Estimation of population differentiation using pedigree and molecular data in Black Slavonian pig

Kristina Gvozdanović1, Dubravko Škorput2, Ivona Djurkin Kušec∗, Krešimir Salajpal2, Goran Kušec1

1 Josip Juraj Strossmayer University of Osijek, Faculty of Agrobiotechnical Sciences Osijek, Department of Animal Production and Biotechnology, Osijek, Croatia
2 University of Zagreb, Faculty of Agriculture, Department of Animal Science, Zagreb, Croatia

The aim of the study was to investigate the genetic differentiation of the Black Slavonian pig population. Two parallel analyses were performed using genealogical records and molecular data. Pedigree information of 6,099 pigs of the Black Slavonian breed was used to evaluate genetic variability and population structure. Additionally, 70 pigs were genotyped using 23 microsatellite markers. Genealogical data showed shrinkage in genetic diversity parameters with effective population size of 23.58 and inbreeding of 3.26%. Expected and observed heterozygosity were 0.685 and 0.625, respectively, and the average number of alleles per locus was 7.826. Bayesian clustering algorithm method and obtained dendrograms based on pedigree information and molecular data revealed the existence of four genetic clusters within the Black Slavonian pig. Wright’s $F_{IS}$, $F_{ST}$ and $F_{IT}$ from pedigree records were 0.017, 0.006, and 0.024, respectively, and did not prove significant population differentiation based on the geographical location of herds, despite the natural mating system. Obtained results indicate that despite the increased number of animals in the population, genetic diversity of Black Slavonian pig is low and conservation programme should focus on strategies aimed at avoiding further loss of genetic variability. Simultaneous use of genealogical and molecular data can be useful in conservation management of Black Slavonian pig breed.

Keywords: autochthonous pig breed, microsatellite, genealogical data, genetic structuring

1 Introduction

There are two main objectives of a conservation programme of a certain population: 1) to reduce the inbreeding levels and 2) to maintain the high level of genetic variability (Toomey et al., 2017). Genetic diversity is an important prerequisite for a successful implementation of conservation programmes. Expected heterozygosity, defined as the probability that two randomly chosen alleles from the population are different (Nei, 1973), is usually used to measure diversity within the population. In local and indigenous pig populations under conservation, the loss of alleles and levels of heterozygosity often occur due to inbreeding or small effective population size. Therefore, the minimization of those occurrences preserves the fluctuation of allelic frequencies within the population (Diniz-Filho et al., 2012). The adoption of appropriate conservation management is essential to preserve local populations since they are often considered to be part of the historical and cultural tradition (D’Alessandro et al., 2019). Population subdivision arises from the joint effects of multiple factors, the most influential being genetic drift, gene flow and selection (Lacy, 1987, Ma et al., 2015).

The assessment of genetic diversity parameters is traditionally based on genealogical information. However, the reliability of such estimations is highly dependent on data quality and integrity. The estimation of genetic diversity parameters from genealogical information assumes that animals in base generations are unrelated. This assumption is not always possible to realize and this leads to the
biased estimation of average inbreeding, inbreeding rate and effective population size. In cases in which information obtained from pedigree is scarce, the combination of pedigree and molecular data is usually beneficial (Wang, 2015).

The implementation of the systems based on DNA analysis is becoming more and more popular in the analysis of the genetic structure of local pig breeds. Different DNA markers have been used to assess genetic diversity in local pig breeds, including mitochondrial DNA analysis (Zhang et al., 2016, Gvozdanović et al., 2019), microsatellites (Cortés et al., 2016, Kramarenko et al., 2018), and single nucleotide polymorphism (SNP; Yang et al., 2017, Muñoz et al., 2019).

Black Slavonian pig is one of the three indigenous pig breeds in the Republic of Croatia. The number of breeding sows increased over last decade, with 2,500 registered breeding sows in 2019 (Croatian Agency for Food and Agriculture, 2020). The genomic effective population size of this breed was estimated to be 33.11 (Muñoz et al., 2019).

Several genetic diversity analyses of Black Slavonian pig breed using different marker systems have been performed so far (Muñoz et al., 2019, Gvozdanović et al., 2020). However, to the best of our knowledge, there are no reports on the genetic structuring of Black Slavonian pig population based on both genealogical and genomic data. Therefore, the aim of the study was to investigate if population differentiation of Black Slavonian pig using different sources of information exists, namely pedigree records and microsatellite markers.

2 Material and methods

2.1 Pedigree data

Pedigree data included 6,099 data records from Black Slavonian pigs farmed in the period from 1994 to 2019. The basic pedigree structure was assessed using CFC 1.0 (Sargolzaei et al., 2006). Coefficients of inbreeding (F) were computed using the algorithm of Meuwissen and Luo (1992) in the CFC software. The inbreeding rate (ΔF) was computed for each generation as:

\[ \Delta F = \frac{F_t - F_{t-1}}{1 - F_{t-1}} \]

where \( F_t \) and \( F_{t-1} \) are the average inbreeding coefficients for the current and the previous generation, respectively.

The effective population size (Ne) was estimated as: \( Ne = \frac{1}{2} \Delta F \).

To calculate the effective number of founders (founder equivalent, \( f_e \); Lacy, 1989), the following formula implemented in ENDOG 4.8. software (Gutiérrez and Goyache, 2005) was used:

\[ f_e = \frac{1}{\sum q_k^2} \]

where \( q_k \) is the probability of gene origin of the k ancestor, and \( f \) is total number of founders.

Effective number of ancestors (\( f_a \)) was calculated according to Boichard et al. (1997):

\[ f_a = \frac{1}{\sum q_j^2} \]

where \( q_j \) is the marginal contribution of an ancestor j, which is the genetic contribution made by an ancestor that is not explained by other ancestors chosen before, and \( a \) is the total number of ancestors.

The average relatedness coefficient (Goyache et al., 2003) of each individual is defined as the probability that an allele randomly chosen from the whole population in the pedigree belongs to a given animal. Genetic distances from pedigree records were defined as 1 minus relatedness coefficient and were used to construct a phylogenetic tree. The population structure was assessed using the F-statistics (Wright, 1931) according to Caballero and Toro (2000). Evaluations were made based on \( F_{ST} \) which estimates heterozygosity loss in subpopulations compared to the total population,
FIS which estimates heterozygosity loss within subpopulations and FIT which estimates heterozygosity loss of the entire population.

The F-statistics were obtained as $F_{IS} = \frac{f - \bar{f}}{1 - \bar{f}}$, $F_{ST} = \frac{f - \bar{f}}{1 - \bar{f}}$, and $F_{IT} = \frac{f - \bar{f}}{1 - \bar{f}}$, where $\bar{f}$ and $\bar{F}$ are the mean coancestry and the inbreeding coefficient for the entire metapopulation, respectively, and $\bar{F}$ is the average coancestry for the subpopulation, so that $(1 - F_{IT}) = (1 - F_{IS})(1 - F_{ST})$. Five subpopulations were predefined according to the geographical location of herds. The formulae for calculation of F-statistics were implemented in ENDOG 4.8. software (Gutiérrez and Goyache, 2005).

The pedigree completeness was expressed as number of fully traced generations defined as the number of generations separating the offspring of the furthest generation, where the 2g ancestors of the individual are known and where g is the number of generations. In addition, the number of equivalent generations defined as the sum over all known ancestors of the terms computed as the sum of $\frac{1}{2^g}$ where n is the number of generations separating the individual to each known ancestor, was calculated.

### 2.2 Molecular data

#### 2.1.1 Animal sampling and DNA analysis

Blood samples were collected from 70 unrelated Black Slavonian pigs and total genomic DNA was extracted using the Gene Jet Genomic DNA Purification Kit (Thermofisher Scientific, Massachusetts, Waltham, MA USA) according to the manufacturer’s protocol. A set of 23 microsatellite markers recommended by FAO (FAO, 2011) were chosen according to their fragment size and annealing temperature and grouped into three multiplex reactions. Multiplex PCR set up and amplification conditions were performed as previously described by Margeta et al. (2016) and Gvozdanović et al. (2020). Obtained PCR products were analysed using GeneScan350 ROX internal standard size marker on the ABI3730XL capillary gene analyser.

#### 2.1.2 Statistical analysis

Expected heterozygosity ($H_{exp}$), observed heterozygosity ($H_{obs}$) and the number of alleles per locus was computed using Genetix 4.05.2 software (Belkhir et al., 2004). Bayesian clustering was performed using the STRUCTURE software version 2.3.4 (Pritchard et al., 2000). The analysis was performed for 10 independent runs for K=1 to K=6, with a burn-in period of 100,000 iterations followed by 100,000 Markov Chain Monte Carlo iterations (MCMC). The number of assumed clusters (K) was estimated according to the Evanno method (Evanno et al., 2005) using the Structure Harvester algorithm (Earl and vonHoldt, 2012). The STRUCTURE results were graphically visualized using POPHELPER (Francis, 2017). The unweighted pair-group method with arithmetic average (UPGMA) based on the matrix of Nei's genetic distances was used to construct phylogenetic tree by adegenet package (Jombart, 2008, Jombart et al., 2010) in the R environment (R Development Core Team, 2018).

### 3 Results and discussion

The main genealogical parameters and population structure based on pedigree data are shown in Table 1. Pedigree completeness parameters, such as the number of fully traced generations and the number of equivalent generations indicate a low pedigree depth. This is, however, a common situation for the populations of domestic animals under conservation programs (Barros et al., 2017). The problem with the identification of common ancestors in shallow pedigrees arises from the lack of information in the pedigree, causing bias in the estimation of inbreeding coefficients, inbreeding rate and effective population size. Thus, conservation decisions such as mating plans can be affected by biased estimates and can cause the increase in inbreeding which is not accounted for by pedigree data. In such situations, the combination of molecular and pedigree data is helpful (Wang, 2015). However, the importance of improving the recording system of genealogical data remains the key task in the monitoring of the population structure.
Table 1 Population parameters obtained from genealogical data

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of animals</td>
<td>6,099</td>
</tr>
<tr>
<td>Number of animals with unknown parents</td>
<td>1,489</td>
</tr>
<tr>
<td>Inbreeding coefficient (F), %</td>
<td>3.26%</td>
</tr>
<tr>
<td>Inbreeding rate (ΔF), %</td>
<td>2.12%</td>
</tr>
<tr>
<td>Effective population size (Ne)</td>
<td>23.58</td>
</tr>
<tr>
<td>Average relatedness (AR), %</td>
<td>1.81</td>
</tr>
<tr>
<td>Effective number of founders (fs)</td>
<td>41</td>
</tr>
<tr>
<td>Effective number of ancestors (fa)</td>
<td>41</td>
</tr>
<tr>
<td>Number of ancestors explaining 50% of variability</td>
<td>30</td>
</tr>
<tr>
<td>Number of fully traced generations</td>
<td>1.20</td>
</tr>
<tr>
<td>Number of equivalent generations</td>
<td>2.05</td>
</tr>
</tbody>
</table>

The FAO recommends ΔF to be maximum 1% and Ne be maintained above 50 animals (FAO, 2000). The calculated F was 3.26%, in accordance with results of Lukić et al. (2015). This value is in line with values observed in some European local pig breeds, such as Retinto line (F=2.50%) and Retinto line (5.80%) of Iberian pig (Casellas et al., 2019), and Blonde (F=3.86%), Swallow-belly (F=3.29%) and Red Mangalitza pig (F=5.02%; Posta et al., 2016). The Ne estimated from genealogical data was 23.58, thus placing the Black Slavonian pig population among breeds with endangered genetic diversity. The Ne from pedigree data was higher than that estimated for Nero di Parma breed (7.68; Mariani et al., 2020), Mora Romagnola breed (10.87; Crovetti et al., 2013) and Gamito line of the Iberian pig (16.0; Silió et al., 2016). On the other hand, higher Ne was observed in Cinta Sinese breed (40.32; Crovetti et al., 2013) and Torbiscal line of the Iberian pig (57.7; Silió et al., 2016). The number of ancestors accounting for 50% of the variability was low, suggesting that some individuals were used more intensively in the population. This could cause a genetic bottleneck which contributes to genetic variability loss and is an important risk factor for the population (Goyache et al., 2003). Similar pattern can be observed with the fs: only small number of founders contributed to the genetic variability of the population. Thus, unequal contribution of founders caused the loss of genetic variability.

Table 2 Genetic differentiation among subpopulations of Black Slavonian pig from pedigree information

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>FIS</td>
<td>0.017</td>
</tr>
<tr>
<td>FST</td>
<td>0.006</td>
</tr>
<tr>
<td>FRT</td>
<td>0.024</td>
</tr>
</tbody>
</table>

FIS - heterozygosity loss within subpopulations
FST - heterozygosity loss in subpopulations compared to the total population
FRT - heterozygosity loss of the entire population

Pedigree information can be used to infer population structure through the Nei’s minimum distance (Nei, 1987) and F-statistics (Wright, 1978) based on the average pairwise coancestry coefficient between individuals of two subpopulations of a given metapopulation. The results of Wright's F-statistics are given in Table 2. Low FRT and FIS values indicate that allele fixation by homozygosis is not occurring. The value of FST confirms the results of FRT and FIS, which indicate that more than 99% of the genetic variability of the population corresponds to existing differences between individuals within the subpopulations. According to Wright's (1978) qualitative guidelines, an FST from 0.15 to 0.25 indicates large differentiation, from 0.05 to 0.15 indicates moderate differentiation and <0.05 indicates little differentiation among subpopulations. In the analysed population, the estimated FST was 0.006, meaning that Black Slavonian pig population does not show strong structuring based on predefined subpopulations, despite the natural mating system and the existence of closed herds. This is probably
due to the lack of nucleus herds and the exchange of genetic material in order to prevent the increase of the inbreeding rate in the population.

Figure 1 depicts the population structure and probable number of clusters assessed by STRUCTURE software. The highest peak of ΔK was identified for K=4, revealing the existence of four genetic clusters. The STRUCTURE analysis classified animals into herds from geographically different areas. Surprisingly, four animals were clustered together into the minor cluster group (Cluster 3, Figure 1). The coat colour of Black Slavonian pig is affected by black allele $E^{D_1}$, which is inherited dominantly and makes phenotypical distinction of pure animals from the crossbreds impossible (Gvozdanović et al., 2020). Although the four animals in Cluster 3 have black coat colour, genetically they are crossbreds with breeds farmed in the same geographical area. This may be a consequence of extensive rearing conditions and uncontrolled mating occurrences.

The differentiation of animals into four genetic clusters from STRUCTURE analysis was confirmed by the UPGMA dendrograms generated using microsatellite markers and genealogical data (Figure 2). The dendrograms obtained from the pedigree information of the whole registered population is characterized by four main clusters and their subdivision into additional clusters, corresponding to the results obtained from molecular data. Similarly to our results, Margeta et al. (2013) genotyped the Black Slavonian pigs using $MC1R$ gene and reported that majority of Black Slavonian pigs were actually crossbreds with modern pig breeds or wild boars, while Druml et al. (2014) reported genetic flow between Black Slavonian pigs and other local pig breeds, such as Turopolje pig. Additionally, using a high-density SNP microarray Muñoz et al. (2019) reported higher genetic heterogeneity within the breed. The same authors reported introgression of alleles 3, 4, and 6 of the $MC1R$ gene in Black Slavonian pigs indicating a contamination of the breed with commercial breeds such as Duroc, Pietrain and Large White.
Figure 2 UPGMA clustering tree obtained according to Nei's genetic distance (Nei et al., 1983) using a) microsatellite markers and b) complete pedigree of Black Slavonian pig population

Table 3 Average number of alleles per locus ($N_L$), expected heterozygosity ($H_{exp}$) and observed heterozygosity ($H_{obs}$) for the Black Slavonian pig breed from molecular data

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_{exp}$</td>
<td>0.685</td>
</tr>
<tr>
<td>$H_{obs}$</td>
<td>0.625</td>
</tr>
<tr>
<td>$N_L$</td>
<td>7.826</td>
</tr>
</tbody>
</table>

The $H_{exp}$ was 0.685 and $H_{obs}$ 0.625, indicating the existence of several genetic lineages and low rates of selection pressure as argued by Lemus-Flores et al. (2001). The average number of alleles per locus was 7.826 (Table 3). Obtained results are higher than those reported by Margeta et al. (2018), with values of 0.61 for $H_{exp}$ and 0.55 for $H_{obs}$. Compared to other local breeds, heterozygosity within the population of the Black Slavonian pig is high. Muñoz et al. (2019) reported that minimum and maximum $H_{exp}$ values of local pig breeds ranged from 0.187 (Turopolje pig) to 0.382 (Sarda), and $H_{obs}$ ranged from 0.195 (Turopolje pig) to 0.363 (Krškopolje pig). Similar to results of the present study, Cortés et al. (2016) reported $H_{exp}$ and $H_{obs}$ for Celta and Bisaro breeds that ranged from 0.60 to 0.63 and 0.50 to 0.55, while Scali et al. (2012) reported $H_{obs}$ of 0.45 for Cinta Senese pig breed.

4 Conclusions
In this study we analysed population differentiation of Black Slavonian pigs with different sources of information, namely microsatellite markers and pedigree records. Microsatellites revealed high heterozygosity and sub-structuring within the population. Genealogical data showed shrinkage in genetic diversity parameters for Black Slavonian pigs. Subdivision within the population follows the pattern obtained by the microsatellite information. However, genealogical data did not show significant structuring between predefined populations. Despite the increase of the number of animals in the
population, genetic diversity of Black Slavonian pig is low and conservation programme should aim at avoiding further loss of genetic variability. Overall, results confirmed that both microsatellites and pedigree records are useful for assessing the population structure and that their simultaneous use can contribute to a better understanding of the population structure. Further conservation efforts in the Black Slavonian pig population should include improved data recording systems to avoid overestimation or underestimation of genetic diversity parameters.

Acknowledgments
This research was partially funded by Croatian Science Foundation, grant number 3396.

References


Commission on Genetic Resources for Food and Agriculture Food and Agriculture Organization. (2011). Molecular genetic characterization of animal genetic resources. FAO.


