DE GRUYTER
OPEN

THE GENETIC STRUCTURE OF FIVE PIG BREEDS MAINTAINED IN POLAND*

Tomasz Szmatoła[1][♦], Katarzyna Ropka-Molik[1], Mirosław Tyra[2], Katarzyna Piórkowska[1], Kacper Żukowski[2], Maria Oczkowicz[1], Tadeusz Blicharski[3]

[1]Department of Animal Genomics and Molecular Biology,
National Research Institute of Animal Production, 32-083 Balice n. Kraków, Poland
[2]Department of Genetics and Animal Breeding,
National Research Institute of Animal Production, 32-083 Balice n. Kraków, Poland
[3]Department of Animal Immunogenetics,
Institute of Genetics and Animal Breeding PAS, Jastrzębiec, Poland
[♦]Corresponding author: tomasz.szmatola@izoo.krakow.pl

Abstract

The aim of this study was to evaluate genetic variability of five pig populations maintained in Poland with the usage of model based clustering method (MCMC approach) on microsatellite data. A high degree of genetic diversity with the values corrected for sample sizes was observed for four breeds: 0.740 for Polish Landrace, 0.697 for Pietrain, 0.692 for Polish Large White and 0.688 for Puławska. Duroc breed, however, was characterized by the lowest genetic diversity (0.589), allelic richness and number of effective alleles. As for Structure software results, each breed was represented by their own cluster while maintaining a possible small admixture of other breeds. The results obtained suggest a moderate gene flow between breeds, especially between Polish Landrace and Polish Large White has occurred. The lowest admixture was presented for Duroc breed, which confirms its high purity. The presented results can be used to study the structure and as a genetic information for the preservation and further genetic improvement of the pig breeds maintained in Poland. It seems also advisable to conduct further studies with a larger number of microsatellites and by analysis of mitochondrial DNA.

Key words: genetic structure, pig breeds, microsatellite

Conservation of local pig breeds is essential for the preservation of biodiversity, especially when many pig breeds are on the edge of extinction. There is a need to estimate genetic variability of various pig breeds, both endangered and those used in breeding, as a first step to achieve this goal (Toro et al., 2011). Recognition of genetic population structure is crucial to identify the differences among breeds and prioritize approaches of their conservation. Puławska breed is a local Polish pig breed originating from the crossing of native pigs maintained in the vicinity of Lublin with

Berkshire pigs. Nowadays it is considered as a genetic reserve. As the first, Babicz et al. (2003) has studied genetic structure of Puławska breed with the use of 14 microsatellites and 6 gene polymorphisms.

Microsatellites can be utilized as an estimator of population structure, inducing genetic relationship among subpopulations (Estoup et al., 1996). Currently, there is a number of software that enables this and one of the most common is Structure (Pritchard et al., 2000). It is a model based clustering method that uses Markov Chain Monte Carlo (MCMC) approach. It allows detecting genetic populations within a group of individuals genotyped at various multiple markers and also identifies the proportion of their genome that derives from inferred populations. With the use of Structure software it is possible to detect estimations of structures by selecting individuals to different clusters (K) without the knowledge to which population or breed a given individual belongs (Druml et al., 2012).

The results for genotype membership coefficients obtained through Structure software may not be completely accurate. This may be due to individual multilocus genotypes being separated into the selected number of clusters through the use of approaches based on stochastic simulation. It can result in a generation of clusters with various membership coefficients, although the initial conditions were identical. Jakobsson and Rosenberg (2007) explained this by "label switching" and "genuine multimodality" phenomena. "Label switching" of clusters happens when different replicates are matched with an identical estimate of membership coefficient, though with a different version of cluster labels that leads to obtaining similar results in different clusters (Jasra et al., 2005; Stephens, 2000). This, after summarizing of the runs with the highest likelihood, results in incorrect calculations of the proportions of clusters. "Genuine multimodality", however, may arise from actual biological factors that may correspond to equally suitable explanations of the data.

The aim of this study was to evaluate genetic variability of four pig populations that were used for meat production (Pietrain, Duroc, Polish Landrace and Polish Large White pure breeds) and one native pig representing a conservative breed (Puławska pigs).

Material and methods

Analysis was performed on animals representing 5 pig breeds used in breeding program as dam-lines (Polish Landrace – PL (282), Polish Large White – PLW(170), sire-lines (Pietrain – 99 , Duroc – 96) and a breed that constitutes a genetic reserve – Puławska pigs (96). All animals were maintained in the Pig Performance Testing Stations (SKURTCh) of the National Research Institute of Animal Production in Pawłowice, Mełno, Chorzelów and Rossocha under the same housing and feeding conditions. The gilts were fed *ad libitum* from 30 kg up to 100 (±2.5) kg when they were slaughtered and dissected.

The genomic DNA was isolated from whole blood using the Genomic Wizard Purification Kit (Promega) according to the instruction provided in the protocol. The

analysis of selected microsatellite sequences (Sw589, S0073, Sw445, S0001, S0097) described previously by Rohrer et al. (1994, 1996) was performed using Fragment Analysis application on sequencer CEQ8000 Genetic Analysis System with Genome-eLab DNA Size Standard Kit – 600 (Beckman Coulter). The microsatellites were amplified in one multiplex reaction with the use of AmpliTaq® 360 DNA Polymerase (Life Technology).

Five microsatellites used in the present study were carefully chosen out of the entire length of chromosome 4. To check if the microsatellites are not in linkage disequilibrium χ^2 and D' were calculated with the use of PowerMarker software and the obtained results were presented in Table 1.

Table 1. Linkage disequilibrium statistics for 5 microsatellites. χ^2 on upper triangle and D' on lower triangle of LD matrix (PowerMarker)

Marker	S0001	Sw589	S0073	Sw445	S0097
S0001		0.003	0.004	0.006	0.005
Sw589	0.141		0.016	0.003	0.005
S0073	0.173	0.356		0.006	0.012
Sw445	0.209	0.176	0.262		0.005
S0097	0.168	0.196	0.258	0.207	

Table 2. Estimates of position, F and other important statistics for 5 microsatellites on chromosome (PowerMarker)

Marker	Position		F statistics			Other statistics				
	bp*	cM**	Fst	Fit	Fis	number of genotypes	number of alleles	frequency of major allele	hetero-zygosity	PIC***
S0001	23809394-23809581	41.8	0.103	0.292	0.210	17	7	0.552	0.439	0.546
Sw589	96936045-96936191	74.4	0.083	0.049	–0.037	26	7	0.281	0.793	0.791
S0073	94938773-94938883	74.4	0.162	0.285	0.146	42	10	0.402	0.576	0.749
Sw445	123094983-123095169	105.8	0.102	0.184	0.092	48	10	0.293	0.701	0.819
S0097	133640993-133641199	120	0.101	0.089	–0.013	28	9	0.279	0.727	0.742

*physical position referred to susscrofa10.2 (Ensembl database).
** genetic position referred to MACR database.
***Botstein et al., 1980.

Program PowerMarker was also used to calculate microsatellite data presented in Table 2 which are as follows: F statistics, number of alleles and genotypes, frequency of major allele, heterozygosity and PIC (Botstein et al., 1980). Moreover, we

used PowerMarker to calculate genetic distances among breeds (Nei et al., 1983). SPAGeDi software was used to calculate differences among breeds presented in Table 3 which are as follows: number of alleles, number of effective alleles (Nielsen et al., 2003), allelic richness and gene diversity corrected for sample size (Nei, 1978).

Table 3. Various important statistics describing characteristics of 5 pig breeds

Breed	Number of individuals	Number of alleles	Number of effective alleles*	Allelic richness	He**
Duroc	96	4.60	2.78	4.60	0.589
Polish Landrace	282	7.40	4.14	7.15	0.740
Pietrain	99	6.80	3.63	6.79	0.697
Puławska	96	6.40	3.34	6.40	0.688
Polish Large White	170	6.60	4.08	6.57	0.692

*(Nielsen et al., 2003).
**gene diversity corrected for sample size (Nei, 1978).

To identify the population structure and the pattern of admixture within populations we used Structure software that was run 10 times per each K, initially from K1 up to K10. Moreover, admixture model was used and the parameters of the runs were as follows: 1 000 000 iterations and 500 000 burn-in period.

The Clumpp software was used to solve the problems of "label switching" and "genuine multimodality" (Jakobsson and Rosenberg, 2007). The Structure Harvester web software (Earl and VonHoldt, 2012) was used to prepare the input files for Clumpp software through the modification of the selected best K output files (10 independent runs which corresponded to the highest likelihood ratio) generated by Structure software. To visualize individual genotype membership coefficients obtained through Clumpp within each cluster, Distruct (Rosenberg, 2004) software was used (Figure 1). Then, Baps (Corander et al., 2008) program was used to confirm the Structure results by the identification of clusters in the dataset with the use of "clustering of groups of individuals" option (presented in Figure 2). We have also used Baps to perform an admixture analysis with similar options as utilized in Structure to simulate comparable conditions.
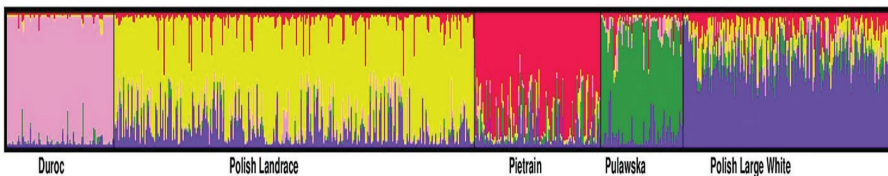

Figure 1. Visualization of individual genotype membership coefficients obtained through CLUMPP software within each cluster
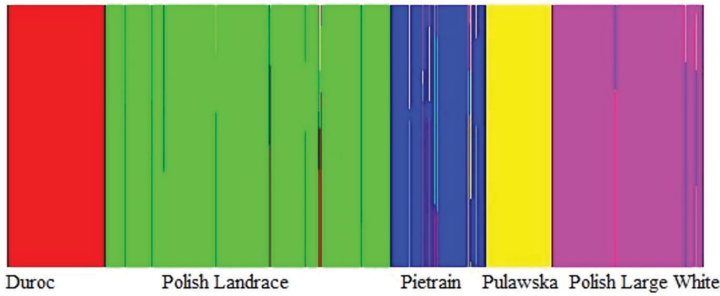
Figure 2. Visualization of genotype membership coefficients obtained through BAPS software within each cluster

Results

In the present study, allele frequencies were analyzed for 743 pigs from 5 pig breeds. Table 2 presents the estimates of position, F and other important statistics for all investigated microsatellites. A total of 43 alleles were observed with the average number of alleles per locus being 8.6 and ranging from 7 (*S0001, Sw589*) to 9 (*S0097*) and 10 (*S0073, Sw445*). The overall F_{IS} values per locus ranged from -0.037 (*Sw589*) up to 0.210 (*S0001*) with the average equal to 0.067, while F_{ST} values ranged from 0.083 (*Sw589*) to 0.162 (*S0073*). The mean F_{ST} calculated for all loci was 0.109 which indicated that 89.1% of genetic variation was associated with differences among individuals while 10.9% was related with diversity among breeds. Heterozygosity of the microsatellites ranged from 0.439 (*S0001*) to 0.793 (*Sw589*). In addition, PIC values obtained in this study ranged from 0.546 (*S0001*) to 0.819 (*Sw445*) and the values obtained for linkage disequilibrium for 5 microsatellites are presented in Table 1.

Table 4. Proportion of membership of each pre-defined population in each of the 5 clusters (Structure results)

Population	Inferred clusters					Number of individuals
	1	2	3	4	5	
Duroc	0.898	0.020	0.014	0.028	0.036	90
PL	0.053	0.734	0.054	0.029	0.131	303
Pietrain	0.034	0.059	0.735	0.103	0.076	106
Puławska	0.057	0.028	0.060	0.764	0.089	69
PLW	0.053	0.131	0.096	0.105	0.616	174
Average distances between individuals*	0.686	0.674	0.719	0.685	0.579	-

*expected heterozygosity.

Table 3 presents genetic diversity with the values corrected for samples sized for all five breeds. Four of the selected breeds (Puławska, Polish Landrace, Polish Large White and Pietrain) represented high values ranging from 0.688 for Puławska to 0.740 for Polish Landrace. Only one breed, Duroc, was characterized by a lower genetic diversity. In addition, the number of alleles and allelic richness represented similar trend – elevated values for Puławska, Polish Landrace, Polish Large White, Pietrain and lower numbers for Duroc. What is more, the highest number of alleles was shown for Polish Landrace (7.40) and the lowest for Duroc (4.60).

The results of population structure and degrees of admixture of 5 pig breeds are presented in Table 4 and Figure 1. The highest likelihood for both Structure and Baps software was obtained for K=5 which corresponded to the number of breeds used in the study.

Discussion

Polymorphic information contents (PIC) are useful to determine the extent of polymorphism of markers for each locus and their values in this study ranged from 0.546 (*S0001*) to 0.819 (*Sw445*). In general, PIC values above 0.5 are considered to be high (Zajc and Sampson, 1999). Four of the five microsatellites used in this study were highly informative with PIC values above 0.7 and only one had lower PIC value but still above 0.5. As shown in Table 1 microsatellites used in the study were not in linkage disequilibrium (LD). This statement, in conjunction with high PIC values, confirms the usefulness of selected microsatellites for analysis of genetic structure of pig breeds. Moreover, the results of F_{ST}, F_{IS} and heterozygosity in our study in general correspond to the results obtained by other authors (Martinez et al., 2000; Sahoo et al., 2016). Similar results were presented by Babicz et al. (2003) regarding Puławska breed. Considering the small size of the population, authors have confirmed the possibility of usage of investigated markers for analysis of genetic profile of Puławska breed, which might be used in future breeding work.

As shown in Table 3 and presented in the results section, high genetic diversity with the values corrected for samples sized was obtained for Polish Landrace (0.740), Pietrain (0.697), Polish Large White (0.692) and Puławska (0.688). In contrast, the lowest genetic diversity (0.589), as well as the lowest allelic richness and the number of effective alleles, was observed for Duroc breed. These findings are comparable with those reported by Kim et al. (2005), who studied genetic diversity among European, Korean and Chinese pig breeds. In their results authors indicated that genetic diversity of European pig breeds was the lowest for Duroc (0.599) and the highest for Landrace (0.702). This also corresponded to the lowest allelic richness in Duroc breed (3.82). Similar results were presented by Druml et al. (2012) whose research was focused on determination of genetic diversity and population structure of 8 Balkan pig breeds. They showed that the highest gene diversity of 0.66 was obtained for Pietrain breed over local breeds which is comparable with the value obtained in our study.

For the selected five breeds used in this study, the lowest number of alleles was shown for Duroc breed (4.60) and the highest for Polish Landrace (7.40) with allelic richness showing the same trend (4.60 and 7.15, respectively). These results correspond to those observed for other European pig breeds but are slightly lower than the ones reported for Asian breeds (Gama et al., 2013). In addition, genetic distances among populations calculated according to Nei et al. (1983) have shown that the population of Duroc pigs occupies the most distant position on the unrooted neighbor joining dendrogram (Figure 3). On the other hand, close genetic distance of PLW and PL as well as Pietrain and Puławska pigs suggests possible genetic exchange among them.
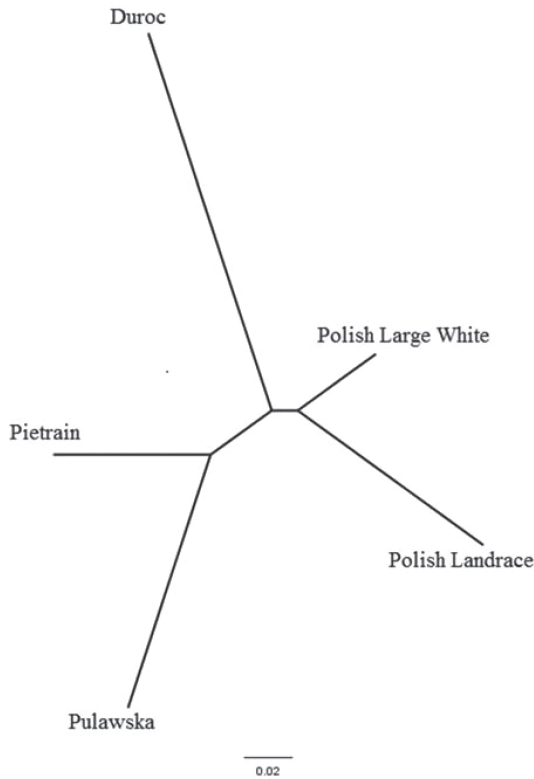

Figure 3. Genetic relationship among five pig populations with the use of neighbor joining tree obtained from Nei83 method based on genetic distance

The population structure and degree of admixture of 5 pig breeds was assessed using Structure algorithm. The highest likelihood was obtained for K=5 allowing for a meaningful interpretation. Ten runs with the highest likelihood were used as an input file for Clumpp software and after the analysis Table 4 and Figure 1 were generated. Each breed is represented by their own cluster while maintaining an ad-

mixture of other breeds. The results obtained suggest that a moderate gene flow between breeds, especially between Polish Landrace and Polish Large White, have occurred. This could be related to a similar direction for use of breeding of both dam-line breeds. The lowest admixture is shown for Duroc breed. The Duroc breed is characterized by high proportion of meat in the carcass, low fat content in carcass as well as high intramuscular fat content and in general high growth rate (Piórkowska et al., 2013; Pierzchała et al., 2011). The present research confirmed high purity of Duroc breed, which may be indicated by the fact that this breed has been created in the USA by crossing of Red Guinea pigs and Iberian pigs with Berkshire breed and was imported to Poland relatively recently – by the 1970s. There is also a visible small admixture of PLW and Pietrain in the Puławska breed. This is probably because Puławska breed was created as a mixture of Polish primitive pigs with English Large White, Berkshire and Tamworth. Moreover, PLW breed was created on the basis of English Large White as well, so these two breeds have the same relative (Szydler-Nędza et al., 2008). More precise data on genetic relationship between breeds might be obtained by mitochondrial DNA analysis.

The results of Baps with the option "clustering of groups of individuals" confirmed the results of Structure, selecting the best K as 5. It is important to mention that admixture model used in Baps is less restrictive than the one used in Structure, which led to obtaining results that were similar, though not as precise and accurate when describing cluster proportions. This was especially visible in the case of Duroc and Puławska breeds which in 100% belong to one cluster for each breed.

To conclude, these findings can be used to study the structure and genetic information for the preservation and further genetic improvement of the pig breeds maintained in Poland. It seems advisable to conduct further studies with a larger number of microsatellites and by mitochondrial DNA analysis.

References

B a b i c z M., K u r y ł J., W a l k i e w i c z A. (2003). Evaluation of the genetic profile of the Pulawska breed. J. Appl. Genet., 44: 497–508.

B o t s t e i n D., W h i t e R.L., S k o l n i c k M., D a v i s R.W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am. J. Hum. Genet., 32: 314–331.

C o r a n d e r J., M a r t t i n e n P., S i r é n J., T a n g J. (2008). Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations. BMC Bioinformatics, 9: 539.

D r u m l T., S a l a j p a l K., D i k i c M., U r o s e v i c M., G r i l z - S e g e r G., B a u m u n g R. (2012). Genetic diversity, population structure and subdivision of local Balkan pig breeds in Austria, Croatia, Serbia and Bosnia-Herzegovina and its practical value in conservation programs. Genet. Sel. Evol., 44: 5.

E a r l D.A., V o n H o l d t B.M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv. Genet. Res., 4: 359–361.

E s t o u p A., S o l i g n a c M., C o r n u e t J.M., G o u d e t J., S c h o l l A. (1996). Genetic differentiation of continental and island populations of *Bombus terrestris* (Hymenoptera: Apidae) in Europe. Mol. Ecol., 5: 19–31.

G a m a L.T., M a r t i n e z A.M., C a r o l i n o I., L a n d i V., D e l g a d o J.V., V i n c e n t e A.A., V e g a - P l a J.L., C o r t e s O., S o u s a C.O., BIOPIG Consortium (2013). Genetic structure, re-

lationship and admixture with wild relatives in native pig breeds from Iberia and its islands. Genet. Sel. Evol., 45: 18.

J a k o b s s o n M., R o s e n b e r g N.A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics, 23: 1801–1806.

J a s r a A., H o l m e s C.C., S t e p h e n s D.A. (2005). Markov chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling. Stat. Sci., 20: 50–67.

K i m T.H., K i m K.S., C h o i B.H., Y o o n D.H., J a n g G.W., L e e K.T., C h u n g H.Y., L e e H.Y., P a r k H.S., L e e J.W. (2005). Genetic structure of pig breeds from Korea and China using microsatellite loci analysis. J. Anim. Sci., 83: 2255–2263.

M a r t i n e z A.M., D e l g a d o J.V., R o d e r o A., V e g a - P l a J.L. (2000). Genetic structure of Iberian pig breed using microsatellites. Anim. Genet., 31: 295–301.

N e i M. (1978). Estimation of average heterozygosity and genetic distance for small number of individuals. Genetics, 89: 583–590.

N e i M., T a j i m a F., T a t e n o Y. (1983). Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. J. Mol. Evol., 19: 153–170.

N i e l s e n R., T a r p y D.R., K e r n R e e v e H. (2003). Estimating effective paternity number in social insects and the effective number of alleles in a population. Mol. Ecol., 12: 3157–3164.

P i e r z c h a ł a M., S h e k h a r P a r e e k C., L i s o w s k i P., U r b a ń s k i P., G o l u c h D., C z a r - n i k U., K a m y c z e k M., R ó ż y c k i M., C o o p e r R.G., K u r y l J. (2011). Expression profile of *MYF5* and *MYF6* genes in skeletal muscles of young growing gilts of five breeds at different ages, based on the most stable reference genes. Anim. Sci. Pap. Rep., 29: 231–246.

P i ó r k o w s k a K., R o p k a - M o l i k K., O c z k o w i c z M., R ó ż y c k i M., Ż u k o w s k i K. (2013). Association study of PIT1 and GHRH SNPs with economically important traits in pigs of three breeds reared in Poland. Anim. Sci. Pap. Rep., 31: 303–314.

P r i t c h a r d J.K., S t e p h e n s M., D o n n e l l y P. (2000). Inference of population structure using multilocus genotype data. Genetics, 155: 945–959.

R o h r e r G.A., A l e x a n d e r L.J., K e e l e J.W., S m i t h T.P.L., B e a t t i e C.W. (1994). A microsatellite linkage map of the porcine genome. Genetics, 136: 231–245.

R o h r e r G.A., A l e x a n d e r L.J., H u Z., S m i t h T.P.L., K e e l e J.W., B e a t t i e C.W. (1996). A comprehensive map of the porcine genome. Genome. Res., 6: 371–391.

R o s e n b e r g N.A. (2004). DISTRUCT: a program for the graphical display of population structure. Mol. Ecol. Notes, 4: 137–138.

S a h o o N.R., N e s s a N., N a s k a r S., B a n i k S., P a n k a j P.K., S a h o o M. (2016). Microsatellite and mitochondrial diversity analysis of native pigs of Indo-Burma Biodiversity Hotspot. Anim. Biotechnol., 27: 52–59.

S t e p h e n s M. (2000). Dealing with label switching in mixture models. J. R. Stat. Soc., 62: 795–809.

S z y d l e r - N ę d z a M., B l i c h a r s k i T., B a j d a Z. (2008). Puławska pig – factors affecting the population size 1932–2007. Wiad. Zoot., 4: 37–40.

T o r o M.A., M e u w i e s s e n T.H.E., F e r n á n d e z J., S h a a t I., M ä k i - T a n i l a A. (2011). Assessing the genetic diversity in small farm animal populations. Animal, 5: 1669–1683.

Z a j c I., S a m p s o n J. (1999). Utility of canine microsatellites in revealing the relationships of pure bred dogs. J. Hered., 1: 104–107.