



Genetic structure and admixture between Bayash Roma from northwestern Croatia and general Croatian population: evidence from Bayesian clustering analysis

Natalija Novokmet¹, Ana Galov², Damir Marjanović^{1,3}, Vedrana Škaro⁴, Petar Projić⁴, Gordan Lauc^{4,5}, Dragan Primorac^{6,7,8} and Pavao Rudan^{1,9}

¹Institute for Anthropological Research, Zagreb, Croatia

natalija.novokmet@inantro.hr

pavao.rudan@inantro.hr

²Division of Biology, Faculty of Science, University of Zagreb, Zagreb, Croatia

anagalov@zg.biol.pmf.hr

³International Burch University, Sarajevo, Bosnia and Herzegovina

damir.marjanovic@ibu.edu.ba

⁴Genos, DNA Laboratory, Zagreb, Croatia

glauc@public.srce.hr

vskaro@genos.hr

pprojić@genos.hr

⁵Faculty of Pharmacy and Biochemistry, University of Zagreb, Zagreb, Croatia

⁶Medical School, University of Split, Split, Croatia

draganprimorac2@gmail.com

⁷Eberly College of Science, Penn State University, University Park, PA, USA

⁸University of New Haven, New Haven, West Haven, Conn, USA

⁹Anthropological Center of the Croatian Academy of Sciences and Arts, Zagreb, Croatia

The first two authors have equally contributed to this work.

With 3 figures and 2 tables

Summary: The European Roma represent a transnational mosaic of minority population groups with different migration histories and contrasting experiences in their interactions with majority populations across the European continent. Although historical genetic contributions of European lineages to the Roma pool were investigated before, the extent of contemporary genetic admixture between Bayash Roma and non-Romani majority population remains elusive. The aim of this study was to assess the genetic structure of the Bayash Roma population from northwestern Croatia and the general Croatian population and to investigate the extent of admixture between them. A set of genetic data from two original studies (100 Bayash Roma from northwestern Croatia and 195 individuals from the general Croatian population) was analyzed by Bayesian clustering implemented in STRUCTURE software. By re-analyzing published data we intended to focus for the first time on genetic differentiation and structure and in doing so we clearly pointed to the importance of considering social phenomena in understanding genetic structuring. Our results demonstrated that two population clusters best explain the genetic structure, which is consistent with social exclusion of Roma

and the demographic history of Bayash Roma who have settled in NW Croatia only about 150 years ago and mostly applied rules of endogamy. The presence of admixture was revealed, while the percentage of non-Croatian individuals in general Croatian population was approximately twofold higher than the percentage of non-Romani individuals in Roma population corroborating the presence of ethnomimicry in Roma.

Key words: population substructure, bottleneck, Bayesian clustering method, ethnomimicry.

Introduction

The Roma are an example of a population that do not share a common homeland, speak different languages and dialects, and consist of individuals of various religions. They are composed of a mosaic of groups differing in cultural and social structures and customs, sometimes separated by strict rules of endogamy. The European Roma are an example of a transnational founder population with common Indian origins (Gresham et al. 2001, Morar et al. 2004, Peričić et al. 2005, Malyarchuk et al. 2006, Gusmao et al. 2008, Martinović Klarić et al. 2008, Peričić Salihović et al. 2011). It is assumed that contemporary Roma originated from a small number of ancestors whose exodus from their original homeland in western India produced a very pronounced primary bottleneck effect about 40 generations ago (Morar et al. 2004). Various social and economic pressures within Europe led to a gradual fragmentation of the population into numerous subgroups as well as to secondary bottleneck effects 6 to 24 generations ago (Morar et al. 2004). The composition of the present-day Roma populations of European countries is the complex product of the early migrations from the Balkans into western Europe, completed by the 15th century, and three subsequent migration waves: the first during the end of the 19th century after the abolition of Gypsy slavery in Romania (Hancock 1987, Fraser 1992, Liégeois 1994); the second out of Yugoslavia during the 1960s and 1970s; and the third during the end of the 20th century following the political and economic changes in eastern Europe (Reyniers 1995). It has been estimated that between 12 and 14 million Roma people from different Roma groups live in Europe. Many of them live in non-registered settlements or are not registered; they move often and it is difficult to enumerate them in the official census (Njegovan 2011). The largest numbers of Roma are concentrated in central and southeastern Europe (Liégeois 1994, Marushiakova & Popov 2001).

The first written record of Roma presence in Croatia dates back to the 14th century court registers in Dubrovnik, southern Croatia (Šlezak 2010). These were most likely sporadic settlers, craftsmen of Romani origins. The 19th century migrations to north-western Croatia involve Koritari (as one of many Bayash groups), as a consequence of migrations out of Romania, after abolition of slavery there after 1856 (Vojak 2004). Newly freed Roma groups, forced by their difficult economic and social position, began to emigrate from Romanian lands to the neighboring countries. The use of Romani language was prohibited during the period of slavery in Romania and some groups adopted Romanian language as their mother tongue. The branch of Roma people in Croatia that speak the archaic dialect of the Romanian language (Ijimba d'bjajš) are the Bayash. Besides Medimurje and Baranja in Croatia, the Bayash are also settled in parts of Hungary, Serbia, Bosnia and Herzegovina, Bulgaria

and Romania, and, to a lesser extent, parts of Macedonia, Greece, Ukraine, Slovakia and Slovenia. Historically, the Bayash pursued traditional trades including crafting of wooden objects (utensils, troughs, spindles), however, today traditional crafts are being replaced by consumer goods trading and the collection and sale of secondary raw materials, especially metals (copper) and plastics. One of the rare crafts that have been preserved only in a few Bayash families is basket-making.

Although some Bayash families live side by side with the majority population of Croatia, most of the population lives in large groups, segregated from the rest of population – usually in the outskirts of villages or small towns without facilities such as sewage, waterworks, garbage collection, roads (Škarić Jurić 2007). The tradition of keeping multi-child families is preserved in the Bayash population in which women marry young and give birth to many children. The traditional assumption of semi-nomadic and nomadic life-style is not evident in the Bayash population which showed a reduced tendency to migrate (Mišetić 2005). According to the last official population census (2011) there are 16.975 registered Roma in Croatia (0.40 % of the total Croatian population). However, by the more realistic estimate accepted by both the Croatian government and Roma NGOs and used by international organizations, there are between 30,000 and 40,000 Roma in Croatia, comprising 1 % of the total population. The reasons for the discrepancy between the census and estimates are complex and mostly arise from the decision by the Roma to ignore the census or declare themselves as member of some other nationality (ethnomimicry) (Škarić Jurić 2007). The largest Roma communities are located in Medimurje County in northwestern Croatia (5,107) where an estimated 30 % of the total Roma population in Croatia lives (UNDP 2005).

Whereas the utility of hypervariable DNA polymorphisms, short tandem repeat (STR) in forensic applications is undisputed, their usefulness in understanding patterns of modern human variations is being widely debated. STR studies of human differentiation providing an overview of population structure and demographic history of human populations (e.g. Deka et al. 1995, Jorde et al. 1997, Rosenberg et al. 2002, Rosenberg et al. 2005, Rosenberg et al. 2006, Perez-Miranda et al. 2005, Handley et al. 2007, Zhivotovsky et al. 2009, Listman et al. 2010), including admixture analyses (e.g. Halder et al. 2009, Kutanan et al. 2011). Recent episodes of admixture and non-random mating contribute significantly to creation of populations with variable proportions of ancestral genetic signals. Therefore, anthropological genetic research is often focused on studying the role of genetic ancestry and admixture in population structuring.

Admixture, the genetic consequence of mixing between genetically differentiated populations resulting from geographical, ecological or cultural diversion, is common in human evolution (Kutanan 2011). However, there are communities with a low extent of admixture with surrounding populations, an unusually high degree of isolation, endogamy and inbreeding. Within isolation regions, a substructure of markedly differentiated endogamous subpopulations is often maintained, as reported in the Eastern Adriatic islands in Dalmatia, Croatia (Rudan et al. 1987, Wadle et al. 1998, Rudan et al. 1999, Martinović Klarić 2005, Vitart et al. 2006, Vitart et al. 2008, Novokmet et al. 2011, Karns et al. 2012), Sardinian region of Ogliastra (Fraumene et al. 2003), in the Daghestan highlands (Bulayeva et al. 2003) and in mountainous areas in Bosnia (Marjanović et al. 2004). One of the most interesting examples of human isolates is the Roma people.

Population-genetic studies of the Roma from different European countries have mostly sought to compare the Roma to autochthonous Europeans and to identify genetic affinities with proposed parental populations and with other Roma populations. A summary of classical genetic data (Kalaydjieva et al. 2001) reconfirmed the Indian origins and stressed internal diversification of the Roma who appear to be more heterogeneous than autochthonous European populations. In most eastern and south-eastern European countries, including the Republic of Croatia, the Roma are one of the national minorities exposed to a high degree of socio-economic, political, spatial and cultural marginalization and stigmatization. The most common reasons cited for the difficult position of Roma are inadequate inclusion in the formal educational system, labor market and health care system. Despite this, a significant rate of admixture of the Roma with the host populations they contacted during their migratory history was suggested (Gusmao et al. 2010). Martinović Klarić et al. (2008) found that almost half of the paternal pool of Croatian Bayash is made of typical European lineages that originate from the more distant part of Roma European history (between the eleventh and fourteenth centuries AD) rather than from contemporary times. Furthermore, based on their field work data, the authors estimated that approximately 90 % of the Bayash people chose partners from the Bayash communities in Croatia, Serbia and, occasionally, Hungary. However, the extent of contemporary genetic admixture between Bayash Roma and non-Romani majority population remains elusive.

The aim of this study is to analyze microsatellite variation at 15 loci (D8S1179, D21S11, D7S820, CSF1PO, D3S1358, TH01, D13S317, D16S539, D2S1338, D19S433, VWA, TPOX, D18S51, D5S818 and FGA) in the population of Bayash Roma from northwestern Croatia (NWC), and in the general Croatian Population in order to: (1) assess the genetic structure using genetic differentiation coefficient (F_{ST}) and a Bayesian clustering method, (2) investigate the extent of admixture between the two studied populations, and (3) investigate whether the historical demographic bottlenecks are accompanied by a genetic signature in NWC Bayash Roma population.

Subjects and methods

Sample and DNA typing

This study included 100 unrelated adult individuals of Bayash Roma from northwestern Croatia (town of Čakovec, Medimurje County) (Fig. 1). DNA extraction from whole blood, PCR amplification of autosomal STRs and DNA typing were performed as described previously (Novokmet & Pavčec 2007). The study also included a representative sample of 195 unrelated individuals born in Croatia, from 14 counties and the city of Zagreb, representing general Croatian population. DNA extraction from buccal swabs, PCR amplification and DNA typing were carried out as described earlier (Projić et al. 2007).

Statistical evaluation

Population comparisons by loci and between subpopulations, based on the estimation of gene diversity (h), coefficients of genetic differentiation (F_{ST}) and locus by locus analysis of molecular variance (AMOVA) were computed according to the methods implemented in Arlequin version 3.5.1.2 software (Excoffier & Lisher 2010). Gene diversity was estimated as

$$H = \frac{n}{n-1} \left(1 - \sum_{i=1}^k p_i^2 \right)$$

where n is the number of gene copies in the sample, k is the number of haplotypes, and p_i is the sample frequency of the i -th haplotype (Nei 1987). Populations structuring was investigated by Bayesian clustering implemented in STRUCTURE 2.3.3 software (Pritchard et al. 2000), which probabilistically assigns individuals to populations based on their multi-locus genotypes. We applied the model allowing for correlated allele frequencies and the admixture model, which permits mixed ancestry within individuals. A total of 1,000,000 Markov Chain Monte Carlo iterations after a burn-in period of 100,000 iterations were run for each number of genetic clusters (K , ranging from 1 to 4). For each value of K we repeated the analysis ten times and estimated the posterior probability of clustering from the average logarithmic probability of data across repeats. The most probable number of clusters was selected by calculating the K value, which is based on the rate of change in the log probability of data between successive K values (Evanno et al. 2005) using STRUCTURE HARVESTER v.0.6 (Earl & von Holdt 2012). Further, the program STRUCTURE 2.3.3 (Pritchard et al. 2000) was used in individual-level admixture analyses where individual admixture proportions (q) were estimated. If individuals were admixed, they were assigned to two or more populations, allowing an estimation of their ancestry.

To test for genetic evidence of recent bottleneck events, we calculated the heterozygosity excess for each locus and population. Calculations were performed by the program Bottleneck, v. 1.2.02. (Piry et al. 1999). When a population has been recently reduced in size, there is a deficit of rare alleles relative to the number expected in a population at equilibrium (because rare alleles are generally first lost through genetic drift). Since rare alleles contribute relatively little to the expected heterozygosity, there will be an excess of observed heterozygosity when compared to a population at equilibrium with an equivalent number of alleles



Fig. 1. Geographic location of the Republic of Croatia and the position of the town of Čakovec in NW Croatia.

(Luikart et al. 1998). Consequently, in a recently reduced population, the gene diversity will be higher than that expected at equilibrium. Gene diversity was estimated under two models of molecular evolution: the two-phase model (TPM) and the stepwise mutation model (SMM). The TPM has been shown to be the most appropriate for microsatellite DNA data (Di Rienzo et al. 1994). We used TPM with 95 % single-step mutations and 5 % multiple-step mutations, and a variance among multiple steps of 12, as recommended by Piry et al. (1999). Ten thousand iterations were used for each mutation model. The probability of significant heterozygosity excess was determined using Wilcoxon signed rank test.

Results and discussion

The average gene diversity across 15 markers was 0.744 (\pm 0.376) for NWC Bayash Roma which was lower than that observed in general Croatian population 0.777 (\pm 0.390). This diversity reduction of 4.25 % was already reported by Gusmao et al. (2010) and the fact that the decrease was higher than that reported for any other Roma-host population pair was attributed to possible stronger effects of drift on Croatian Roma than other European Roma as a consequence of more fragmentation events in their demographic history. The allele frequency distribution of the 15 AmpF/STR Identifiler Loci in NWC Bayash Roma as well as the general Croatian population were previously published in the articles Novokmet & Pavčec (2007) and the article Projic et al. (2007), respectively.

Genetic distances between pairs of those two subpopulations, quantified using F_{ST} measures, revealed statistically significant amount of genetic differentiation (3.4 %) (Table 1). Previous comparisons of NWC Roma with neighboring central and southeastern European non-Roma and available Roma populations revealed lower coefficients of genetic differentiation among Roma population pairs (ranging from 0.7 % to 3 %) and mostly greater coefficients of genetic differentiation among Roma and non-Roma population pairs (ranging from 3.1 % to 9.9 %) (Novokmet & Pavčec 2007) as compared to that obtained in our study (3.4 %). Likewise, locus-by-locus AMOVA pair-wise comparisons revealed significant differences between NWC Bayash and general Croatian populations at all loci using F_{ST} statistics (Table 1).

Bayesian clustering analysis in STRUCTURE identified $K = 2$ as the most likely number of clusters. The maximum value of mean log-likelihood was observed at $K = 2$ and estimates of K revealed the largest increase in the likelihood of the number of clusters also at $K = 2$ (Fig. 2). Most of the individuals of the NWC Roma population were assigned to a single genetic cluster and most of the individuals of the general Croatian population were assigned to another, while no further structuring was detected among either of the clusters (Fig. 3). Our failure to detect signals of further population substructuring among general Croatian population is in line with previous findings. Namely, preliminary Structure analysis of the Croatian population revealed no population structuring, and estimated that the most likely number of subpopulations identified was $K = 1$ (Martinović Klarić et al. 2005). The authors could not cluster individuals from the mainland and five isolated Croatian islands according to their geographic origin. Further, Novokmet (2011) did not detect signals of population structuring in a sample of Croatian population that included mainland and a larger number of isolated island populations. This was expected since the degree of genetic differentiation between the Croatian island populations was smaller (F_{ST} ranged from 0.02 % to 2.8 %; Novokmet et al. 2011) than that detected in this investigation

Table 1. Locus-by-locus AMOVA (p -value) and coefficient of genetic differentiation (F_{ST}) between NWC Bayash Roma and general Croatian population.

| Locus | F_{ST} |
|---------|--------------------------------|
| D8S1179 | 0.025 (0.000) |
| D21S11 | 0.020 (0.001) |
| D7S820 | 0.146 (0.000) |
| CSF1PO | 0.033 (0.000) |
| D3S1358 | 0.016 (0.004) |
| TH01 | 0.011 (0.019) |
| D13S317 | 0.044 (0.000) |
| D16S539 | 0.024 (0.000) |
| D2S1338 | 0.019 (0.000) |
| D19S433 | 0.041 (0.000) |
| vWA | 0.017 (0.000) |
| TPOX | 0.016 (0.002) |
| D18S51 | 0.044 (0.000) |
| D5S818 | 0.030 (0.000) |
| FGA | 0.016 (0.001) |
| Overall | 0.034 ($p = \mathbf{0.000}$) |

Significant values ($p = 0.05$) are marked in bold.

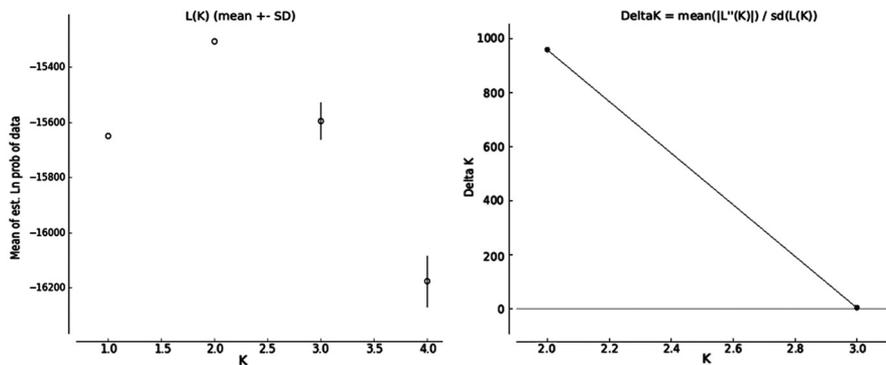


Fig. 2. Mean log likelihood values $\text{LnP}(D)$ (left) and the rate of change in the log probability of data between successive values of K (ΔK) (right) as a function of the number of genetic clusters (K) averaged over 10 independent STRUCTURE runs for each K .

between Bayash Roma and general Croatian population. In addition these results were based on only nine microsatellite loci.

Our results obtained with the clustering algorithm implemented in STRUCTURE illustrated that as low as 15 highly polymorphic loci was sufficient to reveal consistent structure, when differentiation was relatively high ($F_{ST} = 3.4\%$). This is also in

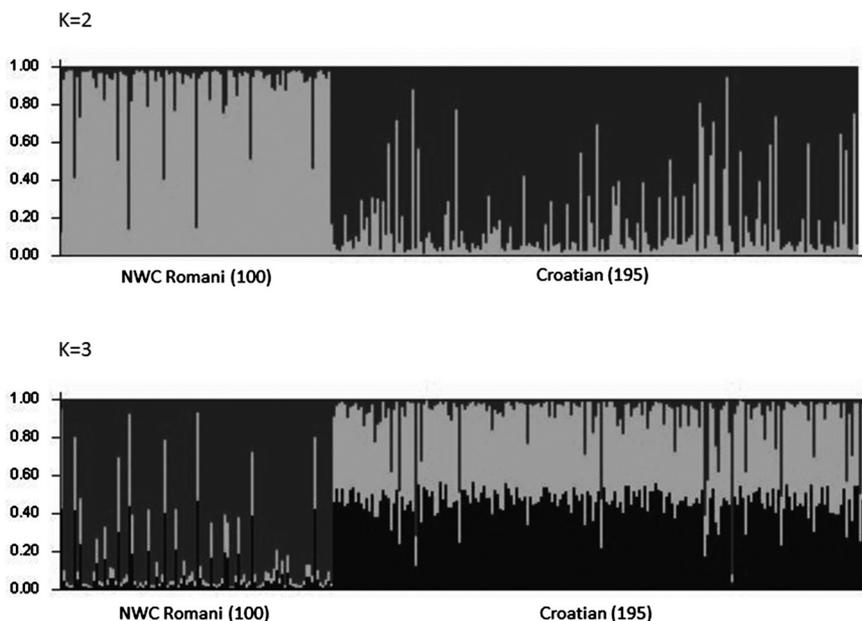


Fig. 3. Probabilistic assignment of NWC Roma and Croatian population to the genetic clusters inferred by Bayesian analysis performed in STRUCTURE, with $K = 2$ and $K = 3$. The most likely number of genetic clusters in the data set was identified as two. Each individual is represented by a vertical column partitioned into K segments that represent the individual's estimated membership fractions in K clusters. Numbers in parentheses indicate sample sizes of self-reported NWC Romani and Croatian populations.

line with the findings of Latch et al. (2006), that at levels of genetic differentiation of $F_{ST} \geq 3\%$, STRUCTURE correctly estimates the number of genetic clusters, K . Further, this is consistent with the demographic history of NWC Bayash Roma who have settled in NW Croatia only about 150 years ago, and whose interactions with majority populations have often taken the form of ethnic stigmatization or even complete social exclusion, in addition to strict rules of endogamy applied in Bayash (Martinović-Klarić et al. 2008).

Further, we estimated admixture proportions (q) for each individual, using the program STRUCTURE. Individuals were assumed as the admixed ones if the admixture proportion to their inferred population was $0.2 < q < 0.8$ which was set as an arbitrary conservative cut-off value. The individuals with $q < 0.2$ were assigned to a cluster opposite to their self-reported population. We further used even more conservative cut-off values of 0.75 and 0.7, as a threshold value for assuming the individuals as either admixed or clustering to the opposite population.

Eighty-seven individuals with self-reported Romani ancestry showed little or no evidence of admixture with the general Croatian population at the 0.8 level (estimated admixture proportions < 0.2) (Table 2). The proportion of the general Croatian population ancestry in ten Romani individuals (10%) ranged from 0.2 to 0.6 (Fig. 3), whereas three individuals with self-reported Romani ancestry (3%) clustered with general Croatian population ($q > 0.8$). Among samples from general Croatian popu-

Table 2. Percentages (%) of individuals self-reported as either Romani or Croatians classified according to their membership proportions inferred by Structure, for 0.8; 0.75 and 0.7 cut-off values (i.e., if an individual self-reported as Roma and the inferred membership proportion was $q >$ cut-off value, he/she was added to the inferred Roma cluster. If an individual's membership proportion was $(1 - \text{cut-off value}) < q <$ (cut-off value), he/she was classified as admixed. If an individual was self-reported as Roma and inferred membership proportion was $q <$ $(1 - \text{cut-off value})$, he/she was added to the inferred Croatian cluster.

| 0.8 level | | inferred by Structure | |
|------------------|------|-----------------------|----------|
| self-reported as | Roma | admixed | Croatian |
| Roma | 87 | 10 | 3 |
| Croatian | 2 | 22 | 76 |
| 0.75 level | | inferred by Structure | |
| self-reported as | Roma | admixed | Croatian |
| Roma | 91 | 6 | 3 |
| Croatian | 2.6 | 17.4 | 80 |
| 0.7 level | | inferred by Structure | |
| self-reported as | Roma | admixed | Croatian |
| Roma | 92 | 5 | 3 |
| Croatian | 4.1 | 12.8 | 83.1 |

lation, there were 149 individuals (76%) assigned to inferred Croatian cluster with $q > 0.8$, while 43 individuals (22%) showed evidence of admixture (estimated admixture proportions ranged from 0.2 to 0.76). Three individuals (2%) sampled as belonging to general Croatian population clustered with inferred Roma ($q > 0.8$). As expected, with lower cut-off values the number of individuals with little or no evidence of admixture increases, while the number of admixed individuals decreases. When comparing the proportion of individuals with admixed ancestry or clustering with the opposite population between inferred Roma and Croatian populations (Table 2), it is evident that the percentage of non-Croatian individuals (as inferred by Structure) in the general Croatian population is approximately twofold higher than the percentage of non-Roma individuals (as inferred by Structure) in the Roma population (24% versus 13% at the 0.8 level). This ratio is evident also at both lower cut-off values (20% versus 9% at the 0.75 level; 16.9% versus 8% at the 0.7 level). Loosely speaking, these results suggest that among individuals that represent the general Croatian population there are twice as many non-Croatian individuals (as inferred by Structure) than there are non-Roma individuals (as inferred by Structure) among individuals that are declared as Roma. This result is congruent with the notion that a great number of the Roma people still declare themselves as members of some other nationalities, most often in accordance with the language of the area in which they live, i.e. according to the dominant ethnic group (Njegovan 2011), in this case Croatian. The Roma may opt not to self-identify for various reasons, such as fear of discrimination (Barany 2001). Moreover, a high percentage of the Bayash population of Medimurje (85%) choose Croatian citizenship, most likely to obtain social welfare support allowance (Škarić-Jurić et al. 2007). Our results are in agreement with

ethnomimicry, a widespread tendency among the Roma to camouflage their ethnic background with a more perceived respectable identity that offers them better prospects of survival, although it seems that they are never fully assimilated into the group with which they identify (Duijzings 2000). Likewise, our results suggest that mixed Roma and non-Roma couples might prefer to leave Roma communities and integrate into the Croatian society. This is not surprising since the Roma have almost always lived worse off than their surrounding non-Roma populations (Friedman 2007).

Our finding of admixed and non-Roma individuals among individuals self-identified as Roma are in line with the finding that Croatia is one of the countries in which a few individuals self-identified as Roma present very long genomic segments of non-Roma origin (Mendizabal et al. 2012), and it further corroborates the shift in the social rules of the acceptance of Roma and non-Roma couples within Roma groups. However, the relatively low levels of admixture and low numbers of non-Roma individuals found in the Roma population (Table 2) suggests endogamous traditions among Bayash Roma in Croatia, as opposed to general Croatian population.

Our re-analysis indicated specific cultural and social issues imprinted in the contemporary genetic structure of one particular Romani group in Croatia – the Bayash. In the case of Bayash genetic structure it is important to consider specific problems with integration with the host population and dilemmas in declaring Roman identity for the fear of discrimination.

Bottleneck testing

For the Bottleneck analysis, Wilcoxon signed-rank tests were significant for NWC Roma population under both mutational models: TPM with 95 % single-step mutations ($p = 0.00754$) and SMM ($p = 0.02063$), providing evidence that the historical demographic bottlenecks are indeed accompanied by a genetic signature in NWC Bayash Roma population. This finding confirms the hypothesis of multiple exposures of Roma populations to the founder and bottleneck effects – the last being the settlement of a groups of Roma in the northwestern Croatia in the 19th century (Vojak 2004). It is assumed that contemporary Roma originated from a small number of ancestors whose exodus from their original homeland in western India produced a very pronounced primary bottleneck effect about 40 generations ago (Morar et al. 2004). Various social and economic pressures within Europe led to a gradual fragmentation of the population into numerous subgroups as well as to secondary bottleneck effects 6 to 24 generations ago (Morar et al. 2004). Contrary to that, Wilcoxon signed-rank test did not provide evidence for a genetic bottleneck in the general Croatian population under either mutational model: $p = 0.42346$ under TPM with 95 % single-step mutations and $p = 0.97604$ under SMM. This result was expected and in line with the historic data, which did not suggest any demographic bottleneck in the recent history of a Croatian population, in spite of the fact that some small Croatian populations experienced bottlenecks (Vitart et al. 2006).

In conclusion, our results obtained by Bayesian clustering showed that as few as 15 highly polymorphic microsatellite markers is sufficient to reveal consistent structure between Roma and a neighboring majority population, if differentiation is relatively high. Our results further reveal the extent of admixture between Bayash Roma and the general Croatian population, with higher percentages of admixed and non-

Croatian individuals found in the general Croatian population and lower percentages of admixed and non-Roma individuals found in the Bayash Roma population, which is in line with the presence of ethnomimicry in Roma. Further studies on other Roma and their respective majority populations would help in gaining insights into extent of a recent or ongoing admixture between Roma and the neighboring majority non-Roma populations and in revealing whether the level and direction of admixture varies geographically and temporally.

Acknowledgements

We express our gratitude to the deceased Zlatko Pavčec, M.D., M.Sc. from Čakovec County Hospital, Department of Radiology and Ultrasound, Čakovec, Croatia for collaboration and providing DNA samples of Roma participants. We would like to thank Irena Martinović Klarić, Ph.D., for helpful discussions and valuable comments and suggestions. We are grateful to all Croatian inhabitants and to all members of Roma community who participated in this project.

Grant sponsorship / Declaration of interest: This study was supported by the Ministry of Science, Education and Sports of the Republic of Croatia, scientific project: “Population structure of Croatia – Anthropogenetic approach”, number: 196-1962766-2751, grant to P.R. and scientific project “Development of Scientific System for DNA Analysis of Archeological Bone Samples” number 309-1300855-2738, grant to D.M. The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

References

- Barany, Z. (2001): *The East European Gypsies: Regime Change, Marginality and Ethnopolitics*. – Cambridge University Press, Cambridge, UK.
- Bulayeva, K., Jorde, L.B., Ostler, C., Watkins, S., Bulayev, O. & Harpending, H. (2003): Genetics and population history of Caucasus populations. – *Hum. Biol.* **75**: 837–853.
- Deka, R., Shriver, M.D., You, L.M., Robert, E.F. & Chakraborty, R. (1995): Intra- and inter-population diversity at short tandem repeat loci in diverse populations of the world. – *Electrophoresis* **16**: 1659–1664.
- Di Rienzo, A., Peterson, A.C., Garza, J.C., Valdes, A.M., Slatkin, M. & Freimer, N.B. (1994): Mutation processes of simple-sequence repeat loci in human populations. – *Proc. Natl. Acad. Sci. United States of America* **91**: 3166–3170.
- Duijzings, G. (2000): *Religion and the Politics of Identity in Kosovo*. – C. Hurst & Co., London
- Earl, D.A. & von Holdt, B.M. (2011): STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. – *Conserv. Gen. Resour.* **4**: 359–361.
- Evanno, G., Regnaut, S. & Goudet, J. (2005): Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. – *Mol. Ecol.* **14**: 2611–2620.
- Excoffier, L. & Lisher, H. (2010): Arlequin ver 3.5.1.2: An Integrated Software Package for Population Genetics Data Analysis. – Population Genetic Lab (CMPG), University of Berne.
- Fraser, A. (1992): *The Gypsies*. – Blackwell, Oxford.
- Fraumene, C., Petretto, E., Angius, A. & Pirastu, M. (2003): Striking differentiation of sub-populations within a genetically homogeneous isolate (Ogliastra) in Sardinia as revealed by mtDNA analysis. – *Hum. Genet.* **114**: 1–10.
- Friedman, E. (2007): A dual challenge for the year of equal opportunities for all : Roms in the Western Balkans. – *JEMIE Journal on ethnopolitics and minority issues in Europe* Vol. 6 (2007): 1–23. <http://nbn-resolving.de/urn:nbn:de:0168-ssoar-6189>
- Gresham, D., Morar, B., Underhill, P.A., Passarino, G., Lin, A.A., Wise, C., Angelicheva, D., Calafell, F., Oefner, P.J., Shen, P.D., 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. Gen.* **69** (6): 1314–1331.
- Gusmão, A., Gusmão, L., Gomes, V., Alves, C., Calafell, F., Amorim, A. & Prata, M.J. (2008): A Perspective on the history of the Iberian Gypsies provided by phylogeographic analysis of Y-chromosome lineages. – *Ann. Hum. Genet.* **72**: 215–227.
- Gusmão, A., Valente, C., Gomes, V., Alves, C., Amorim, A., Prata, M.J. & Gusmão, L.A. (2010): Genetic Historical Sketch of European Gypsies. The Perspective From Autosomal Markers. – *Am. J. Phys. Anthropol.* **141**: 507–514.
- Halder, I., Yang, B.Z., Kranzler, H.R., Stein, M.B., Shriver, M.D. & Gelernter, J. (2009): Measurement of admixture proportions and description of admixture structure in different U.S. populations. – *Hum. Mutat.* **30**: 1299–1309.
- Hancock, I. (1987): *The Pariah Syndrome*. – Karoma Publishers Inc., Ann Arbor, MI.
- Handley, L.J.L. (2007): Going the distance: human population genetics in a clinal world. – *TRENDS in Genetics* **23** (9): 432–439.
- Jorde, L.B., Rogers, A.R., Bamshad, M., Watkins, W.S., Krakowiak, P., Sung, S., Kere, J. & Harpending, H.C. (1997): Microsatellite diversity and the demographic history of modern humans. – *Proc. Natl. Acad. Sci. USA.* **94**: 3100–3103.
- Kalaydjieva, L., Gresham, D. & Calafell, F. (2001): Genetic studies of the Roma (Gypsies): a review. – *BMC Med. Genet.* **2**:5. doi: 10.1186/1471-2350-2-5.
- Kalaydjieva, L. & Morar, B. (2003): Roma (Gypsies). Genetic studies. – *Nature Encyclopedia of the Human Genome*.
- Karns, R., Zhang, G., Sun, G., Rao Indugula, S., Cheng, H., Havas Augustin, D., Novokmet, N., Rudan, D., Durakovic, Z., Missoni, S., Chakraborty, R., Rudan, P. & Deka, R. (2012): Genome-Wide Association of Serum Uric Acid Concentration: Replication of Sequence Variants in an Island Population of the Adriatic Coast of Croatia. – *Ann. Hum. Genet.* **76** (1): 1–7.
- Kutanan, W., Kampuansai, J., Colonna, V., Nakbunlung, S., Lertvicha, P., Seielstad, M., Bertorelle, G. & Kangwanpong, D. (2011): Genetic affinity and admixture of northern Thai people along their migration route in northern Thailand: evidence from autosomal STR loci. – *J. Hum. Genet.* **56** (2): 130–137.
- Latch, E.K., Dharmarajan, G., Glaubitz, J.C. & Rhodes, O.E. (2006): Relative performance of Bayesian clustering software for inferring population substructure and individual assignment at low levels of population differentiation. – *Conservation Genetics.* **7**: 295–302.
- Liegeois, J.P. (1994): *Roma, Gypsies, travellers*. – Council of Europe Press, Strasbourg.
- Listman, J.B., Hasin, D., Kranzler, H.R., Malison, R.T., Mutirangura, A., Sughondhabirom, A., Aharonovich, E., Spivak, B. & Gelernter, J. (2010): Identification of population substructure among Jews using STR markers and dependence on reference populations included. – *BMC Genetics* **11**:48.
- Luikart, G. & Cornuet, J.M. (1998): Empirical evaluation of a test for identifying recently bottlenecked populations from allele frequency data. – *Conservation Biology.* **12** (1): 228–237.
- Malyarchuk, B.A., Grzybowski, T., Derenko, M.V., Czarny, J. & Miscicka Sliwka, D. (2006): Mitochondrial DNA diversity in the Polish Roma. – *Ann. Hum. Genet.* **70**: 195–206.
- Marjanović, D., Kapur, L., Drobnić, K., Budowle, B., Pojskić, N. & Hadžiselimović, R. (2004): Comparative study of genetic variation at 15 STR loci in three isolated populations of the Bosnian mountain area. – *Hum. Biol.* **76**: 15–31.
- Martinović Klarić, I., Peričić, M., Barać Lauc, L., Janičijević, B., Kubat, M., Pavičić, D., Rudan, I., Wang, N., Jin, L., Chakraborty, R., Deka, R. & Rudan, P. (2005): Genetic Variation at nine short tandem repeat loci in the Croatian populations. – *Hum. Biol.* **77**: 471–486.
- Martinović Klarić, I., Peričić Salihović, M., Barać Lauc, L., Zhivotovsky, L.A., Roots, S. & Janičijević, B. (2008): 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.

- Marushiakova, E. & Popov, V. (2001): Historical and Ethnographic Background. Gypsies, Roma, Sinti. – In: Guy, W. (ed.): *Between Past and Future: the Roma of Central and Eastern Europe*. – University of Hertfordshire Press, Hatfield, pp. 33–53. [Updated]
- Mendizabal, I., Lao, O., Marigorta, U.M., Wollstein, A., Gusmão, L., Ferak, V., Ioana, M., Jordanova, A., Kaneva, R., Kouvatsi, A., Kučinskas, V., Makukh, H., Metspalu, A., Netea, M.G., de Pablo, R., Pamjav, H., Radojkovic, D., Rolleston, S.J.H., Sertic, J., Macek, M., Comas, D. & Kayser, M. (2012): Reconstructing the Population History of European Romani from Genome-wide Data – *Current Biology*. <http://dx.doi.org/10.1016/j.cub.2012.10.039>.
- Mišetić, A. (2005): The Socio-cultural Features of the Romani Population. – In: Štambuk, M. (ed.): *Kakožive Hrvatski Romi / How Do Croatian Roma Live*. — Institute of Social Sciences “Ivo Pilar”, Zagreb, pp. 343–363.
- Morar, B., Gresham, D., Angelicheva, D., Tournev, I., Gooding, R., Guerguelcheva, V., Schmidt, C., Abicht, A., Lochmüller, H., Tordai, A., Kalmar, L., Nagy, M., Karcagi, V., Jeanpierre, M., Herczegfalvi, A., Beeson, D., Venkataraman, V., Carter, K.W., Reeve, J., de Pablo, R., Kucinskas, V. & Kalaydjieva, L. (2004): Mutation history of the Roma/Gypsies. – *Am. J. Hum. Gen.* **75** (4): 596–609.
- Nei, M. (1987): *Molecular Evolutionary Genetics*. – Columbia University Press, New York, NY, USA.
- Novokmet, N. (2011): Meta-Analysis of the influence of Substructuring of statistical forensic parameters on genetic STR Markers in the Area of South-Eastern Europe. – Doctoral Thesis, Faculty of Science, University of Zagreb, Croatia (in Croatian).
- Novokmet, N. & Pavčec, Z. (2007): Genetic Polymorphisms of 15 AmpF/STR identifier loci in Romani population from Northwestern Croatia. – *Forensic Sci. Int.* **168**: 43–6.
- Novokmet, N., Marjanović, D., Škaro, V., Projić, P., Lauc, G., Grahovac, B., Ostojčić, S., Kapović, M. & Rudan, P. (2011): Genetic polymorphisms of 15 STR loci in the population of the island of Cres (Croatia). – *Ann. Hum. Biol.* **38** (1): 12–21.
- Njegovan, B.R. (2011): Social integration of Roma people – the importance and remit of Roma media: a case study. – *Trames*. **15**: 102–119.
- Pérez-Miranda, A.M., Alfonso-Sánchez, M.A., Kalantar, A., García-Obregón, S., de Pancorbo, M.M., Peña, J.A. & Herrera, R.J. (2005): Microsatellite data support subpopulation structuring among Basques. – *J. Hum. Genet.* **50**: 403–414.
- Peričić Salihović, M., Barešić, A., Martinović Klarić, I., Cukrov, S., Barać Lauc, L. & Janičević, B. (2011): The role of the Vlax Roma in shaping the European Romani maternal genetic history. – *Am. J. Phys. Anthropol.* **146**: 262–270.
- Piry, S., Luikart, G. & Cornuet, J.M. (1999): BOTTLENECK: A Computer Program for Detecting Recent Reductions in the Effective Population Size Using Allele Frequency Data. – *J. Heredity* **90**: 502–503.
- Pritchard, J.K., Stephens, M. & Donnelly, P. (2000): Inference of population structure using multilocus genotype data. – *Genetics* **155**: 945–959.
- Projić, P., Škaro, V., Šamija, I., Pojskić, N., Durmić-Pašić, A., Kovačević, L., Bakal, N., Primorac, D. & Marijanović, D. (2007): Allele Frequencies for 15 Short Tandem Repeat Loci in Representative Sample of Croatian Population. – *Croat. Med. J.* **48**: 473–77.
- Reyniers, A. (1995): Gypsy Populations and their Movements within Central and Eastern Europe and Towards some OECD Countries. *International Migration and Labour Market – Policies Occasional Papers*, No. 1. OECD: Paris.
- Rosenberg, N.A. (2006): Standardized subsets of the HGDP-CEPH Human Genome Diversity Cell Line Panel, accounting for atypical and duplicated samples and pairs of close relatives. – *Ann. Hum. Genet.* **70**: 841–847.
- Rosenberg, N.A., Pritchard, J.K., Weber, J.L., Cann, H.M., Kidd, K.K., Zhivotovsky, L.A. & Feldman, M.W. (2002): Genetic structure of human populations. – *Science* **298**: 2381–2385.

- Rosenberg, N.A., Mahajan, S., Ramachandran, S., Zhao, C., Pritchard, J. & Feldman, M.W. (2005): Clines, clusters, and the effect of study design on the inference of human population structure. – *PLoS Genet.* **1** (6): 660–671.
- Rudan, P., Šimić, D., Smolej-Narančić, N., Bennet, L.A., Janičijević, B., Jovanović, V., Lethbridge, M.F., Šimić, J., Roberts, D.F., Sujoldžić, A. & Szivovicsza, L. (1987): Isolation by distance in Middle Dalmatia – Yugoslavia. – *Am. J. Phys. Anthropol.* **74**: 417–426.
- Rudan, I., Campbell, H. & Rudan, P. (1999): Genetic epidemiological studies of eastern Adriatic island isolates, Croatia: Objectivities and strategies. – *Coll. Antropol.* **23** (2): 531–546.
- Škarić Jurić, T., Martinović Klarić, I., Smolej Narančić, N., Drmić, S., Peričić Salihović, M., Barać Lauc, L., Miličić, J., Barabalić, M., Zajc, M. & Janičijević, B. (2007): Trapped between Tradition and Transition Anthropological and Epidemiological Cross-sectional Study of Bayash Roma in Croatia. – *Croat. Med. J.* **48**: 708–719.
- Šlezak, H. (2010): Demogeografska i sociokulturna obilježja romske populacije u Meimurju. – Master Thesis, Faculty of Science, University of Zagreb, Croatia (in Croatian).
- United Nations Development Programme UNDP (2005): UNDP releases largest ever survey of Roma. – Available at: <http://vulnerability.undp.sk>.
- Vitart, V., Biloglav, Z., Hayward, C., Janičijević, B., Smolej-Narančić, N., Barać, L., Peričić, M., Martinović Klarić, I., Škarić-Jurić, T., Barbalić, M., Polašek, O., Kolčić, I., Carothers, A., Rudan, P., Hastie, N., Wright, A., Campbell, A. & Rudan, I. (2006): 3000 years of solitude: extreme differentiation in the island isolates of Dalmatia, Croatia. – *Eur. J. Hum. Genet.* **14**: 478–487.
- Vitart, V., Rudan, I., Hayward, C., Gray, N.K., Floyd, J., Palmer, C.N., Knott, S.A., Kolčić, I., Polašek, O., Graessler, J., Wilson, J.F., Marinaki, A., Riches, P.L., Shu, X., Janičijević, B., Smolej-Narančić, N., Gorgoni, B., Morgan, J., Campbell, S., Biloglav, Z., Barać-Lauc, L., Peričić, M., Klarić, I.M., Zgaga, L., Skarić-Jurić, T., Wild, S.H., Richardson, W.A., Hohenstein, P., Kimber, C.H., Tenesa, A., Donnelly, L.A., Fairbanks, L.D., Aringer, M., McKeigue, P.M., Ralston, S.H., Morris, A.D., Rudan, P., Hastie, N.D., Campbell, H. & Wright, A.F. (2008): SLC2A9 is a newly identified urate transporter influencing serum urate concentration, urate excretion and gout. – *Nat. Genet.* **40** (4): 437–42.
- Vojak, D. (2004): Romsko stanovništvo u popisima stanovništva u Hrvatskoj i Slavoniji u razdoblju 1850–1910. – *Časopis za suvremenu povijest* **36**: 701–728.
- Waddle, D.M., Sokal, R.R. & Rudan, P. (1998): Factors affecting population variation in Eastern Adriatic isolates (Croatia). – *Hum. Biol.* **70**: 845–864.
- Zhivotovsky, L.A., Malyarchuk, B.A., Derenko, M.V., Wozniak, M. & Grzybowski, T. (2009): Developing STR databases on structured populations: The native South Siberian population versus the Russian population. – *Forensic Sci. Int.: Genetics* **3**: e111–e116.

Submitted: 23 May 2014;
accepted: 17 November 2014.

Addresses for correspondence: N. Novokmet, Institute for Anthropological Research, Gajeva 32, 10000 Zagreb, Croatia
A. Galov, Division of Biology, Faculty of Science, University of Zagreb, Rooseveltov trg 6, 10000 Zagreb, Croatia
natalija.novokmet@inantro.hr
anagalov@zg.biol.pmf.hr