., 2009; Mirabal et al., 2010). The Vojvodina I chromosomes, in contrast, belong to the I1-M253 sub-clade, which occurs at high frequencies in Northern Europe but has also been observed sporadically in the Balkan Peninsula (Battaglia et al., 2009). The relatively high frequency of M253 Y-chromosomes in the Vojvodina Roma (15%) is likely the result of genetic admixture with neighboring Balkan populations and subsequent isolation and/or founder effect.

Haplogroup J arrived in Europe from the Middle East and its sub-haplogroups marked complex migrations during and after the Neolithic period (Cinnioğlu et al., 2004). J2 is thought to be the most important Y-chromosome marker for the spread of farming into southeast Europe (Semino et al., 2000; Di Giacomo et al., 2004). The presence of sub-haplogroup J2a4b-M67 in the populations of Belgrade (10%) and Kosovo (21%), as assessed by Haplogroup Predictor algorithms, is reminiscent of the frequencies found in other Romani groups throughout Europe (Gresham et al., 2001; Gusmão et al., 2008; Klarić et al., 2009). As previously suggested, it is likely that the presence of this sub-haplogroup in Romanis is the result of infiltration from local populations. Although probably this incorporation took place early before their migration into different regions of Europe (Gusmão et al., 2008), the present as well as previous studies cannot rule out the possibility that the introduction occurred independently in different locales.

With the exception of the Vojvodina population, genetic ties between the Roma from Kosovo and Belgrade, and the other European Romanis are apparent in the MDS plot. These two Roma collections plot closer to the Balkan Roma (Bulgarian, Macedonian and Croatian) populations than to those from Western Europe (SPA and POR). Instead, the Spanish and Portuguese Roma as well as the Feredjelli collection from Bulgaria group together with the Indian populations in the SCA cluster, most likely the result of similar frequencies of haplogroup H-M69. It is not clear whether the partitioning of the Iberian Romani with the collections from the Indian subcontinent reflects random drift resulting from bottleneck effects during their east-west expansion to the western fringes of the Romani range or if it signals stronger phylogenetic relationships with sources populations. In addition, the statistically insignificant correlations observed between geography and genetics in the AMOVA indicate that contrary to other Europeans (Rosser et al., 2000) geography has not played a major role in shaping the genetic landscape of the Romani people (Gresham et al., 2001). These findings suggest that regardless of the distant geographic locations of Romani populations throughout Europe, their Y-chromosome lineages remain shared, most likely the result of a common origin and limited levels of genetic admixture with the exception of Vojvodina. The significant preservation of ancestral H1a patrilineages in the Romani is likely the result of considerable but variable levels of endogamy and isolation. The most recent European strata is dominated by the E1b1b1a2 lineages in Vojvodina, I in Belgrade, and J in Kosovo, reflecting differential gene flow, admixture with the local populations, and/or genetic drift.

Overall, our results indicate that the most pronounced difference among the three Roma collections examined in this study lies in the contrasting distributions of Y haplogroups H and E. The high proportions of haplogroup H-M69 lineages in the Kosovo and Belgrade populations, along with relatively reduced haplotype diversity and mean variance compared to the Indian subcontinent, lend support to the plausible Indian ancestry of these groups. In addition, the coalescent times based on the Romani modal haplotype for the Kosovo and Belgrade populations are younger than their Indian coun-

terpart. Conversely, the low frequency of H-M69 derived chromosomes, in combination with elevated levels of haplogroups E-V13 and I-M258, supports a higher degree of gene flow and/or admixture of the Vojvodina Roma with neighboring Balkan groups.

LITERATURE CITED

Abicht A, Stucka R, Karcagi V, Herczegfalvi A, Horvath R, Mortier W, Schara U, Ramaekers V, Jost W, Brunner J, Janssen G, Seidel U, Schlotter B, Muller-Felber W, Pongratz D, Rudel R, Lochmuller H. 1999. A common mutation (epsilon1267delG) in congenital myasthenic patients of Gypsy ethnic origin. Neurology 53:1564–1569.

Ackerley, FG. 1941. Bosnian Romani: prolegomena. J Gypsy Lore Soc Third Series 20:84–99.

Ali M, McKibbin M, Booth A, Parry DA, Jain P, Riazuddin SA, Hejtmancik JF, Khan SN, Firasat S, Shires M, Gilmour DF, Towns K, Murphy AL, Azmanov D, Tournev I, Cherninkova S, Jafri H, Raashid Y, Toomes C, Craig J, Mackey DA, Kalaydjieva L Riazuddin, S, Inglehearn CF. 2009. Null mutations in LTBP2 cause primary congenital glaucoma. Am J Hum Genet 84:664–671.

Álvarez A, del Castillo I, Villamar M, Aguirre LA, González-Neira A, López-Nevot A, Moreno-Pelayo MA, Moreno F. 2005. High prevalence of the W24X mutation in the gene encoding connexin-26 (GJB2) in Spanish Romani (gypsies) with autosomal recessive non-syndromic hearing loss. Am J Med Genet Part A 137A:255–258.

Applied Biosystems. 2006. AmpFlSTR Yfiler PCR amplification kit user's manual. Foster City, CA: Applied Biosystems.

Athey WT. 2006. Haplogroup predictions from Y-STR values using a Bayesian-Allele-frequency approach. J Genet Geneal 2:34–39.

Bandelt HJ, Forster P, Röhl A. 1999. Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol 16:37–48.

Barać L, Pericić M, Klarić IM, Rootsi S, Janićijević B, Kivisild T, Parik J, Rudan I, Villems R, Rudan P. 2003. Y chromosomal heritage of Croatian population and its island isolates. Eur J Hum Genet 11:535–542.

Battaglia V, Fornarino S, Al-Zahery N, Olivieri A, Pala M, Myres NM, King RJ, Rootsi S, Marjanovic D, Primorac D, Hadziselimovic R, Vidovic S, Drobnic K, Durmishi N, Torroni A, Santachiara-Benerecetti AS, Underhill PA, Semino O. 2009. Y-chromosomal evidence of the cultural diffusion of agriculture in southeast Europe. Eur J Hum Genet 17:820–830.

Boretzky N. 1986. Zur Sprache der Gurbet von Priština (Jugoslawien). Giessener Hefte für Tsiganologie 3:195–216.

Boretzky N. 1999. Die Verwandtschaftsbeziehungen zwischen den Südbalkanischen Romani-Dialekten. Mit einem Kartenanhang. Frankfurt am Main: Peter Lang.

Bosch E, Calafell F, González-Neira A, Flaiz C, Mateu E, Scheil HG, Huckenbeck W, Efremovska L, Mikerezi I, Xirotiris N, Grasa C, Schmidt H, Comas D. 2006. Paternal and maternal lineages in the Balkans show a homogeneous landscape over linguistic barriers, except for the isolated Aromuns. Ann Hum Genet 70:459–487.

Bouwer S, Angelicheva D, Chandler D, Seeman P, Tournev I, Kalaydjieva L. 2007. Carrier rates of the ancestral Indian W24X mutation in GJB2 in the general Gypsy population and individual subisolates. Genetic Test 11:455–458.

Brown KA, Janjua AH, Karbani G, Parry G, Noble A, Crockford G, Bishop DT, Newton VE, Markham AF, Mueller RF. 1996. Linkage studies of non-syndromic recessive deafness (NSRD) in a family originating from the Mirpur region of Pakistan maps DFNB1 centromeric to D13S175. Hum Mol Genet 5:169–173.

Chiaroni J, Underhill PA, Cavalli-Sforza LL. 2009. Y chromosome diversity, human expansion, drift, and cultural evolution. Proc Natl Acad Sci USA 106:20174–20179.

Cinnioğlu C, King R, Kivisild T, Kalfoğlu E, Atasoy S, Cavalleri GL, Lillie AS, Roseman CC, Lin AA, Prince K, Oefner PJ, Shen P, Semino O, Cavalli-Sforza LL, Underhill PA. 2004. Excavating Y-chromosome haplotype strata in Anatolia. Hum Genet 114:127–148.

Croxen R, Newland C, Betty M, Vincent A, Newsom-Davis J, Beeson D. 1999. Novel functional epsilon-subunit polypeptide generated by a single nucleotide deletion in acetylcholine receptor deficiency congenital myasthenic syndrome. Ann Neurol 46:639–647.

Cruciani F, Fratta R, Torroni A, Underhill PA, Scozzar R. 2006. Molecular dissection of the Y chromosome haplogroup E-M78 (E3b1a): a posterior evaluation of a microsatellite-network-based approach through six new biallelic markers. Hum Mutat 27:831–832.

Cruciani F, La Fratta R, Trombetta B, Santolamazza P, Sellitto D, Colomb EB, Dugoujon JM, Crivellaro F, Benincasa T, Pascone R, Moral P, Watson E, Melegh B, Barbujani G, Fuselli S, Vona G, Zagradisnik B, Assum G, Brdicka R, Kozlov AI, Efremov GD, Coppa A, Novelletto A, Scozzari R. 2007. Tracing past human male movements in northern/eastern Africa and western Eurasia: new clues from Y-chromosomal haplogroups E-M78 and J-M12. Mol Biol Evol 24:1300–1311.

Di Giacomo F, Luca F, Popa LO, Akar N, Anagnou N, Banyko J, Brdicka R, Barbujani G, Papola F, Ciavarella G, Cucci F, Di Stasi L, Gavrila L, Kerimova MG, Kovatchev D, Kozlov AI, Loutradis A, Mandarino V, Mammi' C, Michalodimitrakis EN, Paoli G, Pappa KI, Pedicini G, Terrenato L, Tofanelli S, Malaspina P, Novelletto A. 2004. Y chromosomal haplogroup J as a signature of the post-neolithic colonization of Europe, Hum Genet 115:357–371.

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

Firasat S, Khaliq S, Mohyuddin A, Papaioannou M, Tyler-Smith C, Underhill PA, Ayub Q. 2007. Y-chromosomal evidence for a limited Greek contribution to the Pathan population of Pakistan. Eur J Hum Genet 15:121–126.

Fraser A. 1992. The Gypsies. Oxford: Blackwell Publishers.

Gayden T, Regueiro M, Martinez L, Cadenas AM, Herrera RJ. 2008. Human Y-chromosome haplotyping by allele-specific polymerase chain reaction. Electrophoresis 29:2419–2423.

Goedbloed M, Vermeulen M, Fang RN, Lembring M, Wollstein A, Ballantyne K, Lao O, Brauer S, Krüger C, Roewer L, Lessig R, Ploski R, Dobosz T, Henke L, Henke J, Furtado MR, Kayser M. 2009. Comprehensive mutation analysis of 17 Y-chromosomal short tandem repeat polymorphisms included in the AmpFlSTR Yfiler PCR amplification kit. Int J Legal Med 123:471–482.

Gresham D, Morar B, Underhill PA, Passarino G, Lin AA, Wise C, Angelicheva D, Calafell F, Oefner PJ, Shen P, Tournev I, de Pablo R, Kucinskas V, Perez-Lezaun A, Marushiakova E, Popov V, Kalaydjieva L. 2001. Origins and divergence of the Roma (Gypsies). Am J Hum Genet 69:1314–1331.

Gusmão L, Butler JM, Carracedo A, Gill P, Kayser M, Mayr WR, Morling N, Prinz M, Roewer L, Tyler-Smith C, Schneider PM. 2006. DNA Commission of the International Society of Forensic Genetics (ISFG): an update of the recommendation on the use of Y-STRs in forensic analysis. Forensic Sci Int 157:187–197.

Gusmão A, Gusmão L, Gomes V, Alves C, Calafell F, Amorim A, Prata MJ. 2008. A perspective on the history of the Iberian Gypsies provided by phylogeographic analysis of Y-chromosome lineages. Ann Hum Genet 72:215–227.

Hammer MF, Horai S. 1995. Y chromosomal DNA variation and the peopling of Japan. Am J Hum Genet 56:951–962.

Hancock, I. 2002. We are the Romani people. Hartfordshire: University of Hartfordshire Press.

Kalaydjieva L, Calafell F, Jobling MA, Angelicheva D, de Knijff P, Rosser, ZH, Hurles ME, Underhill P, Tournev I, Marushiakova E, Popov V. 2001. Patterns of inter- and intragroup genetic diversity in the Vlax Roma as revealed by Y chromosome and mitochondrial DNA lineages. Eur J Hum Genet 9:97–104.

Kalaydjieva L, Morar B. 2003. Roma (Gypsies): genetic studies. In: Cooper N, editor. Nature encyclopedia of the human Genome, Vol. 5. London: Nature Publishing Group. p 160–165.

Kalaydjieva L, Morar B, Chaix R, Tang H. 2005. A newly discovered founder population: the Roma/Gypsies. BioEssays 27:1084–1094.

Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL, Hammer MF. 2008. New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. Genome Res 18:830–838.

Kayser M, Krawczak M, Excoffier L, Dieltjes P, Corach D, Pascali V, Gehrig C, Bernini LF, Jespersen J, Bakker E, Roewer L, de Knijff P. 2001. An extensive analysis of Y-chromosomal microsatellite haplotypes in globally dispersed human populations. Am J Hum Genet 68:990–1018.

Kelsell DP, Dunlop J, Stevens HP, Lench NJ, Liang JN, Parry G, Mueller RF, Leigh IM. 1997. Connexin 26 mutations in hereditary non-syndromic sensorineural deafness. Nature 387:80–83.

Kivisild T, Rootsi S, Metspalu M, Mastana S, Kaldma K, Parik J, Metspalu E, Adojaan M, Tolk HV, Stepanov V, Gölge M, Usanga E, Papiha SS, Cinnioğlu C, King R, Cavalli-Sforza L, Underhill PA, Villems R. 2003. The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. Am J Hum Genet 72:313–332.

Klarić IM, Salihović MP, Lauc LB, Zhivotovsky LA, Rootsi S, Janićijević B. 2009. Dissecting the molecular architecture and origin of Bayash Romani patrilineages: genetic influences from South-Asia and the Balkans. Am J Phys Anthropol 138:333–342.

Li CC. 1976. First course in population genetics. Pacific Grove: Boxwood Press.

Liégeois JP. 1994. Roma, Gypsies, Travellers. Strasbourg: Council of Europe Press.

Long JC, Williams RC, McAuley JE, Medis R, Partel R, Tregellas WM, South SF, Rea AE, McCormick SB, Iwaniec U. 1991. Genetic variation in Arizona Mexican Americans: estimation and interpretation of admixture proportions. Am J Phys Anthropol 84:141–157.

Maheshwari M, Vijaya R, Ghosh M, Shastri S, Kabra M, Menon PS. 2003. Screening of families with autosomal recessive nonsyndromic hearing impairment (ARNSHI) for mutations in GJB2 gene: Indian scenario. Am J Med Genet Part A 120A:180–184.

Malyarchuk BA, Grzybowski T, Deremko MV, Czarny J, Miscicka-Sliwka D. 2006 Mitochondrial DNA diversity in the Polish Roma. Ann Hum Genet 70:195–206.

Malyarchuk BA, Perkova MA., Derenko MV, Vanecek T, Lazur J, Gomolcak P. 2008. Mitochondrial DNA variability in Slovaks, with application to the Roma origin. Ann Hum Genet 72:228–240.

Marjanović D, Fornarino S, Montagna S, Primorac D, Hadziselimovic R, Vidovic S, Pojskic N, Battaglia V, Achilli A, Drobnic K, Andjelinovic S, Torroni A, Santachiara-Benerecetti AS, Semino O. 2005. The peopling of modern Bosnia-Herzegovina: Y-chromosome haplogroups in the three main ethnic groups. Ann Hum Genet 69:757–763.

Martinez L, Underhill PA, Zhivotovsky LA, Gayden T, Moschonas NK, Chow CE, Conti S, Mamolini E, Cavalli-Sforza LL, Herrera RJ. 2007. Paleolithic Y-haplogroup heritage predominates in a Cretan highland plateau. Eur J Hum Genet 15:485–493.

Marushiakova E, Popov V. 1997. Gypsies (Roma) in Bulgaria. Frankfurt am Main: Peter Lang.

Minárik G, Ferák V, Feráková E, Ficek A, Poláková H, Kádasi L. 2003. High frequency of GJB2 mutation W24X among Slovak Romany (Gypsy) patients with non-syndromic hearing loss (NSHL). Gen Physiol Biophys 22:549–556.

Mirabal S, Regueiro M, Cadenas AM, Cavalli-Sforza LL, Underhill PA, Verbenko DA, Limborska SA, Herrera RJ. 2009. Y-chromosome distribution within the geo-linguistic landscape of northwestern Russia. Eur J Hum Genet 17:1260–1273.

Mirabal S, Varljen T, Gayden T, Regueiro M, Vujovic S, Popovic D, Djuric M, Stojkovic O, Herrera RJ. 2010. Human Y-chromosome short tandem repeats: a tale of acculturation and migrations as mechanisms for the diffusion of agriculture in the Balkan Peninsula. Am J Phys Anthropol 142:380–390.

Morar B, Gresham D, Angelicheva D, Tournev I, Gooding R, Guergueltcheva V, Schmidt C, Abicht A, Lochmuller H, Tordai A, Kalmar L, Nagy M, Karcagi V, Jeanpierre M, Herczegfalvi A, Beeson D, Venkataraman V, Warwick Carter K, Reeve J, de Pablo R, Kucinskas V, Kalaydjieva L. 2004. Mutation history of the Roma/Gypsies. Am J Hum Genet 75:596–609.

Nei M. 1987. Molecular evolutionary genetics. New York: Columbia University Press.

Perićić M, Lauc LB, Klarić IM, Rootsi S, Janićijevic B, Rudan I, Terzić R, Colak I, Kvesić A, Popović D, Sijacki A, Behluli I, Dordevic D, Efremovska L, Bajec DD, Stefanović BD, Villems R, Rudan P. 2005. High-resolution phylogenetic analysis of Southeastern Europe traces major episodes of paternal gene flow among Slavic populations. Mol Biol Evol 22:1964–1975.

Rootsi S, Magri C, Kivisild T, Benuzzi G, Help H, Bermisheva M, Kutuev I, Barać L, Perićić M, Balanovsky O, Pshenichnov A, Dion D, Grobei M, Zhivotovsky LA, Battaglia V, Achilli A, Al-Zahery N, Parik J, King R, Cinnioğlu C, Khusnutdinova E, Rudan P, Balanovska E, Scheffrahn W, Simonescu M, Brehm A, Goncalves R, Rosa A, Moisan JP, Chaventre A, Ferak V, Füredi S, Oefner PJ, Shen P, Beckman L, Mikerezi I, Terzić R, Primorac D, Cambon-Thomsen A, Krumina A, Torroni A, Underhill PA, Santachiara-Benerecetti AS, Villems R, Semino O. 2004. Phylogeography of Y-chromosome haplogroup I reveals distinct domains of prehistoric gene flow in Europe. Am J Hum Genet 75:128–137.

Rosser ZH, Zerjal T, Hurles ME, Adojaan M, Alavantic D, Amorim A, Amos W, Armenteros M, Arroyo E, Barbujani G, Beckman G, Beckman L, Bertranpetit J, Bosch E, Bradley DG, Brede G, Cooper G, Côrte-Real HB, de Knijff P, Decorte R, Dubrova YE, Evgrafov O, Gilissen A, Glisic S, Gölge M, Hill EW, Jeziorowska A, Kalaydjieva L, Kayser M, Kivisild T, Kravchenko SA, Krumina A, Kucinskas V, Lavinha J, Livshits LA, Malaspina P, Maria S, McElreavey K, Meitinger TA, Mikelsaar AV, Mitchell RJ, Nafa K, Nicholson J, Nørby S, Pandya A, Parik J, Patsalis PC, Pereira L, Peterlin B, Pielberg G, Prata MJ, Previderé C, Roewer L, Rootsi S, Rubinsztein DC, Saillard J, Santos FR, Stefanescu G, Sykes BC, Tolun A, Villems R, Tyler-Smith C, Jobling MA. 2000. Y-chromosome diversity in Europe is clinal and influenced primarily by geography, rather than by language. Am J Hum Genet 67:1526–1543.

Sahoo S, Singh A, Himabindu G, Banerjee J, Sitalaximi T, Gaikwad S, Trivedi R, Endicott P, Kivisild T, Metspalu M, Villems R, Kashyap VK. 2006. A prehistory of Indian Y chromosomes: evaluating demic diffusion scenarios. Proc Natl Acad Sci USA 103:843–848.

Scott DA, Kraft ML, Carmi R, Ramesh A, Elbedour K, Yairi Y, Srisailapathy CR, Rosengren SS, Markham AF, Mueller RF, Lench NJ, Van Camp G, Smith RJ, Sheffield VC. 1998. Identification of mutations in the connexin 26 gene that cause autosomal recessive non-syndromic hearing loss. Hum Mutat 11:387–394.

Semino O, Magri C, Benuzzi G, Lin AA, Al-Zahery N, Battaglia V, Maccioni L, Triantaphyllidis C, Shen P, Oefner PJ, Zhivotovsky LA, King R, Torroni A, Cavalli-Sforza LL, Underhill PA, Santachiara-Benerecetti AS. 2004. Origin, diffusion, and differentiation of Y-chromosome haplogroups E and J: inferences on the neolithization of Europe and later migratory events in the Mediterranean area. Am J Hum Genet 74:1023–1034.

Semino O, Passarino G, Oefner PJ, Lin AA, Arbuzova S, Beckman LE, De Benedictis G, Francalacci P, Kouvatsi A, Limborska S, Marcikiae M, Mika A, Mika B, Primorac D, Santachiara-Benerecetti AS, Cavalli-Sforza LL, Underhill PA. 2000. The genetic legacy of Paleolithic *Homo sapiens sapiens* in extant Europeans: a Y chromosome perspective. Science 290:1155–1159.

Sengupta S, Zhivotovsky LA, King R, Mehdi SQ, Edmonds CA, Chow CE, Lin AA, Mitra M, Sil SK, Ramesh A, Usha Rani MV, Thakur CM, Cavalli-Sforza LL, Majumder PP, Underhill

PA. 2006. Polarity and temporality of high-resolution y-chromosome distributions in India identify both indigenous and exogenous expansions and reveal minor genetic influence of Central Asian pastoralists. Am J Hum Genet 78:202–221.

Thanseem I, Thangaraj K, Chaubey G, Singh VK, Bhaskar LV, Reddy BM, Reddy AG, Singh L. 2006. Genetic affinities among the lower castes and tribal groups of India: inference from Y chromosome and mitochondrial DNA. BMC Genet 7:42.

Underhill PA, Jin L, Lin AA, Mehdi SQ, Jenkins T, Vollrath D, Davis RW, Cavalli-Sforza LL, Oefner PJ. 1997. Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. Genome Res 7:996–1005.

Underhill PA, Passarino G, Lin AA, Shen P, Mirazón Lahr M, Foley RA, Oefner PJ, Cavalli-Sforza LL. 2001. The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. Ann Hum Genet 65:43–62

Wells RS, Yuldasheva N, Ruzibakiev R, Underhill PA, Evseeva I, Blue-Smith J, Jin L, Su B, Pitchappan R, Shanmugalakshmi S, Balakrishnan K, Read M, Pearson NM, Zerjal T, Webster MT, Zholoshvili I, Jamarjashvili E, Gambarov S, Nikbin B, Dostiev A, Aknazarov O, Zalloua P, Tsoy I, Kitaev M, Mirrakhimov M, Chariev A, Bodmer WF. 2001. The Eurasian heartland: a continental perspective on Y-chromosome diversity. Proc Natl Acad Sci USA 98:10244–10249.

Zhivotovsky LA, Underhill PA. 2005. On the evolutionary mutation rate at Y-chromosome STRs: comments on paper by Di Giacomo et al. (2004). Hum Genet 116:529–532.

Zhivotovsky LA, Underhill PA, Cinnioğlu C, Kayser M, Morar B, Kivisild T, Scozzari R, Cruciani F, Destro-Bisol G, Spedini G, Chambers GK, Herrera RJ, Yong KK, Gresham D, Tournev I, Feldman MW, Kalaydjieva L. 2004. The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. Am J Hum Genet 74:50–61.