Differentiation of the domestic pig and wild boar using genotyping-by-sequencing

Anna Koseniuk,* Grzegorz Smolucha, Artur Gurgul, Tomasz Szmatola, Maria Oczkowicz and Anna Radko

Accepted January 19, 2023 Published online February 23, 2023 Issue online March 30, 2023

Domestic pigs and wild boars have undergone frequent interspecies crossbreeding; therefore, the presence of hybrids makes it challenging to find genetic markers that distinguish both subspecies. The aim of this research is to identify the DNA regions that underwent strong selection during the domestication of the pig and to give an insight into the genetic diversity of the Polish wild boar and domestic pigs by implementing the genotyping-by-sequencing (GBS) technique. We studied two groups of animals: one consisted of domestic pigs (Landrace, Large White, Duroc, Puławska and Pietrain), while the second group included wild boars from Poland. The filtered single nucleotide polymorphisms (SNP) panel used in this study included 7,298 markers that were spread across 18 porcine autosomes and unmapped contigs. The maximum-likelihood phylogenetic trees and multidimensional scaling (MDS) clearly separated the populations of pigs from the wild boars. We also detected genome regions that demonstrated the most significant genetic differences between the domestic pigs and wild boars. These regions were distributed on eight different autosomes and overlapped with 48 different pig RefSeq genes. The KEGG pathway, Reactome and GO terms were further used to assign a functional significance to the identified genes that were associated with inter alia muscle development (MYOG, MEOX2), pre-weaning mortality stress (MYO7A) and sensory perception (TAS1R3).

Key words: GBS, population study, species identification, SNP.

Anna Koseniuk,* Grzegorz Smolucha, Tomasz Szmatola, Maria Oczkowicz, Anna Radko, National Research Institute of Animal Production, Department of Animal Molecular Biology.
E-mail: anna.koseniuk@iz.edu.pl
Artur Gurgul, Tomasz Szmatola, University of Agriculture in Krakow, Center for Experimental and Innovative Medicine.

For many reasons, pigs are one of the most important livestock species for humans. First, pork is a major source of protein in global animal production. In addition, the pig is an excellent animal model to study the molecular components and processes of some human diseases including cancer (Groenen et al. 2012).

Food adulteration practices commonly include the substitution of wild boar meat products with widely accessible and cheaper, but poorer quality, pig meat (Koutsogiannouli et al. 2010). Therefore, the differentiation of both the pig (Sus scrofa domesticus) and the wild boar (Sus scrofa scrofa) is of huge importance for the identification and
Molecular tests are commercially available that identify animal components in human food, as well as in animal feed. Based on PCR techniques (e.g., RFLP, Real-Time) it is possible to identify the DNA of cattle, horses and poultry in food and in feed for animals (Natonek-Wisniewska et al. 2013; Safdar et al. 2014). Nonetheless, although a porcine component can be identified in a simple manner, the differentiation between a wild boar and a domestic pig in meat products remains challenging (Koutsogiannouli et al. 2010; Lorenzini et al. 2020; Kaltebrunner et. al. 2019, 2020).

Fajardo et al. 2008 conducted studies for the first time that were aimed at the differentiation of wild boars and commercial pig breeds based on combining both nucleus (melanocortin 1 receptor, MC1R) and mitochondrial DNA (D-loop). The analysis of the MC1R gene proved to be more effective than the most common method used for species identification based on the polymorphism of mitochondrial DNA (Fajardo et al. 2008). The NR6A1 gene has also been studied, which is a strong candidate gene for a vertebral QTL. A proline to leucine substitution at NR6A1 codon 192 (p.Pro192Leu) resulted in an increased number of vertebrae in commercial pig breeds (Mikawa et al. 2007). A method of differentiating pigs and wild boars was also tested using short tandem repeats (STRs), a marker type commonly used in parentage verification and population studies (Radko et al. 2021). The STRs alone (Rębala et al. 2016), combined with MC1R polymorphism (Nikolov et al. 2017) and MC1R, NR6A1 or mitochondrial D-loop polymorphism, revealed the hybridisation that exists between pigs and wild boars (Lorenzini et al. 2020; Scandura et al. 2008). Szemethy et al. 2021 tested another approach based on identifying the insertions or deletions (InDels) specific to wild boars. In their case, the bioinformatics pipeline indicated three InDels to marker development during the performed routine genetic testing.

Introducing single nucleotide polymorphism (SNP) based molecular techniques to (sub-)species discrimination has many benefits. The most important of them is the significant number of markers (up to tens of thousands) that can be analysed for one individual. SNP-based molecular techniques have been implemented to seek the SNP markers which, under certain conditions (sufficient genotyping density and relevant reference populations), can differentiate between a species and subspecies (Mary et al. 2022).

Rubin et al. (2012) used a whole-genome resequencing approach. They identified 33 selective sweeps localised on 18 loci covering the NR6A1, PLAG1, LCORL, and OSTN genes. All these genes have a known effect on the phenotype and are potential marker candidates. Another technology, based on a Porcine 60K SNP chip, was used to study the diversity of European pig breeds and to differentiate between domestic pigs and wild boars (Wilkinson et al. 2013). The genomes mostly differed in the regions covering the genes for bone formation (on SSC1), growth (on SSC7) and fat deposition (on SSC7 and SSCX). However, the ability to genetically distinguish between phylogenetically close taxa like pigs and wild boar with a proven interspecies gene flow needs further elucidation.

Although microarrays are the most commonly used technology in genomic studies for small projects, their use is financially unjustified. A microarray study also requires details on the SNP type and the surrounding DNA sequence. For studying a poorly annotated species, the discovery of genome-wide SNPs and population-wide genotyping, genotyping-by-sequencing (GBS) could be an alternative to microarrays (Gurgul et al. 2019).

Genotyping-by-sequencing (GBS), which was originally developed for use in plants (crops), is a relatively simple approach that is applied in population genomic studies. GBS is both versatile and can be used for any species. It is based on the high-throughput, next generation sequencing (NGS) of genomic loci that are targeted by restriction enzymes (Elshire et al. 2011). The technique involves a digestion of the genomic DNA with one of the available frequently-cutting endonucleases and, following the adapters ligation, the sequencing of the obtained short DNA fragments using NGS. This approach reduces the target genome regions for NGS, thus allowing for an analysis of multiple samples (multiplexing abilities) in a single sequencing run. The GBS technique can be used to identify many markers with a relatively low error rate, and the detected polymorphisms are distributed throughout the whole genome. Another advantage of the technique is that it allows for the simultaneous discovery and genotyping of SNPs within a large number of samples. These features facilitate different types of genome-wide analyses such as relatedness studies, genome-wide association studies and genomic selection (Kim et al. 2016), as well as population studies, including among other
things the identification of the selection signatures (Baaazou et al. 2020; Wang et al. 2018). However, the analysis of GBS-derived SNP datasets is complicated by frequent cases of missing genotypes. This is especially relevant in highly heterozygous organisms, due to a dependence on the read depth to correctly identify both alleles (Dodds et al. 2015).

For the first time, we used a GBS approach to study the genetic differentiation of wild boars and domestic pigs in Poland. The aim of the research is to identify the DNA regions that underwent strong selection during the domestication of the pig and to give an insight into the genetic diversity of the Polish wild boar and domestic pigs. The study also focuses on the identification of admixed individuals in both subspecies.

Materials and Methods

Material

The material for the study was genomic DNA extracted with the use of the Sherlock AX Kit for DNA purification (A&A Biotechnology, Poland). The DNA was extracted from the ear tissues of 20 wild boars (WB) from the South region of Poland, as well as from the blood samples of unrelated and randomly selected pigs of the following breeds: Polish Landrace (PBZ, n=2), Large White Polish (WBP, n=4), Duroc (DUR, n=1), Puławska (PUL, n=3) and Pietrain (n=2). The wild boars were legally hunted in accordance with the national regulations during the 2009-2011 hunting seasons. The sample collection was performed by licensed hunters as a part of routine wildlife management measures. Blood samples from the pigs were taken by the representatives of the Polish Pig Breeders Association, “POLSUS”, and were collected in accordance with the routine parentage controls conducted at the National Research Institute of Animal Production in Balice, Poland; therefore, no permission from the local ethics commission was needed.

Methods

Library preparation and sequencing

We prepared the GBS libraries according to the previously described protocol (Elshire et al. 2011) with minor modifications. In short, the PstI endonuclease was used for the DNA digestion and ligation of 48 indexed adapters designed with the use of GBS Barcode Generator software. Before the library preparation process, the ligation-ready double-stranded adapters were prepared, i.e. the oligonucleotides (forward and reverse complements) were syntheisised with unique barcodes (4-9 nt long) and annealed with a common adapter. The laboratory workflow covered the following steps: overnight incubation of the genomic DNA (200 ng in total) with a tenfold PstI-HF enzyme excess (100,000 U/ml, New England Biolabs, USA) at 37°C; digestion and ligation of the products at 22°C for 60 min with 4.8 ng of each adapter, using the T4 DNA ligase (400 U/µl; New England Biolabs, USA); clean-up step; amplification of the libraries by PCR in 18 cycles, using universal primers (12.5 pmol/µl each) and 2x Taq Master Mix (New England Biolabs, USA); purification of the amplified libraries; quality assessment using Agilent’s TapeStation2200 system (Agilent Technologies, USA) and quantification using a DNA binding dye (Qubit DS DNA assay; Thermo Fisher Scientific, USA). The oligonucleotide sequences used here were the same as the ones used in our previous study (Gurgul et al. 2019).

The library pools were ultimately sequenced in a single-end 50-bp run on the HiScanSQ system (Illumina, USA) using the TruSeq v3 HiSeq flowcell and v3 sequencing reagents (Illumina, USA). The raw reads were deposited in the NCBI Sequence Read Archive (SRA) database under the BioProject Number ID PRJNA658252.

SNPs calling and filtering

The quality of the generated sequencing reads was assessed with the use of FastQC software. The subsequent analysis of the reads was done using the TASSEL 5 GBS v2 Pipeline (Glaubitz et al. 2014) with the default parameters, except maximum kmer Length that was set to 40 and the minimum count of reads for a tag to be output that was set to 5. The pig Sscrofa11.1 genome assembly was used as a reference for the TASSEL pipeline including the read tags matching with a Bowtie2 aligner (Langmead & Salzberg 2012). Polymorphisms resulting from the default TASSEL GBS v2 production pipeline were initially filtered as following: indels and multiallelic markers were removed; individual genotypes were removed if the obtained coverage was low (< 5 reads); and finally, SNPs located on the X and Y chromosomes were removed due to a lack of knowledge of the wild boars’ sexes.

We also removed from the analysis markers with a minor allele frequency (MAF) lower than 0.01, missing genotypes in more than 40% of the
which allowed us to account for a random locus-by-locus variation in the allele frequency. Subsequently, the FST-windows falling into the top 1% of the highest FST observations were analysed in detail for their gene content, in order to identify the genes that were strongly affected by the wild boar domestication and the artificial selection of pigs. The genes within regions with the highest FST were identified using the UCSC Genome Browser (Haeussler et al. 2019) and were analysed in terms of their functions using Kobas3.0 (Xie et al. 2011). The overrepresentation tests in the separate GO, KEEG and Reactome pathways categories were conducted with respect to all known pig genes using Fisher’s exact test and the FDR (Benjamini & Hochberg 1995) correction for multiple testing. Some annotation analyses were completed using the PANTHER Classification System (Mi et al. 2013).

Results

Sequencing read statistics

The sequencing of the libraries with the HiScanSQ System generated a total of 84.3 million (M) raw reads for all the animals. Of the reads, 36,285,567 (1.3 M on average, per sample) had a proper barcode and PstI cut site. These reads were collapsed into 463,756 different sequence tags matching the reference genome. Finally, an average of about 1.1 M reads was mapped to the identified sequence tags and were used for SNP calling in the separate samples (Table 1).

SNPs panel characteristics

The initial TASSEL GBS pipeline allowed for the identification of 29,647 variants common for the domestic pigs and wild boars. Filtering of the markers for the average read depth and variants character (keeping only the biallelic SNPs) resulted in 7,785 SNPs, which were additionally evaluated
between the studied populations. The MDS allowed for the identification of clearly separated populations of domestic pigs and wild boars, with a visibly lower genetic variation present within the wild boar cluster. Among the domestic pigs, a visible sub-clustering was observed that largely corresponded to the included breeds (Figure 1). Similar observations were made when the maximum-likelihood method was applied to create a phylogenetic tree. This analysis clearly separated the two analysed populations (Figure 2). To further validate this result, the Structure software was used to fully reveals the pattern of individual admixtures. For both the groups studied, we verified our hypothesis that more than two parental populations existed (at least one for the domestic representatives and a putative two parental lines, i.e. Asian and European, for the wild boar group). Therefore, we tested different K values ranging from one to six. However, the most appropriate K=2 was established using the Delta K parameter (Suppl. Mat. 1 – Fig. 1). This analysis showed a clear subdivision of the analysed animals into two subpopulations corresponding to the domestic pig and the wild boar populations (Figure 3).

Genetic differentiation of the pig and the wild boar captured using GBS

The final filtered panel of SNPs was used to analyse and visualise the genetic differentiation between the studied populations. The MDS allowed for the identification of clearly separated populations of domestic pigs and wild boars, with a visibly lower genetic variation present within the wild boar cluster. Among the domestic pigs, a visible sub-clustering was observed that largely corresponded to the included breeds (Figure 1). Similar observations were made when the maximum-likelihood method was applied to create a phylogenetic tree. This analysis clearly separated the two analysed populations (Figure 2). To further validate this result, the Structure software was used to fully reveal the pattern of individual admixtures. For both the groups studied, we verified our hypothesis that more than two parental populations existed (at least one for the domestic representatives and a putative two parental lines, i.e. Asian and European, for the wild boar group). Therefore, we tested different K values ranging from one to six. However, the most appropriate K=2 was established using the Delta K parameter (Suppl. Mat. 1 – Fig. 1). This analysis showed a clear subdivision of the analysed animals into two subpopulations corresponding to the domestic pig and the wild boar populations (Figure 3).

Table 2
SNP panel statistics

<table>
<thead>
<tr>
<th>CHR</th>
<th>SNPs</th>
<th># of shared SNPs</th>
<th>Distance between All Pig Wild boar Pig/Wild boar difference* SNP panel statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>566</td>
<td>426</td>
<td>483,995</td>
</tr>
<tr>
<td>2</td>
<td>383</td>
<td>273</td>
<td>396,869</td>
</tr>
<tr>
<td>3</td>
<td>487</td>
<td>333</td>
<td>272,844</td>
</tr>
<tr>
<td>4</td>
<td>390</td>
<td>284</td>
<td>333,230</td>
</tr>
<tr>
<td>5</td>
<td>348</td>
<td>261</td>
<td>290,874</td>
</tr>
<tr>
<td>6</td>
<td>587</td>
<td>399</td>
<td>209,683</td>
</tr>
<tr>
<td>7</td>
<td>379</td>
<td>282</td>
<td>302,387</td>
</tr>
<tr>
<td>8</td>
<td>377</td>
<td>266</td>
<td>362,862</td>
</tr>
<tr>
<td>9</td>
<td>442</td>
<td>319</td>
<td>315,460</td>
</tr>
<tr>
<td>10</td>
<td>334</td>
<td>259</td>
<td>204,321</td>
</tr>
<tr>
<td>11</td>
<td>258</td>
<td>205</td>
<td>305,994</td>
</tr>
<tr>
<td>12</td>
<td>287</td>
<td>186</td>
<td>213,221</td>
</tr>
<tr>
<td>13</td>
<td>441</td>
<td>314</td>
<td>473,241</td>
</tr>
<tr>
<td>14</td>
<td>510</td>
<td>358</td>
<td>278,147</td>
</tr>
<tr>
<td>15</td>
<td>386</td>
<td>283</td>
<td>364,286</td>
</tr>
<tr>
<td>16</td>
<td>226</td>
<td>156</td>
<td>353,030</td>
</tr>
<tr>
<td>17</td>
<td>307</td>
<td>218</td>
<td>205,745</td>
</tr>
<tr>
<td>18</td>
<td>203</td>
<td>160</td>
<td>272,869</td>
</tr>
<tr>
<td>Cotgs</td>
<td>387</td>
<td>289</td>
<td>NA</td>
</tr>
<tr>
<td>All</td>
<td>7298</td>
<td>5243</td>
<td>326,918</td>
</tr>
</tbody>
</table>

CHR – chromosome; MAF – minor allele frequency; Ho – observed heterozygosity; He – expected heterozygosity; * – negative values point to the parameters which are lower in pigs than in wild boar; SNPs – polymorphic in both pig and wild boar populations; Fixed SNPs – monomorphic in at least one population.
Fig. 1. Multidimensional scaling for the analysed individuals based on the IBS distance matrix. PBZ – Polish Landrace; WBP – Large White Polish; DUR – Duroc; PUL – Puławska; PIETR – Pietrain.

Fig. 2 Maximum-likelihood phylogenetic trees constructed with MEGA-X software. A – The evolutionary history was inferred by using the Maximum Likelihood method and the Jukes-Cantor model (Jukes & Cantor 1969). The tree with the highest log likelihood (-65198.14) is shown. The percentage of trees in which the associated taxa are clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying the Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with the superior log likelihood value. The tree is drawn to scale, with the branch lengths measured according to the number of substitutions per site. B – The bootstrap consensus tree inferred from 1000 replicates is shown to represent the evolutionary history of the taxa that were analysed (Felsenstein 1985). The branches corresponding to partitions reproduced in less than 50% of the bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa were clustered together in the bootstrap test (1000 replicates) are shown next to the branches (Felsenstein 1985). Initial tree(s) for the heuristic search were obtained automatically by applying the Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with the superior log likelihood value. PBZ – Polish Landrace; WBP – Large White Polish; DUR – Duroc; PUL – Puławska; PIETR – Pietrain.
The overall global genetic differentiation of the analysed population captured by the GBS SNPs was relatively low, with mean and weighted FST values of 0.099 and 0.118, respectively, meaning that only about 10% of the global genetic variation can be attributed to the differentiation between the populations.

The overlapping top 5-SNP windows were merged and the covered genome regions were analysed in detail. This approach allowed for the detection of the genome regions with fixed genetic differences between the pig and wild boar (Suppl. Mat. 1 – Fig. 2). The regions were distributed on 17 different autosomes, with the highest number located on SSC1, SSC6, SSC9 and SSC17 (3 regions each). The size of the regions ranged from 210 kb to 4.6 Mb, with a mean of 1.7 Mb. In the overlapped regions with 50 different RefSeq genes, 48 genes were annotated (Suppl. Mat. 2). Together, the genes enriched GO terms (FDR<0.05) associated inter alia with DNA-binding transcription activator activity, RNA polymerase II-specific and the regulation of oogenesis; KEEGps was associated with the Wnt signalling pathway, arginine and proline metabolism and pyrimidine metabolism; and the Reactome pathways connected with, e.g., interconversion of nucleotide di- and triphosphates, metabolism of nucleotides and amino acids, VEGFR2 mediated cell proliferation and many other pathways (Table 3; Suppl. Mat. 2). In addition, the gene functions were supported by an extensive literature search.

**Fig. 3.** Patterns of admixture among the studied populations revealed using Structure software. 1 – pig population; 2 – wild boar population; K – number of expected populations; K calculated from 2 to 6. The analysis was of locus-specific genetic differentiation.

**Table 3**

Top enriched GO teams for genes found in the genome regions the most divergent between wild boar and pig

<table>
<thead>
<tr>
<th>GO term</th>
<th>GO ID</th>
<th>FDR</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>axonal transport of mitochondrion</td>
<td>GO:0019896</td>
<td>0.016194556</td>
<td>NEFL, UCHL1</td>
</tr>
<tr>
<td>UTP biosynthetic process</td>
<td>GO:0006228</td>
<td>0.016194556</td>
<td>NME1, NME2</td>
</tr>
<tr>
<td>GTP biosynthetic process</td>
<td>GO:0006183</td>
<td>0.016194556</td>
<td>NME1, NME2</td>
</tr>
<tr>
<td>CTP biosynthetic process</td>
<td>GO:0006241</td>
<td>0.016194556</td>
<td>NME1, NME2</td>
</tr>
<tr>
<td>endomembrane system</td>
<td>GO:0012505</td>
<td>0.016194556</td>
<td>RAB22A, PGRMC2, STX16</td>
</tr>
<tr>
<td>DNA-binding transcription activator activity, RNA polymerase II-specific</td>
<td>GO:0001228</td>
<td>0.016194556</td>
<td>WT1, PAX6, MYOG, NOBOX, MEOX2</td>
</tr>
<tr>
<td>nucleoside diphosphate phosphorylation</td>
<td>GO:0006165</td>
<td>0.016194556</td>
<td>NME1, NME2</td>
</tr>
<tr>
<td>positive regulation of transcription by RNA polymerase II</td>
<td>GO:0045944</td>
<td>0.01881921</td>
<td>WT1, NOBOX, NME2, AHR, PAX6, MAVS, MYOG</td>
</tr>
<tr>
<td>nucleoside diphosphate kinase activity</td>
<td>GO:004550</td>
<td>0.01881921</td>
<td>NME1, NME2</td>
</tr>
<tr>
<td>positive regulation of protein kinase activity</td>
<td>GO:0045860</td>
<td>0.023408697</td>
<td>CDC25B, MMD</td>
</tr>
<tr>
<td>negative regulation of microtubule depolymerization</td>
<td>GO:0007026</td>
<td>0.023579161</td>
<td>SPEF1, APC</td>
</tr>
<tr>
<td>negative regulation of viral genome replication</td>
<td>GO:0045071</td>
<td>0.032865802</td>
<td>MAVS, ISG15</td>
</tr>
<tr>
<td>axon cytoplasm</td>
<td>GO:1904115</td>
<td>0.040083018</td>
<td>NEFL, UCHL1</td>
</tr>
<tr>
<td>cytoplasm</td>
<td>GO:0005737</td>
<td>0.042170286</td>
<td>SPEF1, CDH11, NME2, WT1, CDC25B, NPE-PL1, ISG15, APC, UCHL1, MEOX2, BZW2</td>
</tr>
</tbody>
</table>
Discussion

We used the GBS approach for the first time, with an aim to differentiate between closely related subspecies. GBS is a technique based on the random digestion of genomic DNA with one frequently-cutting restriction enzyme (REs) and ligating the cut site-specific adapters, including unique indexes following a common primer binding site. The selection of appropriate REs is a pivotal step in the GBS lab protocol, but it also generates the highest financial cost. The PstI enzyme was selected based on our previous experiences of farm animals as producing sufficient fragments for a genome-wide analysis (Gurgul et al. 2019). GBS takes advantage of a simplified hands-on lab protocol, which was achieved mostly by reducing the sequencing library preparation steps and simplifying the procedure for the generation of restriction fragments with barcoded adapters (Elshire et al. 2011). Finally, the time and cost were diminished by reducing the DNA purification steps and omitting DNA size selection (Gurgul et al. 2019).

While considering the GBS data analysis, the identification of the polymorphism and sample genotyping does not require a reference genome. This gives the GBS technique a broader purpose than more commonly used microarrays.

Genetic diversity and identification of hybrids and interspecies admixtures

The European and Asian wild boar populations diverged around 1 million years ago (mya), resulting in very different allele frequencies at millions of genomic loci and over a million loci representing alternative alleles (Groenen et al. 2012). While wild boars are the ancestors of modern domestic pigs, it is believed that the domestication process occurred separately for the species in Asia and Europe. Studies of the genetic diversity of wild boars and pigs can give insights into the phylogenetic background of the subspecies.

In our study, we found a low level of genetic diversity in the Polish wild boar population compared to the domestic pig population. Domestic pigs appear more divergent, as they originated from a highly diverse range of wild ancestors and were bred for several traits resulting in a large number of phenotypically diverse breeds (Yang et al. 2017). All the methods (MDS, Structure) we implemented to study the genetic structure of both subspecies proved that the analysed wild boars are not admixed with the gene pool of the domestic pigs. In the domestic pig group, two samples (one from the Duroc breed and one Landrace) had a negligible proportion of the wild boar genome. This observation is also clearly demonstrated in the phylogenetic tree. Our studies did not coincide with that research on the European population that reported the frequent hybridisation of these two subspecies (Fontanesi et al. 2014; Iacolina et al. 2018; Scandura et al. 2008). There are farms (e.g. in Bulgaria and Sardinia) where pigs are kept in semi-wild conditions, and it is highly likely that in these cases the domestic pigs sporadically crossbred with wild boars (Scandura et al. 2008). The frequent hybridisation of pigs and wild boar was found in studies on the European population (Scandura et al. 2008; Fontanesi et al. 2014; Iacolina et al. 2018), while studies of the MC1R gene in the Polish population of wild boars also proved the existence of admixed genotypes (Dzialuk et al. 2018; Babicz et al. 2013). However, these results are probably due to the limited number of samples representing wild boars, which were restricted to one region of Poland. The breeding of Polish pigs also prevents crossing between domestic pigs and wild boars.

Genetic differentiation of the pig and the wild boar

In the present study, we identified genome regions that were differentially selected between the pig and the wild boar. Within these regions, 48 different genes were identified. Three databases – the KEGG pathway, Reactome and GO terms – were further used to assign a functional significance to the identified genes determined to have the greatest differences between the two groups. A detailed analysis of the enriched GO terms allowed for the identification of genes that are associated with the axonal transport of mitochondria (NEFL, UCHL1), DNA-binding transcription activator activity, RNA polymerase II-specific (WT1, PAX6, MYOG, NOBOX, MEOX2) and the positive regulation of transcription by RNA polymerase II (WT1, NOBOX, NME2, AHR, PAX6, MAYS, MYOG). The KEGG pathway and Reactome grouped the selected genes into biological processes: interconversion of nucleotide di- and triphosphates, myogenesis, RIG-I-like receptor signalling pathway, extra-nuclear oestrogen signalling and muscle contraction. These are processes that are presumably under selection during domestication and animal breeding.

Furthermore, we found several genes that were potentially under selection in early domestication, as well as in the later artificial selection of pigs. These are associated with inter alia muscle development (MYOG, MEOX2), pre-weaning mortality stress (MYO7A), immune response (CSTF3, APC) and sensory perception (TAS1R3).
The breeding selection of pigs was aimed at increasing the meat production efficiency through both changes in the bone and muscle development processes and changes in the metabolism that affects the growth rate (Amaral et al. 2011). We found the MYOG and MEOX2 genes in divergent gene regions, which influence muscle development (Otto et al. 2010; Stupka et al. 2012; Verner et al. 2007). MYOG belongs to the MYOD family of genes, which determines the lean meat content capacity in pork (Verner et al. 2007). The MYO7A gene has been thoroughly studied in pigs, owing to its potential effect on different production and meat quality traits (Roman et al. 2012; Verner et al. 2007). The meat quality is positively correlated with the lean meat content, which relies on the muscle fibres. The key moment for developing muscle fibres is the period of embryogenesis (Verner et al. 2007). The MEOX2 gene belongs to the mesoderm/mesenchyme homebox family of genes with a pivotal role in the development and patterning of somites (Mankoo et al. 2003). In mice, it was found that a genetic deletion of MEOX2 led to a serious reduction of the adult muscle size, as well as a modification of the myofibre composition, muscle architecture and physiological features (Otto et al. 2010).

Parameters related to reproduction affect the efficiency of the slaughter production. Derks et al. (2021) studied a stop-gain variant (g.11280403C>T) in the MYO7A gene. They proved a significant 23% decline in the pre-weaning survival while analysing 31 past carrier-by-carrier litters. In humans, a variant of the MYO7A gene is associated with Usher syndrome 1 (deaf-blindness) (Liu et al. 1997). MYO7A knocked-out mice suffered from various neurological diseases and reduced male fertility (Williams 2008).

To date, comparative studies the genomes of both pigs and wild boars have identified different genes related to the immune system, host defence and sensory perception (Amaral et al. 2011; Groenen et al. 2012). The genes associated with the immune system and olfactory receptors belong to the most rapidly evolving regions of the genome (Amaral et al. 2011). The immune response mechanism is always faced with strong challenges caused by changing environments and pathogens (Iacolina et al. 2016). Among the differentiating pigs and wild boar genes, we found several genes related to the immune response. These genes code the receptors for interleukins (e.g. CSTF3) or play a significant role in the immune response to viruses (e.g. APC). (Groenen et al. 2012) pointed out that among the porcine evolutionary breakpoint regions (EBPs) is a group related to the sensory perception of taste in pigs (mostly related to salty, umami and sweet tastes). Among the genes coding taste receptors, we found the one related to the sweet taste (TAS1R3) in the divergent gene regions.

Despite a limited sample number, our results proved that the GBS as a whole-genome screening technique is a useful and valuable tool for population genetic studies and for the differentiation of two closely-related species and their possible hybrids. A limited number of SNPs (5K SNPs) was sufficient to grasp the genetic distance between pigs and wild boars. However, the number of SNPs might be improved in further experiments regarding this subject. We revealed novel genes with different patterns of variation in domestic pigs and wild boars. The genes differentiating both subspecies were related to both breeding selection (fertility and meat efficiency) and to biological processes associated with adaptation to the altered environment (host defence and sensory perception.). The polymorphism of these genes has the potential to differentiate wild boars from domestic pigs, but further studies with more samples representing both subspecies and their hybrids are needed.

Author Contributions

Research concept and design: A.K.; Collection and/or assembly of data: G.S., A.R.; Data analysis and interpretation: A.K., G.S., A.G., T.S., M.O.; Writing the article: A.K., M.O.; Critical revision of the article: G.S., A.G., T.S., M.O., A.R.; Final approval of article: A.K.

Conflict of Interest

The authors declare no conflict of interest.

Supplementary Material

Supplementary Material to this article can be found online at:

Supplementary files:
Suppl. Mat. 1 – Fig. 1. Delta K values for different numbers of populations assumed in the Structure analysis.
Suppl. Mat. 1 – Fig. 2. Sliding-window averaged FST values between pigs and wild boars.
Suppl. Mat. 2. Data for genetic differentiation analysis.
References
Fajardo V., González I., Martín I., Rojas M., Hernández P.E., García T., Martín R. 2008. Differentiation of European wild boar (Sus scrofa scrofa) and domestic swine (Sus scrofa domestica) meats by PCR analysis targeting the mitochondrial D-loop and the nuclear melanocortin receptor 1 (MC1R) genes. Meat Sci. 78: 314-322. https://doi.org/10.1016/j.meatsci.2007.06.018
Pig and wild boar identification using genotyping-by-sequencing technology


Stamatis C., Mamuris Z. 2010. Detection of hybrids between Wild Boars (Sus scrofa) and East Balkan Sus scrofa scrofa in the context of phylogenetic trees. Nucleic Acids Res. 38: 4655-4664. https://doi.org/10.1038/s41598-020-02646-4


