

Brief Report

Identification of Breed-Specific SNPs of Danish Large White Pig in Comparison with Four Chinese Local Pig Breed Genomes

Xudong Wu ¹ , Decai Xiang ², Wei Zhang ¹, Yu Ma ³, Guiying Zhao ^{4,*} and Zongjun Yin ^{3,*}

¹ Anhui Provincial Key Laboratory of Livestock and Poultry Product Safety Engineering, Institute of Animal Husbandry and Veterinary Medicine, Anhui Academy of Agricultural Sciences, Hefei 230001, China

² Yunnan Academy of Animal Husbandry and Veterinary Sciences, Kunming 650224, China

³ Anhui Provincial Laboratory of Local Animal Genetic Resource Conservation and Bio-Breeding, College of Animal Science and Technology, Anhui Agricultural University, Hefei 230036, China

⁴ College of Animal Science and Technology, Yunnan Agricultural University, Kunming 650201, China

* Correspondence: zhaoguiying2006@163.com (G.Z.); yinzongjun@ahau.edu.cn (Z.Y.)

Abstract: Genetic variation facilitates the evolution, environmental adaptability, and biodiversity of organisms. Danish Large White (LW) pigs have more desirable phenotypes compared with local Chinese pigs, which have difficulty adapting to the modern swine industry. However, the genome-wide mutational differences between these pig breeds are yet to be evaluated. Therefore, this study aimed to evaluate genomic variation and identify breed-specific SNPs in Danish LW pigs. Here, 43 LW, 15 Diqing Tibetan (DQZ), and 15 Diannan small-ear (DN) pigs whose genomes were resequenced with 5× depth were selected. This was followed by a conjoined analysis of our previous resequencing data of 24 Anqing six-end white (AQ) and six Asian wild (SS) pigs. In total, 39,158,378 SNPs and 13,143,989 insertion–deletions were obtained in all breeds. The variation number of LW pigs was the lowest, with 287,194 breed-specific and 1289 non-synonymous SNPs compared with Chinese breeds. Functional analysis of the breed-specific non-synonymous SNPs indicated that these mutations were mainly associated with the reproductive performance, feed intake, and feed conversion ratio of LW pigs. These findings provide a theoretical basis for genetic improvements in the Chinese swine industry.

Keywords: Large White pig; genome; breed-specific; functional analysis



Citation: Wu, X.; Xiang, D.; Zhang, W.; Ma, Y.; Zhao, G.; Yin, Z.

Identification of Breed-Specific SNPs of Danish Large White Pig in Comparison with Four Chinese Local Pig Breed Genomes. *Genes* **2024**, *15*, 623. <https://doi.org/10.3390/genes15050623>

Academic Editor: Miloš Macholán

Received: 11 April 2024

Revised: 3 May 2024

Accepted: 10 May 2024

Published: 14 May 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Genetic variation changes the genetic information of organisms. This affects phenotypic traits, which are not only the genetic basis of evolution and environmental adaptation but also the main source of biodiversity [1]. The rapid development of next-generation sequencing in recent years has provided an important technology for acquiring extensive genetic information. Whole-genome resequencing has been widely used in swine genomic research. For example, it is used for mutation site comparisons and analyses, such as SNP, insertion–deletion (InDel), and structural variation, based on the generation sequencing platform to re-sequence representative individuals of different groups within a species containing a known genome [2]. Understanding genetic variability and its influence on economically important traits is critical for pig breeding and production. High-quality, richly annotated reference genome sequences are key resources. They provide an important framework for discovering and analyzing genetic variation and linking genotypes to function. The new assembly of a Duroc pig (*Sus scrofa* 11.1) whole-genome sequence has provided a good reference [3].

Son et al. (2020) re-sequenced 20 micro-pigs to detect 144,507 non-synonymous SNPs, revealing that *RHEB* and *FRAS1* may affect traits of interest in these pigs [4]. Whole-genome re-sequencing data revealed that the number of genomic variants was similar in Danish Duroc, Landrace, and Large White pigs (LW), with almost half of the variants

segregated in all three breeds. Moreover, LW pigs had 3,883,826 variants that are not found in Duroc or Landrace pigs [5]. The European and Asian wild boar populations diverged approximately 1 mya, and different selective directions resulted in millions of different minor allele frequencies of genomic locations [6]. As one of the most famous commercial lean pig breeds, Large White pigs (LW) were originally bred in Yorkshire, Britain, and then widely distributed in European and American countries. They have been selectively bred and improved according to different directions, resulting in various characteristic strains such as the Danish LW, American LW, Canadian LW, and French LW. There are significant phenotypic differences between LW and Chinese native pig breeds, with several hundred years of history of evolution and breeding. Compared with traditional breeding methods, marker-assisted selection can accurately and quickly analyze the genetic composition of individuals from the molecular level to realize the selection of genotypes with more information and higher selection accuracy. In animal breeding, the use of marker genotypes can accurately estimate the breeding value of quantitative traits. Danish LW pigs have a higher feed conversion ratio, growth rate, and lean meat percentage than local Chinese pigs. Thus, the molecular markers of their germplasm characteristics warrant further study to enable genetic improvement and hybrid breeding in the Chinese swine industry. This study aimed to investigate specific SNPs and their functions in LW pigs using Chinese local pig genome data as a control.

2. Materials and Methods

In this study, all animal experimental procedures were performed according to the Guide for the Care and Use of Laboratory Animals (Ministry of Science and Technology of China, 2006). Genetic variation information of 103 pigs was detected. Of these, 43 ear tissues of LW pigs were collected from Huaibei City, Anhui province, People's Republic of China. Fifteen blood samples each of Diqing Tibetan (DQT) and Diannan small-ear (DN) pigs were collected from the Diqing Tibetan Autonomous Prefecture and Xishuangbanna Dai Autonomous Prefecture, Yunnan province, China, respectively, and stored in a 2 mL anticoagulation tube at -20°C . Genomic DNA was extracted from the ear and blood samples using the standard phenol–chloroform method [7]. The quality and concentration of the DNA were assessed using a 1% agarose gel (run for >25 min at 120 V) and a NanoDrop spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). Genomic DNA was randomly fragmented, and the fragmented genomic DNA was selected to a certain average size. The fragments were subjected to end-repair and then were 3' adenylated. Then, adaptors were ligated to the ends of these 3' adenylated fragments. After adapter ligation and DNA cluster preparation, the fragments were sequenced on the DNB-SEQ platform with $5\times$ depth (Beijing Genomics Institution, Wuhan, China). The whole-genome data of 30 pigs, including 24 Anqing six-end white (AQ) and 6 Asian *Sus scrofa* (SS) pigs, were obtained from our previous work [8].

The comparison software bwa 0.7.17 (<http://bio-bwa.sourceforge.net/> (accessed on 28 December 2023)) was used to compare the data to the reference genome (*Sus scrofa* 11.1) and the comparison samtools software 1.17 (<https://github.com/samtools/samtools> (accessed on 28 December 2023)) was used to sort the results and mark PCR duplications. We used the GATK software 4.4.0 (<https://gatk.broadinstitute.org> (accessed on 28 December 2023)) to find SNP and Indel sites. The detected variation was filtered using VariantFiltration, and quality control standards were set as $-\text{Window } 4, -\text{filter } \text{"QD} < 4.0 \text{ || FS} > 60.0 \text{ || MQ} < 40.0\text{"}$, $-\text{G_filter } \text{"GQ} < 20\text{"}$ (QD:Variant Confidence/Quality by Depth; FS:Phred-scaled p -value using Fisher's exact test to detect strand bias; MQ:RMS Mapping Quality; GQ:Genotype Quality). Then, ANNOVAR software 2020-06-07 (<https://annovar.openbioinformatics.org/> (accessed on 28 December 2023)) was used to annotate the detected SNP, Indel, and other genomic variations with external databases.

Breed-specific SNPs were detected as follows in LW pigs, which were compared with four local Chinese pig breed genomes: (i) SNP only in the LW pig population, (ii) SNP only in autosomes, and (iii) SNP allele frequency > 0.1 . The encompassing gene

was annotated using the annotation database by NCBI (<https://www.ncbi.nlm.nih.gov> (accessed on 23 November 2023)), and to further analyze the functions of the identified genes, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses were performed. Additionally, an extensive literature review was performed to gather pertinent information on gene functions for exploratory investigations.

3. Results

In this study, 103 pigs from 5 pig populations were selected for whole-genome re-search, and approximately 10 billion reads (1406 Gb) were generated. Average data sets consisted of 13.65 Gb (5.46 X) each, and reads were aligned to the pig reference genome with an average alignment rate of $98.09 \pm 0.02\%$ and coverage ($1\times$) with an average rate of $95.81 \pm 0.40\%$ (Supplementary Table S1). Of the 39,158,378 SNPs and 13,143,989 InDels detected in all animals after alignment with the reference genome 11.1 and variant calling, 14,914,614 SNPs and 7,898,289 InDels were detected in the LW pig genome, 24,334,002 SNPs and 9,552,558 InDels were detected in the DQZ pig genome, 24,458,193 SNPs and 9,907,047 InDels were detected in the AQ pig genome, 24,283,900 SNPs and 9,548,787 InDels were detected in the DN pig genome, and 24,879,296 SNPs and 8,950,602 InDels were detected in the SS pig genome (Figure 1a,b). Functional annotation results showed that 51.75%, 40.08%, and 5.30% of the SNPs in LW pigs were found in genome intergenic, intronic, and exonic regions, respectively (Figure 1c). Moreover, the ratio of exonic region SNPs to total SNPs in local Chinese pigs was higher than that in LW pigs, and the average transition/transversion (ts/tv) ratio of LW pigs was the lowest among the five pig breeds (Figure 1d).

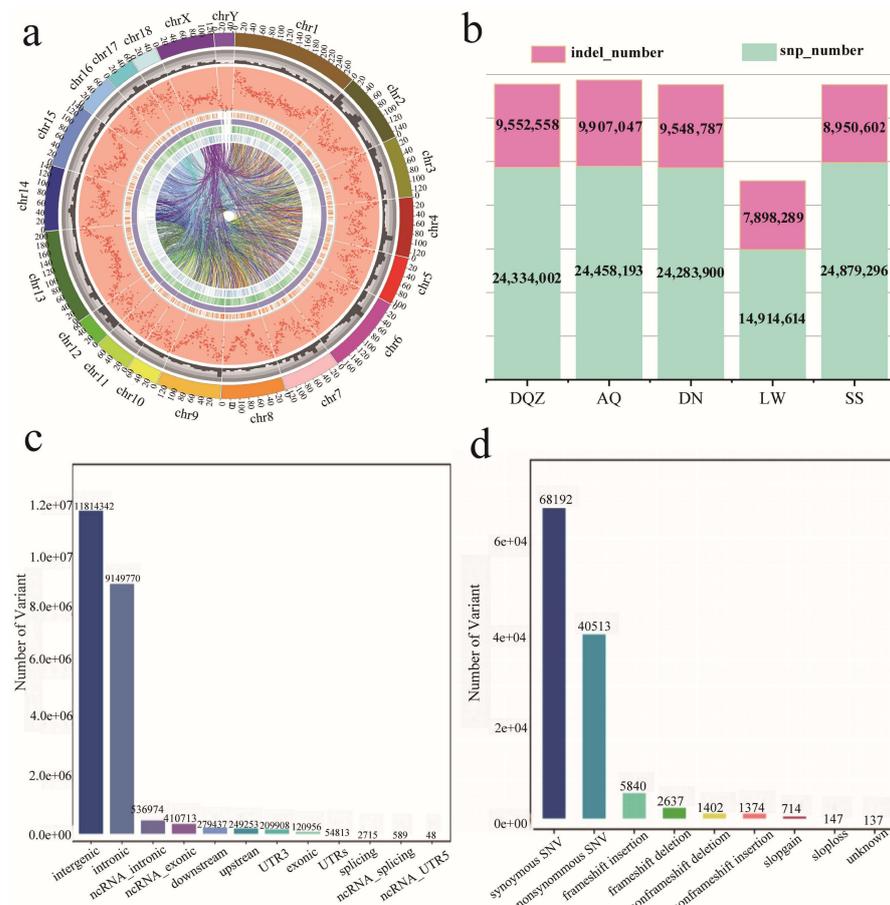


Figure 1. (a) The variation distribution in five pig breed genomes. Rings represent the genomic information of five pig breeds at each level. From the outside: length of each chromosome (in Mb).

The orange, purple, green, and blue lines indicate the locations of tandem duplications (DUPs), large-segment deletions (DELS), insertion-type structural variations (INSSs), and inverted-type structural variations (INVs) on the chromosome, respectively. The line in the inner circle indicates the position of BND (structural variant of the inter-chromosomal translocation type) on both chromosomes. (b) Number of single-nucleotide polymorphism (SNP) and insertion–deletion (InDel) statistics in five pig breed genomes. (c) Number of functional site statistics in the Large White (LW) pig genome. (d) Number of annotated site statistics in the LW pig genome.

We obtained 287,194 breed-specific SNPs in LW pigs, with *Sus scrofa* chromosomes (SSC) 13 and 18 showing the highest and lowest numbers of breed-specific SNPs, respectively (Figure 1a,b). The 2794 SNPs found in the exonic regions included 1486 synonymous, 1289 non-synonymous, 17 stop-gain, and 2 stoploss SNPs. We annotated 1131 genes of the LW pig breed-specific non-synonymous SNPs to further evaluate the functions of the identified genes using GO and KEGG analyses (Supplementary Table S2). These genes included 7575 GO items that were mainly enriched in the cytoskeleton, basement membrane, extracellular matrix, and external encapsulating structure terms (Figure 2c). KEGG pathway analysis revealed that the identified genes were enriched in protein digestion and absorption, ECM–receptor interaction, peroxisomes, adherens junction, and taste transduction pathways (Figure 2d). An extensive literature search revealed that *COL5A1*, *CSF1R*, *CEBPA*, *CYP4B1*, and *NQO1* were associated with back fat thickness and fat deposition; *COL27A1*, *PDZRN4*, *LTBP2*, *TBC1D1*, *CCSER1*, *KDM6B*, *GUCY2D*, *CPS1*, and *GCK* were associated with growth, body size, and carcass traits; *SPAG6*, *NUAK1*, *ZDHHC18*, and *ADAT1* were associated with reproductive performance; and *RAC1*, *AQP4*, *CYP2J2*, and *CDH19* were associated with feed intake and feed conversion ratio.

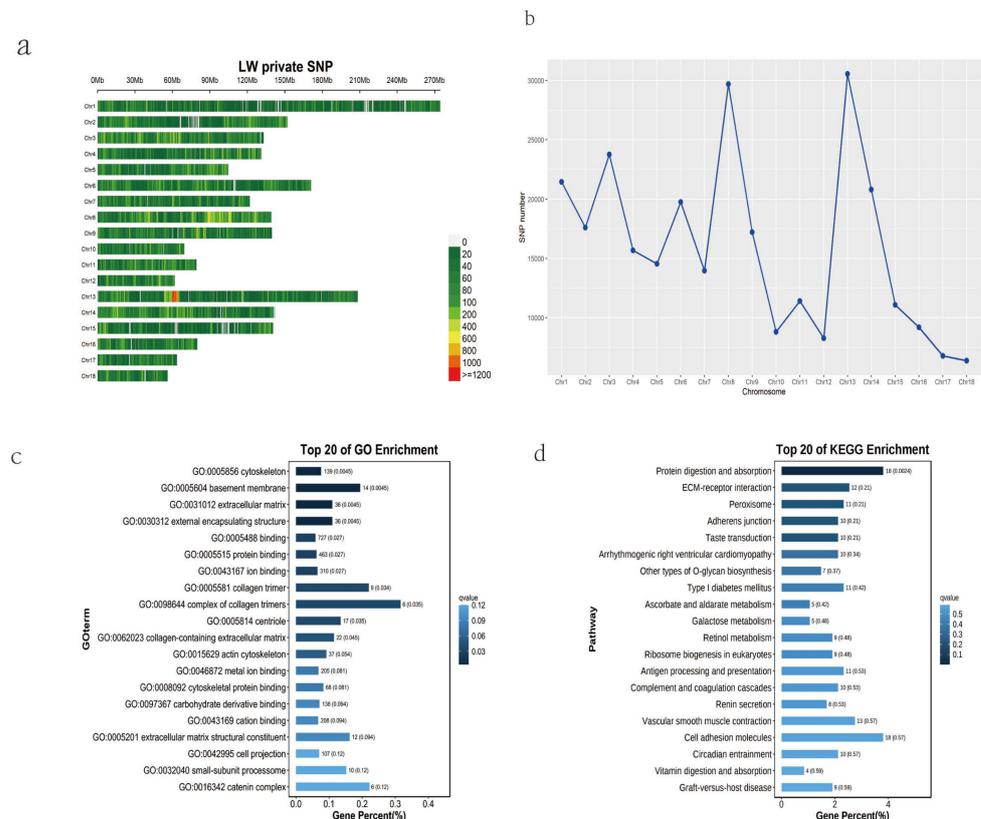


Figure 2. Breed-specific single-nucleotide polymorphisms (SNPs) and their functional analysis in LW pigs. (a) Distribution of breed-specific SNPs in the LW autosome. (b) Number statistics of breed-specific SNPs in the LW autosome. (c) Gene Ontology (GO) analysis of breed-specific non-synonymous SNPs. (d) Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis of breed-specific non-synonymous SNPs.

4. Discussion

Pigs are important domestic animals in China. The national pork production was 55.41 million tons in 2022, increasing by 4.6% year on year (<https://www.stats.gov.cn/> (accessed on 28 December 2023)). In addition, pork consumption has always occupied a large proportion of the daily diet. Local pig breeds are abundant in China, with 83 native breeds (<https://www.moa.gov.cn> (accessed on 28 December 2023)) that exhibit good meat quality, reproductive performance, and disease resistance, providing valuable genetic resources. However, most local breeds have difficulty adapting to the modern swine industry because of their body size, lean percentages, and growth rates. Duroc, Landrace, and LW pigs are the main commercial breeds on large-scale pig farms in China. Therefore, the molecular marker identification of key traits and crossbreeding between lean and local breeds are important directions for pig production.

The genome is a complete collection of DNA, and whole-genome re-sequencing enables comprehensive genome-wide studies that elucidate all classes of genome variation across the allele frequency spectrum [9]. Many studies have revealed the breed characteristics of LW pigs using performance differences, the transcriptome, selective sweeps, and ROH analysis, identifying many candidate genes [10–12]. However, a genome-wide comparison of mutational differences between LW pigs and local Chinese pigs has not been performed. Major genomic and phenotypic differences have been observed between Asian and European pig breeds. However, the current reference genome belongs to the European Duroc pig, which is a lean-type breed similar to the LW pig. This explains why we obtained fewer variation loci in LW pigs than in local Chinese pigs. In the present study, most variations were in the intergenic and intronic regions of the genomes of all five pig breeds, which is consistent with previous reports. Breed-specific SNPs can elucidate the genomic landscape related to the unique characteristics of a breed compared with other breeds, which implies that they may be excellent genetic markers [13].

Non-synonymous SNPs in the protein-encoding regions of genes result in single amino acid substitutions in proteins, which causes considerable changes owing to the alteration in protein structure and function [14]. This may represent the genetic basis of characteristic differences in organisms. In the present study, 1131 identified genes possessing LW pig breed-specific non-synonymous SNPs were enriched in pathways related to transport, catabolism, metabolism, and digestion. Feed efficiency has a major ecological impact on pig production. Therefore, we focused on the candidate genes associated with the feed conversion ratio in this study. RAC1 proteins induce actin polymerization, which was upregulated in the high-feed-efficiency group of Duroc × (Landrace × Yorkshire) commercial pigs [15]. Endurance exercise increases *AQP4* expression levels in muscle cell membranes to regulate metabolic needs during physical activity. These expression levels have been linked to feed efficiency traits in GWAS results [16]. Vitamin A metabolism in the liver affects feed efficiency through energy metabolism and the promotion of fatty acid biosynthesis in pigs. *CYP2J2* is the key gene associated with fatty acid biosynthesis, and its expression is significantly upregulated in the livers of high-feed-efficiency pigs [17]. Similarly, *CDH19* plays an important role in the formation and maintenance of the integrity of various tissues. This gene has also been described on SSC1, with functions related to feed efficiency traits [18].

5. Conclusions

This study preliminarily revealed the genomic variation differences between LW and local Chinese pigs. We further described breed-specific genomic SNPs and their functions in LW pigs. Our results revealed 287,194 breed-specific genomic SNPs in LW pigs, which contained 1289 non-synonymous SNPs. Functional analysis of non-synonymous SNPs provides a reference for the genetic basis of LW pig characteristics, such as growth and feed conversion ratio. Finally, our findings may help develop useful molecular marker sites for application in the swine industry.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/genes15050623/s1>, Table S1: The re-sequencing information of 103 samples. Table S2: The annotated genes information of the LW pig breed-specific non-synonymous SNPs.

Author Contributions: Conceptualization, X.W. and D.X.; methodology, X.W.; software, X.W. and W.Z.; validation, X.W., D.X. and W.Z.; formal analysis, X.W.; investigation, Y.M.; resources, D.X.; data curation, X.W.; writing—original draft preparation, X.W.; writing—review and editing, X.W.; visualization, X.W.; supervision, G.Z.; project administration, Z.Y.; funding acquisition, Z.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Major Special Science and Technology Project of Anhui Province, grant numbers 202203a06020031 and 202103a06020013; the National Natural Science Foundation of China, grant number 31972531; the Major Special Science and Technology Project of Yunnan Province, grant number 202302AE090016; and the Yunnan Provincial Science and Technology Talents and Platform Plan, grant number 202305AF150168).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The WGS datasets of LW, DQZ, and DN pigs obtained during the present study are not publicly available because of intellectual property considerations but are available from the corresponding author upon reasonable request. The WGS datasets of the AQ and SS pigs were submitted to the NCBI (National Center for Biotechnology Information, <https://www.ncbi.nlm.nih.gov/>) (accessed on 26 October 2023)) under accession number PRJNA699491.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Agashe, D.; Sane, M.; Singhal, S. Revisiting the role of genetic variation in adaptation. *Am. Nat.* **2023**, *202*, 486–502. [[CrossRef](#)] [[PubMed](#)]
2. Xiong, X.; Liu, J.; Rao, Y. Whole genome resequencing helps study important traits in chickens. *Genes* **2023**, *14*, 1198. [[CrossRef](#)]
3. Warr, A.; Affara, N.; Aken, B.; Beiki, H.; Bickhart, D.M.; Billis, K.; Chow, W.; Eory, L.; Finlayson, H.A.; Flicek, P.; et al. An improved pig reference genome sequence to enable pig genetics and genomics research. *Gigascience* **2020**, *9*, giaa051. [[CrossRef](#)] [[PubMed](#)]
4. Son, D.H.; Hwang, N.H.; Chung, W.H.; Seong, H.S.; Lim, H.; Cho, E.S.; Choi, J.W.; Kang, K.S.; Kim, Y.M. Whole-genome resequencing analysis of 20 Micro-pigs. *Genes Genom.* **2020**, *42*, 263–272. [[CrossRef](#)]
5. Cai, Z.; Sarup, P.; Ostersen, T.; Nielsen, B.; Fredholm, M.; Karlskov-Mortensen, P.; Sørensen, P.; Jensen, J.; Guldbandsen, B.; Lund, M.S.; et al. Genomic diversity revealed by whole-genome sequencing in three Danish commercial pig breeds. *J. Anim. Sci.* **2020**, *98*, skaa229. [[CrossRef](#)]
6. Groenen, M.A.; Archibald, A.L.; Uenishi, H.; Tuggle, C.K.; Takeuchi, Y.; Rothschild, M.F.; Rogel-Gaillard, C.; Park, C.; Milan, D.; Megens, H.J.; et al. Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* **2012**, *491*, 393–398. [[CrossRef](#)] [[PubMed](#)]
7. Sambrook, J.; Russell, D.W. *Molecular Cloning: A Laboratory Manual*, 3rd ed.; Cold Spring Harbor Laboratory Press: New York, NY, USA, 2001.
8. Wu, X.; Zhou, R.; Wang, Y.; Zhang, W.; Zheng, X.; Zhao, G.; Zhang, X.; Yin, Z.; Ding, Y. Genome-wide scan for runs of homozygosity in Asian wild boars and Anqing six-end-white pigs. *Anim. Genet.* **2022**, *53*, 867–871. [[CrossRef](#)] [[PubMed](#)]
9. Lappalainen, T.; Scott, A.J.; Brandt, M.; Hall, I.M. Genomic analysis in the age of human genome sequencing. *Cell* **2019**, *177*, 70–84. [[CrossRef](#)]
10. Seoni, E.; Battacone, G.; Ampuero Kragten, S.; Dohme-Meier, F.; Bee, G. Impact of increasing levels of condensed tannins from sainfoin in the grower-finisher diets of entire male pigs on growth performance, carcass characteristics, and meat quality. *Animal* **2021**, *15*, 100110. [[CrossRef](#)] [[PubMed](#)]
11. Wang, J.; Xiao, J.; Liu, X.; Gao, Y.; Luo, Z.; Gu, X.; Zhang, J.; Wu, D.; Geng, F. Tandem mass tag-labeled quantitative proteomic analysis of tenderloins between Tibetan and Yorkshire pigs. *Meat Sci.* **2021**, *172*, 108343. [[CrossRef](#)]
12. Zhang, D.; Wu, W.; Huang, X.; Xu, K.; Zheng, C.; Zhang, J. Comparative analysis of gene expression profiles in differentiated subcutaneous adipocytes between Jiaying Black and Large White pigs. *BMC Genom.* **2021**, *22*, 61. [[CrossRef](#)]
13. Lee, Y.S.; Son, S.; Lee, H.K.; Lee, R.H.; Shin, D. Elucidating breed-specific variants of native pigs in Korea: Insights into pig breeds' genomic characteristics. *Anim. Cells Syst.* **2022**, *26*, 338–347. [[CrossRef](#)]
14. Yadav, A.K.; Murthy, T.P.K.; Divyashri, G.; Prasad, N.D.; Prakash, S.; Vaishnavi, V.V.; Shukla, R.; Singh, T. Computational screening of pathogenic missense nsSNPs in heme oxygenase 1 (HMOX1) gene and their structural and functional consequences. *J. Biomol. Struct. Dyn.* **2023**, 1–20. [[CrossRef](#)]
15. Wu, J.; Wang, X.; Ding, R.; Quan, J.; Ye, Y.; Gu, T.; Xu, Z.; Zheng, E.; Cai, G.; Wu, Z.; et al. Identification of important proteins and pathways affecting feed efficiency in DLY pigs by iTRAQ-based proteomic analysis. *Animals* **2020**, *10*, 189. [[CrossRef](#)] [[PubMed](#)]

16. Reyer, H.; Shirali, M.; Ponsuksili, S.; Murani, E.; Varley, P.F.; Jensen, J.; Wimmers, K. Exploring the genetics of feed efficiency and feeding behaviour traits in a pig line highly selected for performance characteristics. *Mol. Genet. Genom.* **2017**, *292*, 1001–1011. [[CrossRef](#)]
17. Zhao, Y.; Hou, Y.; Liu, F.; Liu, A.; Jing, L.; Zhao, C.; Luan, Y.; Miao, Y.; Zhao, S.; Li, X. Transcriptome analysis reveals that Vitamin A metabolism in the liver affects feed efficiency in pigs. *G3 Genes Genomes Genet.* **2016**, *6*, 3615–3624. [[CrossRef](#)]
18. Silva, É.F.; Lopes, M.S.; Lopes, P.S.; Gasparino, E. A genome-wide association study for feed efficiency-related traits in a crossbred pig population. *Animal* **2019**, *13*, 2447–2456. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.