THE THESIS OF THE PHD DISSERTATION

NGUYEN ANH THI

Kaposvár

2025

HUNGARIAN UNIVERSITY OF AGRICULTURE AND LIFE SCIENCES
KAPOSVÁR CAMPUS

DOCTORAL SCHOOL OF AGRICULTURAL AND FOOD SCIENCES

Head of Doctoral School:

PROF. DR. MELINDA KOVÁCS, DSc.

Supervisor:

PROF. DR. ISTVÁN NAGY, DSc.
Professor

Co-supervisor:

DR. GYÖRGY KÖVÉR, PhD.
Associate Professor

ANALYSIS OF THE INBREEDING EFFECTS ON THE POPULATION STRUCTURE AND GENETIC DIVERSITY OF THE MANGALICA PIG BREEDS

Written by:

NGUYEN ANH THI

Kaposvár
2025

Appendix 1.

The PhD School

Name: Doctoral School of Agricultural and Food Sciences

Discipline: Animal Sciences

Head of the Doctoral School:
Prof. Dr. Melinda Kovács
University Professor, Full Member of the Hungarian Academy of Sciences

Supervisor(s):
Dr. István Nagy
Head of Department, Professor, DSc.
Hungarian University of Agriculture and Life Sciences,
Institute of Animal Sciences, Department of Animal Breeding

Dr. György Kövér
Associate Professor, PhD.
Hungarian University of Agriculture and Life Sciences,
Institute of Animal Sciences, Department of Animal Breeding

....

Approval of the Head of Doctoral School Approval of the Supervisor(s)

TABLE OF CONTENTS

1. BACKGROUND AND OBJECTIVES OF THE RESEARCH

Maintaining genetic variability of a certain population is one of the crucial strategies for an effective genetic conservation program as it ensures the adaptation of the population to environmental change and genetic selection. Genetic variability is basically measured by the expected heterozygosity Nei (1973) known as the probability that two randomly chosen alleles from the population are different. The heterozygosity and homozygosity are two opposing states involved in genetic diversity (GD) in the process by which alleles frequency is passed down from a certain generation to subsequent generations. The reduction of heterozygosity is due to various reasons from which inbreeding is probably the most substantial because inbreeding is recognized to increase homozygosity (Curik et al., 2017). According to Woolliams (2007) inbreeding even causes higher rate loss of alleles in a closed and small population, since the contribution of founders to the gene pool is not equal. Therefore, studying genetic variability and structure of a population provides valuable information for a conservation program to approach dual targets: inbreeding reduction and high genetic variation maintenance.

The inbreeding levels are traditionally estimated according to the method proposed by Wright (1922) where inbreeding coefficient is defined as the probability that the two alleles at any locus in an individual are identical by descent. In recent decades, some improved and extended methods for inbreeding coefficient estimation have been published such as ancestral inbreeding coefficient (e.g. Ballou, 1997). This method approached the inbreeding with consideration of whether any allele in an individual has been autozygous at least once in previous generations. It is assumed that an inbred progeny of non-inbred ancestors is more vulnerable to negative effect of inbreeding than the inbred progeny of inbred ancestors (Suwanlee et al., 2007).

Among the conservation programs in Hungary probably the most characterizing is that of the Mangalica pigs, which is an autochthon Hungarian pig breed group preserved for its biological diversity after having held a vital role in the commercial lard supply chain (Egerszegi et al., 2003, Rátky et al., 2007). Currently, the Mangalica pig exists in three well-known breeds - Blonde, Red, Swallow-Belly which molecular genetic analysis by Zsolnai et al. (2006) confirmed as distinct breeds. Posta et al. (2016) studied the population structure and GD based on the data of the pedigree up to 2011 of the Mangalica populations in Hungary. However, during the last decade, the population of these pig breeds

has changed significantly in numbers and an internal change in genetic variation was highly probable. A comprehensive analysis of updated data from the herdbook up to 2021 is important to elucidate these changes in population structure, refine our understanding of the genetic parameters related to these native breeds, and serve as an information basis for further study. In addition, in terms of breed loss, it is necessary to study the population substructure and migration in order to make an appropriate assessment of the management and conservation of genetic variability (Cervantes et al., 2008) as all breeds are kept in several herds, these herds can be interpreted as subpopulations.

1.1. Research objectives

1. To assess the genetic variability in Mangalica pig breeds using pedigree data, key metrics such as inbreeding coefficients, effective population size will be calculated. This evaluation will help identify levels of genetic diversity, detect bottlenecks and uncover historical breeding patterns that support preservation of genetic diversity and control of inbreeding rates.

2. To calculate the probability of gene origin, pedigree information will be used to trace the ancestral contributions within a population. The distribution of gene origins can be estimated, providing valuable insights into the genetic structure and the influence of different founders on the current gene pool.

3. To calculate the inbreeding coefficients, different approaches will be employed, allowing for a comparison of the results and an assessment of a new-inbreeding coefficient, which is expected to have more negative effects. In addition, the correlation between these various inbreeding coefficients will be evaluated.

4. To monitor the potential population subdivision among different herds of Mangalica pigs, pedigree data will be analysed to assess the potential for genetic differentiation between the herds. This will involve calculating genetic distances using Wright's F_{ST}. The results will provide valuable insights into population dynamics and potential isolation.

5. To analyze migration patterns between herds of Mangalica pigs, the birth identified herds and current identified herds recorded in the pedigree will be used to track movements among different herds. This approach provides insights into how these movements affect genetic diversity and population structure across various herds.

3. MATERIALS AND METHODS

3.1. Genealogical data

Data for analyzing in this study were provided by the Hungarian National Association of Mangalica Breeders. The organization recorded the data of the registered Mangalica pigs in the Herdbook born between 1981 and 2023, respectively. Genealogy analysis was restricted to breeding animals.

From the examined Mangalica pig breeds the pedigree of the Blonde Mangalica was the largest containing 12,808 individuals (up to 2021). The Swallow Belly and the Red Mangalica pedigrees were smaller, and they consisted of 2,393 and 3,963 pigs, respectively. The pigs born in the period from 2016 to 2021 (REF2021) were determined as reference population which was used to assess the genetic variability. Besides, several reference populations were defined for pigs born between 1981-1985, 1986-1991, 1992-1997, 1998-2003, 2004-2009 and 2010-2015, respectively to characterize the evolution of pedigree-based parameters.

For population subdivision and mingration assessment, the genealogical data were extended up to 2023.

3.2. Pedigree completeness

The pedigree completeness was evaluated by: (1) the number of full generations traced, (2) the maximum number of generations traced (3) the complete generations equivalent (CGE) for each animal in the pedigree data (4) the pedigree completeness

3.3. Diversity parameters

3.3.1. Average relatedness coefficient (AR) and inbreeding coefficient

AR 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 (Gutiérrez and Goyache, 2005)

Inbreeding coefficient was defined as the probability that the two alleles at any locus in an individual are identical by descent (Wright, 1922) and was computed following Suwanlee at al. (2007).

$$F_{W(x)} = \sum \left(\frac{1}{2}\right)^{n+n'+1} \left(1 + f_{a(c)}\right)$$

In which x is the individual, n and n' are the number of generations from sire and dam respectively to the ancestor in question, and $f_{a(c)}$ is the inbreeding coefficient of the common ancestors.

Ballou (1997) defined inbreeding as the cumulative proportion of an individual's genome that has been previously exposed to inbreeding in its ancestors. The Ballou's inbreeding coefficient was calculated as the following formula.

$$F_{Bal(x)} = \frac{\left[F_{Bal(s)} + \left(1 - F_{Bal(s)}\right) * F_{W(s)} + F_{Bal(d)} + \left(1 - F_{Bal(d)}\right) * F_{W(d)}\right]}{2}$$

Where $F_{Bal(x)}$ is the ancestral inbreeding coefficient for an individual *(x)*, F_W is the Wright's inbreeding coefficient and the subscripts *(s)* and *(d)* represent for the sire and dam, respectively.

Kalinowski et al. (2000) defined inbreeding coefficient ($F_{Kal(x)}$) as the probability that any allele in an individual is currently autozygous and has been autozygous in previous generations at least once. The Kalinowski "new" inbreeding coefficient was defined as the probability that any alleles in an individual is autozygous for the first time and was calculated as the deduction of $F_{W(x)}$ and ($F_{Kal(x)}$) (Kalinowski et al., 2000; Schäler et al., 2020).

$$F_{New(x)} = F_{W(x)} - F_{Kal(x)}$$

3.3.2. Generation interval (GI)

The GI determined the average age of parents at the birth of their progeny that were subsequently kept for reproduction (Gutiérrez and Goyache, 2005), being computed for the 4 pathways (sire – son, sire – daughter, dam – son, and dam – daughter).

3.3.3. Effective population size (N_e)

N_e is defined as the number of breeding animals that would lead to the actual increase in inbreeding if they contributed equally to the next generation (Wright, 1931). In this study, the N_e was estimated from individual increase in

inbreeding (ΔF_i) called as "realized effective population size - $\overline{N_e}$) (Cervantes et al., 2008a; Gutiérrez et al., 2009).

$\overline{N_e} = \frac{1}{2\overline{\Delta F}}$ in which $\overline{\Delta F}$ represents the average increase in inbreeding across individuals in a reference subpopulation, calculated from individual inbreeding increases (ΔF_i)

3.3.4. Predicting future inbreeding

The inbreeding coefficients in the next 25 years were predicted and the relevant populations were classified in the five categories of endangerment according to (Alderson, 2009). The predicted inbreeding coefficients were the sum of initial inbreeding coefficients (t = 0) of reference populations and the expected inbreeding increase in the next 25 years (t = 25). Using the annual inbreeding rate (ΔF_y) the inbreeding increase was calculated as $F_t = 1 - (1-\Delta F_y)^{t-1}$ (Falconer and Mackay, 1996; Gutiérrez et al., 2009).

3.3.5. Effective number of founders (f_e)

The effective number of founders (f_e) means the number of founders contributing equally that would be expected to produce the same amount of GD as in the studied population (Lacy, 1989). This parameter was calculated by the equation: $f_e = \frac{1}{\sum_{k=1}^{f} q_k^2}$

where q_k is the genetic contribution of the *k*th founder (the proportion of genes in the current living, descendant population that originated from founder k) reference population and f is the total number of founders. Total number of founders (f) was defined as ancestors with unknown parents. Under the selection, the contribution of reproductive individuals is usually unequal so that another parameter known as effective number of ancestors was calculated to complement for the loss of genetic variability.

3.3.6. Effective number of ancestors (f_a)

The effective number of ancestors (f_a) is the minimum number of ancestors (not necessarily founders), explaining the complete genetic diversity of the population (Boichard et al., 1997). This parameter is calculated as $f_a = \frac{1}{\sum_{j=1}^{a} q_j^2}$

where qj 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.

3.3.7. Effective number of founder genomes (f_g)

Effective number of founder genomes (founder genome equivalents (f_g)) is the number of founders that would produce the same expected heterozygosity in the studied population if the founders equally contributed and were not lost of alleles (Ballou and Lacy, 1995). This parameter was calculated by the method of Caballero and Toro (2000).

3.3.8. Genetic diversity (GD)

The loss of GD was considered in two aspects: the unequal contribution of founders and the genetic drift. The bottleneck effect was characterized by the ratio f_a / f_e, occurring when the effective number of ancestors is much smaller than the effective number of founders or the ratio is much smaller than one.

Total GD of the reference population was estimated according to (Lacy, 1995) with GD = $1–1/(2f_g)$. The loss of genetic variation in both aspects was inferred as 1– GD. Similarly, the total amount of genetic diversity in the reference population possibly disappeared due to unequal founder contribution was calculated as GD* = $1–1/(2f_e)$ and the actual GD loss with this reason was measured as 1–GD*. Finally, the difference of GD* and GD was calculated to expose the genetic diversity loss because of genetic drift (Caballero and Toro, 2000; Honda et al., 2004).

3.4. Population subdivision

The genealogical data were used to analyse the structure of the subpopulations using the concept of Wright's F–statistics (Wright, 1978), calculated according to Caballero and Toro (2000) for each specified subpopulation. The F_{ST} matrix was visualised using a heat map. The normality of the F_{ST} coefficients was evaluated by an Anderson–Darling normality test. We checked all F_{ST} values for every herd from which the largest value for each herd was taken. Then, we sorted these (maximum) F_{ST} values from lowest to highest and calculated the proportion of the total population they represent. Finally, the results were depicted as histograms and the cumulative population proportion according to the maximum F_{ST} values per each herd.

3.5. Migration assessment

The actual migration of pigs among herds derived from the stud book was visualised by the chord diagram. Every individual's herd ID was documented from birth to the most recent assessment. Those with incomplete ID information at either birth or the present assessment were excluded from the parameter analysis. Chord diagrams were employed to illustrate the transition of individuals from their birth herds to another current herd, with separate diagrams for males and females, as well as a combined one. However, to enhance clarity, only individual chord diagrams for male and female migrations were presented.

3.6. Programme used

ENDOG software programs version 4.8 (Gutiérrez and Goyache, 2005 was utilized to calculate all pedigree parameters and the F_{ST} matrix based on the differentiation between the herds (pairwise F_{ST} coefficients). The function heatmap.2 of the R package "gplots" was used to create a heatmap to visualise the pairwise F_{ST} matrix.

Different inbreeding coefficients were determined using the GRain software version 2.2 (Doekes et al., 2020, Baumung et al. 2015). Spearman correlations and their significance among the different inbreeding coefficients were tested by the Performance Analytics-package in R-studio.

The chordDiagram function from the R package "circlize" (Gu, 2014) was used to create the chord diagram of migration intensity.

4. RESULTS AND DISCUSSION

4.1. Characterization of the pedigree structures and completeness

The Blonde and Red shared similarity of more than 90.00% of pedigree completeness up to the sixth generation while that proportion of the Swallow-Belly was observed in the fifth generation.

The CGE values of the evaluated Mangalica populations ranged between 4.29 and 9.73 and between 6.78 and 9.73 in the whole period and in REF2021, respectively. The relatively large CGE (> 4.00) in the current study indicates that all of the analyzed Mangalica pedigrees were sufficiently deep and complete making reliable genetic variability and population structure analyses possible.

4.2. Average relatedness coefficient (AR) and inbreeding coefficients

In general, most parameters increased continuously except for F_{New} which decreased in all breeds in the last decade. The rate of increase was the highest for F_{Bal} which reached high values by the end of the examined period, especially for the Blonde and for the Red population where their mean values were 23.04% and 27.14%, respectively. Compared to those breeds the Swallow-Belly Mangalica breed showed substantially lower mean F_{Bal} value (13.88% in the REF2021).

The inbreeding level calculated by different methods was generally the highest in the Red and smallest in the Swallow-Belly. For examples, the F_W in the REF2021 was 7.54%, 5.50% and 4.30% for the Red, Blonde, and Swallow-Belly populations while F_{Kal} was 4.25%, 2.68%, 1.35%, respectively. The Red population also has highest F_{New} value (3.29%) compared to the lowest one (2.82%) of the Blonde population. F_{New} showed a decreasing tendency in the final decade (2010-2021) with the corresponding increase of F_{Kal}, which is favourable for the conservation program of this Red population from the aspect of possible inbreeding depression.

AR and F_W coefficients increased in parallel during the last two decades for all breeds and similarly to that of F_{Bal} the highest and lowest values were observed for the Red and for the Swallow Belly Mangalica breeds, respectively. Nevertheless, in the REF2021 68.00% of the F_W of the Swallow Belly Mangalica breed was composed of F_{New} while in the other two breeds (Blonde and Red) the proportion of ancestral inbreeding F_{Kal} exceeded (51.00% and 56.00%, respectively) that of the new inbreeding.

For the Swallow-Belly population, the inbreeding coefficient F_W was mainly composed of the F_{New}, meaning recent inbreeding events predominate, which likely exacerbates inbreeding depression. Besides, in this population the F_{New} showed a very strong correlation with F_W while these correlations were lower on the two other breeds.

The average relatedness (AR) sharply increased in the Red population from 1991 to 2003, ranking first in REF2021 with 12.67% comparing to 8.50% and 6.59% of the Blonde and Swallow-Belly populations, respectively. These parameters explain the highest and lowest inbreeding coefficients of the Red and the Swallow-Belly, respectively, because mating of closer relatives causes higher inbreeding levels.

The estimated correlation coefficients among the different inbreeding coefficients were all positive except for F_{New} and F_{Bal} which were significantly negative in all breeds, but the strength of correlation was negligible. In general F_W showed strong correlation with F_{New} and low correlation with F_{Bal} while the correlation coefficients among the two types of ancestral inbreeding coefficients (F_{Bal} and F_{Kal}) ranged between moderately strong and strong in every breed.

The weak correlation between F_{New}/F_W and F_{Bal} indicates that recent and ancestral inbreeding metrics provide distinct information, which could be useful for comprehensive assessment of inbreeding load in conservation or breeding programs. The slight negative correlation between F_{New} and F_{Bal} suggests a possible trade-off between recent and historical inbreeding, but its practical impact is small.

4.3. Generation interval (GI)

The average GI for Blonde population was the shortest of 3.06 years and the longest one belonged to Swallow-Belly population with 3.45 years. The difference in average GI between pathways was not significant.

The breeding animals were well maintained in this Mangalica conservation program such that the GI increased nearly by one-year (3.27 vs 2.46 years) compared to another research for the same Mangalica breeds.

4.4. Effective population size (N_e)

The Blonde and the Red Mangalica breeds showed the largest and the smallest population sizes throughout the whole examination period. Compared to

their starting values the effective population size doubled in the Red (20 – 45) and Swallow-Belly (46 – 95) and tripled in the Blonde (45 – 156) population. Then all three breeds showed erratic changes where their effective population sizes were subsequently increased and decreased several times. In the REF2021 of all Mangalica breeds the estimated effective population sizes were within the ranges of 50 and 100, respectively.

None of three effective population sizes is smaller than 50 individuals but also none of them is bigger than 500 individuals, expecting that the studied Mangalica populations can avoid inbreeding depression in a short term. The effective population size of every breed has increased during the last decade, which also signals the efficiency of the conservation programs conducted of these breeds. The increasing of N_e in the REF2021 comparing to the next previous stage contributed to lowering the genetic variation loss in the Swallow-Belly (62 – 77) and Red populations (53 – 58) but not in the Blonde (88 – 86) since the effective number of founders genome decreased in this population.

4.5. Predicting future inbreeding

It can be seen that the future inbreeding level for the next 25 years only exceeded the 10% threshold in the Red Mangalica breed which falls the least severe endangered "Transitional" category while the other two Mangalica breeds are non-endangered.

Among the reference populations, F_W of the Red population was the highest (7.54%) and the rate of inbreeding was also higher around 0.87% per generation. For the Blonde and Swallow-Belly Mangalica populations, although they showed similar inbreeding rate (approximately 0.60% per generation), the Blonde has higher current inbreeding coefficient (5.50% vs. 4.30%). Thus, as predicting the Blonde might be likely to enter the endangered category in the future than the Swallow-Belly population. However, at present none of these two populations have reached any category of endangerment regarding only inbreeding coefficient, meaning that the existence and genetic diversity of these populations are in optimistic perspective at least in the next 25 years.

4.6. Probability of gene origin

All populations showed increased number of founders, effective number of founders, effective number of ancestors and effective number of founder genomes, respectively. Then all population witnessed decreases from 2009 for all

of the above-mentioned parameters where the largest values were recorded for Blonde (138, 49, 21, 8.91, respectively) and the smallest ones for the Red Mangalica (76, 26, 14, 5.72, respectively).

The number of effective founders (f_e) were much smaller than the total number of founders in all three REF2021, accounting for 35.51%, 39.56% and 34.21% in the Blonde, Swallow-Belly and Red, respectively. After 1986 the effective number of founder genomes (f_g) were shown a decreased tendency in the Blonde and Red populations. Concerning the Swallow-Belly population, this decrease was only detected from 2004, giving the highest f_g for the Swallow-Belly REF2021. The Blonde and Swallow-Belly reference populations shared the similar number of effective ancestors (f_a) of 21 and 22 individuals, respectively while this number of the Red was smaller with only 14 individuals.

The ratio f_a/f_e was 1.00 between 1981–1985 in the Blonde and Red, meaning that the bottleneck had happened in these two populations afterwards. On the contrary, signals for bottlenecks were already observed in the Swallow-Belly population during the first five years of the examined period. The ratios f_a/f_e in the reference populations (2016 –2021) were ranging from 0.43 to 0.61 in the various breeds meaning that the bottleneck effects were not so severe. The Blonde Mangalica pigs were expected to be the most vulnerable to the bottleneck effects.

The whole gene pool in the Red and Swallow-Belly REF2021 could be explained only by 57 and 78 ancestors, respectively while this parameter in the Blonde REF2021 was almost double with 123. The number of ancestors contributes to 50.00% of gene pool is very small only 6 individuals in the Red, 7 individuals in the Blonde, and 8 individuals in the Swallow-Belly REF2021. From these parameters, the Red population would have suffered the most substantial genetic variability loss.

The genetic diversity of the in the Blonde and Red populations both showed similar decreasing tendencies after an initial increase while this parameter fluctuated throughout the Swallow-Belly breed. However, the GD of Swallow-Belly and Red populations was higher in their REF2021 than at the beginning of the evaluated period. In the REF2021, the Blonde and Swallow-Belly had biggest genetic diversity of approximately 94.00% comparing to the lowest of 91.30% in the Red. This means that the genetic diversity loss in the REF2021 of the Red Mangalica pigs was accounted for the highest proportion of 8.70% and these proportions of the Blonde and Swallow-Belly were 5.60 % and 5.70%,

respectively. The loss of GD was dominated by random genetic drift in comparison to the unequal founder contributions.

The total number of founders (*f)* increased remarkably from 1981 to 2009 in the Blonde and Red populations and up to 2003 in the Swallow-Belly. This indicated that new founders had continuously been migrated into these three populations during a relatively long period, making population structure of the Mangalica breeds temporarily open. Consequently, the contribution of these migrated founders could have enriched the GD of these populations and may have lowered the inbreeding coefficients supposing they had not related each other and to the other members of these populations. It must be noted however, that these migrated "founders" might have caused underestimating of inbreeding coefficients to some extent if they violated the before mentioned assumptions. Only molecular genetic analyzes could clarify the possible effect of the described migration on the estimated parameters. Nevertheless, it can be noted that the CGE within the three Mangalica reference populations (2016–2021) were still quite long ranging between 6.87 and 9.73. Thus, these Mangalica pedigrees still provide sufficient information for estimating inbreeding level and genetic variability.

The f_e of three REF2021 was much smaller than the total number of founders, showing the unequal genetic contribution of founders in the current populations. The imbalance of founder contributions was highest in the Red, followed by the Blonde population, whereas the less unequal contribution was found in the Swallow-Belly population. The ratios between f_e/f in this study range from 34.21% (Red) to 39.56% (Swallow-Belly)

The effective number of ancestors (f_a) followed similar pattern of (f_e) that they increased in some early periods and then decreased. If the f_a is much lower than the f_e, the ratio between f_a/f_e is much lower than 1, that population experienced bottleneck effects. Among the three Mangalica populations, the Blonde reference population has got the highest bottleneck effects due to the lowest ratio f_a/f_e of 0.43, following by the Red (0.54) and the Swallow-Belly (0.61).

Only seven ancestors in the Blonde and six ancestors in the Red and eight ones in the Swallow-Belly contributed to 50% of the current reference gene pool. The 100% of reference gene pool were responsible by 57, 78, 123 ancestors in the Red, Swallow-Belly, and Blonde populations, respectively. Based on the genetic contribution of ancestors, the Blonde population is expected to have higher genetic variability than the Red and Swallow-Belly breeds.

The GD of the Blonde retained 94.40% of its original value which was higher than the Swallow-Belly and Red with 94.30% and 91.30%, respectively. In other words, the GD loss of the Red was the highest, followed by the Swallow-Belly and the Blonde with the loss ranging between 8.70%, 5.70% and 5.60%, respectively. The results showed the GD loss due to random genetic drift accounted for higher proportion than the GD loss due to unequal founder contributions. Although the Blonde Mangalica population experienced more severe bottleneck effect ($f_a/f_e \sim 0.43$), the GD loss of the Red was still the largest among the Mangalica breeds. This can be explained that in a smaller population, the bottleneck causes more severe genetic diversity loss. In general, smaller populations are most likely to experience loss of genetic diversity.

4.7. Population subdivision

The average F_{ST} coefficients were 0.04 for the Blonde and 0.047 for the Swallow-Belly and the Red, which are significant smaller than 0.05 (p-value < 0.05), while these parameters were even smaller (0.03 and 0.04, respectively) for the active herds. The distribution of the F_{ST} coefficients was not normal in any breed ($p < 0.001$). The proportion of herds with $F_{ST} > 0.15$ was 15.96% for the Blonde, 12.41% for the Swallow-Belly and 12.40% for the Red. In addition, the proportion of animals with an F_{ST} bigger than 0.15 was 1.21%, 0.81% and 0.38%, respectively. In the currently active herds, highly differentiated herds with large distances ($F_{ST} > 0.15$) were only observed in the Blonde and Red breeds, accounting for 6.41% and 3.64%, which represents a reduction compared to the total herds. A very small proportion of animal with an F_{ST} bigger than 0.15 was found in the Blonde active herds, with 0.14%, and in the Red one, with 0.09%.

Within the Blonde Mangalica, three active herds (1645, 1630 and 1358) show considerable differentiation from each other. Despite the presence of these widely separated herds, the proportions of animals with F_{ST} values bigger than 0.15 were 0.38% and 0.14% in the overall herds and in the active herds, respectively.

In the Swallow-Belly breed, all active herds in this population had an $F_{ST} < 0.15$, indicating that there was no significant genetic differentiation between the current active herds.

The active herds in this breed showed a significant differentiation between herds 1436, 1646 and 1664. The proportions of animals with a maximum F_{ST} by

0.15 were 0.81% and 0.09% in the overall herds and in the active herds, respectively.

The three Mangalica breeds show a uniform F_{ST} pattern between the herds. The small F_{ST} group ($F_{ST} < 0.05$) accounts for the largest proportion of more than 58% of the total, namely, 71.26%, 61.29% and 58.83% for the Blonde, Swallow-Belly and Red breeds, respectively. Conversely, the large F_{ST} set ($F_{ST} > 0.15$) represents a consistently minimal percentage of around 1.00%. The Red breed has the highest percentage (40.33%) of moderate differentiation between herds ($0.05 < F_{ST} < 0.15$), followed by the Swallow-Belly breed with 37.27% and the Blonde breed with 27.55%. It is noteworthy that the Red breed shows a tendency to separate herds, with the majority of moderately differentiated herds. However, the strong stratification between herds was found in a very small proportion of only 0.84%.

While the proportion of the strongly differentiated herds was below 2.00% for the three breeds, the Blonde breed stands out with the highest average distance value (average F_{ST}) within this group, which is 0.24 (between 0.15 and 0.35). In comparison, the Swallow-Belly and Red breeds are smaller, with an average F_{ST} of 0.20 (between 0.15 and 0.35) and 0.21 (between 0.15 and 0.34), respectively. Conversely, the average F_{ST} values for the small and moderate groups were approximately 0.03 and 0.07, respectively.

Of the total herds, approximately 30% were active, with proportions of 30.23%, 32.98% and 35.71% for the Blonde, Swallow-Belly and Red breeds, respectively. Looking at the genetic distances between active herds, over 99.70% fall into the small and medium F_{ST} groups. This leads to a remarkable decrease in the proportion of large F_{ST} groups, which account for less than 0.30% in all three breeds, except for the Swallow-Belly breed, where the percentage is 0.00%.

The Swallow-Belly and Red breeds showed a stronger tendency towards internal differentiation, with a larger percentage of herds showing moderate genetic differences than in the Blonde breed. Nevertheless, the average F_{ST} coefficients between herds remained similar for all three breeds (0.04). This phenomenon may be explained by the smaller population sizes of the Swallow-Belly and Red breeds.

The populations analysed, which have been listed in pedigrees since 1981, include both active and previously inactive herds. Analysing entire populations provides a comprehensive overview, but precise information on genetic

subdivision depends on the active herds. Genetic differentiation was observed in certain herds across the entire populations analysed. However, among the active herds, these differentiated herds make up a tiny proportion, less than 0.30% (8 out of 3,003). When analysing these herds, e.g., 1645, 1630 and 1358 in the Blonde breed and 1436, 1646 and 1664 in the Red breed, each herd had only one selected sire. When calculating the average herd coancestry, the predominance of self-coancestry contributes to high F_{ST} coefficients. However, despite this observed differentiation, the details of the substructure within the breeds remain unclear with the applied approaches.

4.8. Migration assessment

The migration of individuals within herds was considerable, affecting over 60.00% of the total current herds, specifically impacting 61.63% of the Blonde breed, 75.53% of the Swallow-Belly breed, and 63.64% of the Red breed. A consistent pattern emerged across all three breeds, suggesting that an enormous number of females were transferred between herds, while in comparison, the number of males in every moving remained relatively low. Within the three breeds, the herd numbered 872 was the most active and dominant in providing sires to neighbouring herds.

In the Blonde breed, the maximum number of male pigs migrating from a particular farm to a particular herd was 10, exceeding the numbers for the Swallow-Belly and Red breeds, at 6 and 4, respectively. In contrast, the range for female pigs is much wider, reaching up to around 270 animals. In contrast the numbers for the Swallow-Belly and Red breeds were lower, with 86 and 78 individuals, respectively.

The connectivity among migrating herds revealed that more than 80% were connected by a single sire. More specifically, this percentage was 80.72% in the current Blonde herds, 87.00% in the current Swallow-Belly herds and 90.34% in the current Red herds. At the same time, 90% of these herds established connections involving more than two sows, and this pattern applied to all three breeds.

The results showed a strong migration between the herds of the three breeds, as about 60.00% of the herds are connected to other herds in some way. In addition, more than 90.00% of the migration involved one sire and more than two sows. The extensive exchange of animals between individual herds could be the reason for genetic similarity between herds in this study. Both male and female

individuals play crucial roles in creating robust connections between herds within breeds. It clearly shows that more females were exchanged in the breeds analysed. However, it must be noted that the two sexes show different migration characteristics. The exchange of boars between herds is a continuous process, and generally, it consists of one or few animals. On the contrary, the female exchange is occasional, and its aim can be establishing a new herd, the herd size enlargement of an existing herd or the re-establishment of a previously closed herd. In addition, the Mangalica farms do not use artificial insemination (personal communication with HNAMB), but the boars are moved between herds under control. This could support the contribution to the gene flow between the herds of sires.

5. CONCLUSIONS AND RECOMMENDATIONS

The pedigree analysis of Blonde Mangalica (6 generations), Swallow-Belly Mangalica (5 generations), and Red Mangalica (6 generations) showed sufficient depth and completeness (CGE values of 7.08, 4.29, and 6.17, respectively) to reliably estimate key genetic diversity parameters.

In the reference population (REF2021), the number of ancestors explaining the entire gene pool was 57, 78, and 123 for the Red, Swallow-Bellied, and Blonde Mangalica breeds, respectively, with corresponding genetic diversity losses of 8.70%, 5.70%, and 5.60%. Wright's inbreeding coefficient (F_w) were 5.50 for reference population of Blonde, 4.30 for Swallow-Belly, and 7.54 for Red Mangalica while Kalinowski's new inbreeding coefficients were 2.82, 2.95, and 3.29, respectively.

The Red population has the highest genetic diversity loss, smallest effective population size, and highest inbreeding coefficient that needs to be monitored. In addition, mating plans should be focused on reducing new inbreeding in all Mangalica populations, especially in the Swallow-Belly Mangalica. The estimate inbreeding coefficients over the next 25 years were 9.98, 8.73, and 13.67 for the Blonde, Swallow-Belly, and Red Mangalica, respectively, bringing only Red Mangalica into transitional category.

The correlations between different inbreeding coefficients were calculated. Wright's inbreeding coefficient is moderately to highly correlated with F_{Kal} and F_{New} but shows a low correlation with F_{Bal}. F_{Bal} and F_{Kal} are moderately correlated, whereas F_{Bal} is negatively correlated with F_{New}. These findings indicated that different approaches to estimate inbreeding coefficients capture different aspects of inbreeding.

Analyzing Wright's F_{ST} coefficients by heatmap and calculation cumulative proportion of population relative to different levels of F_{ST}, the substructure within the Blonde, Swallow-Belly and Red breeds could not be found.

The frequency of extensive animal exchange between individual herds and the uniformity of mating strategies confirm genetic homogeneity within these breeds. The patterns observed indicate that the breeds studied, with the aim of

maintaining genetic diversity and minimising the risk of inbreeding, show positive signs consistent with conservation objectives.

Pedigree analysis generally assumes that founder individuals are unrelated and non-inbred, which can underestimate inbreeding coefficients by ignoring potential background relatedness. Integrating genomic data (e.g. single nucleotide polymorphism data) overcomes this limitation, yielding more accurate estimates and revealing undocumented inbreeding or relatedness events absent from the pedigree.

Further research can decompose the inbreeding coefficient into partial inbreeding coefficients to identify ancestors with a high inbreeding load in Mangalica populations. This approach could add more information that enables breeders to develop smarter mating strategies by selecting pairs with lower inbreeding load.

6. NEW SCIENTIFIC RESULTS

1. The genealogical data, updated through 2023, now includes a more comprehensive complete generation equivalent (CGE), reported as 9.73, 6.87, and 9.03 for Blonde, Swallow-Belly, and Red Mangalica, respectively. This updated dataset provides a robust foundation for analyzing potential population struture and genetic patterns with accuracy.

2. In the reference population (REF2021), the number of ancestors explaining the entire gene pool was 57, 78, and 123 for the Red, Swallow-Bellied, and Blonde Mangalica breeds, respectively, with corresponding genetic diversity losses of 8.70%, 5.70%, and 5.60%.

3. Wright's inbreeding coefficient (F_w) were 5.50 for reference population of Blonde, 4.30 for Swallow-Belly, and 7.54 for Red Mangalica while Kalinowski's new inbreeding coefficients were 2.82, 2.95, and 3.29, respectively. Kalinowski's new inbreeding coefficient showed an increasing trend in the Swallow-Belly Mangalica reference population but presented highest value in the Red Mangalica reference population. The estimate inbreeding coefficients over the next 25 years were 9.98, 8.73, and 13.67 for the Blonde, Swallow-Belly, and Red Mangalica, respectively, bringing only Red Mangalica into transitional category.

4. Wright's inbreeding coefficient (F_w) is moderately to highly correlated with F_{Kal} and F_{New} but shows a low correlation with F_{Bal}. F_{Bal} and F_{Kal} are moderately correlated, whereas F_{Bal} is negatively correlated with F_{New}. These findings indicated that different approaches to estimate inbreeding coefficients capture different aspects of inbreeding.

5. The research did not detect the actual genetic differentiation within the breeds of Mangalica. The average Wright's F_{ST} coefficients were 0.04 for the Blonde breed and 0.047 for the Swallow-Belly and Red Mangalica breeds, while these parameters were lower in the active herds at 0.03 and 0.04, respectively.

6. The research examined migration patterns within the Mangalica breeds. The migration of individuals between herds was 61.63% for the Blonde breed and 75.53% and 63.64% for the Swallow-Belly and Red Magalica breeds, respectively.

10. PUBLICATIONS AND PRESENTATIONS

10.1 Publications on the thesis topic

1. **NGUYEN, A. T**., KÖVÉR, G., FARKAS, J., BOKOR, Á., TÓTH, P., NAGY, I. (2023): Analysis of genetic variability and population structure of the Mangalica pig breed using pedigree data. *Livestock Science*, *273*, 105265.

2. NAGY, I., **NGUYEN, T. A.** (2023): Characterizing and Eliminating the Inbreeding Load. *Veterinary Sciences*, *11*(1), 8.

3. **NGUYEN, A. T.**, KÖVÉR, G., TÓTH, P., CURIK, I., BOKOR, Á., NAGY, I. (2024): Population Subdivision and Migration Assessment of Mangalica Pig Breeds Based on Pedigree Analysis. *Animals*, *14*(4), 653.

4. **NGUYEN, A. T.**, NAGY, I. (2024): Physiological and Genetic Aspects of some Fitness Traits Performance in Pigs. *Agriculturae Conspectus Scientificus*, *89*(2), 95-103.

10.2 Oral presentations

1. **NGUYEN, A. T.,** KÖVÉR, G., NAGY, I. (2023): Insight the population structure and genetic diversity of the Red Mangalica pig by pedigree analysis. *Program and Abstract Book of the 32nd Annual Meeting of DAGENE*. 26 p. pp. 15-15, 1 p.

2. **NGUYEN, A. T.,** NAGY, I., FARKAS, J., CURIK, I., KOVER, G. (2023): Ancestral inbreeding and inbreeding-purging models' comparative analysis based on their classification efficiency of the Pannon white rabbit kits' survival at birth. *Book of Abstracts of the 1st Regional Meeting of the European Federation of Animal Science.* 102 p. pp. 90-90, 1 p.

3. **NGUYEN, A. T.**, KÖVÉR, G., TÓTH, P., BOKOR, A., NAGY, I. (2024): Analyzing population subdivision of the Blonde Mangalica breed. *7th Fatty pig & 12th Mediterranean pig meeting*. 74p. pp. 12-12, 1p.