RESEARCH ARTICLE
JIVA 22 (2) August 2024
https://doi.org/10.55296/JIVA/22.2.2024.35-41

Received on 11/07/2024
Accepted on 24/07/2024

IDENTIFICATION OF GENOME-WIDE VARIANTS IN ANKAMALI PIGS OF KERALA BY WHOLE GENOME SEQUENCING

Michelle Elezabeth Roy[1], M. Manoj*[2], Tina Sadan[3], P.M. Rojan[2], T.V. Aravindakshan[4], A.P. Usha[5] and M.P. Unnikrishnan[6]

[1]M.V.Sc Scholar, [2]Assistant Professor, [*3] Ph.D Scholar, [4]Senior Professor and Head (Retd.),
Department of Animal Genetics and Breeding,
College of Veterinary and Animal Sciences, Mannuthy
[5]Senior Professor and Head (Retd.), [6]Assistant Professor, CPPR, Mannuthy
Kerala Veterinary and Animal Sciences University

* Corresponding author: manojm@kvasu.ac.in

ABSTRACT

Ankamali pigs are a domesticated indigenous pig variety from Kerala, recognised for their resistance to diseases, production of lean meat and ability to thrive in tropical humid climates. However, there is still limited understanding of Ankamali pig genome variation, genetic relationships with other pig breeds and the process of domestication. Here, we focus on elucidating the genome-wide variants in Ankamali pig through whole genome sequencing. Whole genome sequencing of the Ankamali pig genome generated 205.66 Gb of raw data. By GATK Haplotype Caller algorithm we identified a total of 26.6 million single nucleotide variants (SNVs) including 21.3 million single nucleotide polymorphisms (SNPs) and 5.3 million InDels. Out of the total number of SNPs obtained, 69.59 per cent were transitions and 30.41 per cent were transversions. During functional annotation of SNVs, 66.19 per cent of the mutations were silent, 33.53 per cent were missense and 0.29 per cent were nonsense mutations. The potential variants identified in this study can facilitate future research into the positive attributes of Ankamali pigs, thereby aiding in the development of more effective conservation strategies.

INTRODUCTION

According to the United Nations Food and Agriculture Organisation, pork is the most widely consumed meat in the world, accounting for 36 per cent of global meat consumption. In India, as of the 20th livestock census in 2019, the pig population was estimated at 9.06 million. However, the number of indigenous pigs is declining due to extensive crossbreeding with exotic germplasm. Indigenous pig population comprises a small number of recognised breeds distinguished by specific breed characteristics, while the majority are classified as non-descriptive or desi

pigs. Among these, the Ankamali pig is a domesticated variety that is primarily raised in Kerala. Although Ankamali pigs exhibit slower growth rates and less efficient feed conversion ratios compared to exotic or crossbred pigs (Gaur *et al*., 1997), they possess distinct advantages such as disease resistance, heat tolerance and the ability to produce leaner meat in comparison to exotic breeds (Behl *et al*., 2006). This leanness makes the meat of Ankamali pigs preferred in the market, commanding higher prices than meat from exotic or crossbred breeds. To establish a cost-effective conservation strategy, it is essential to assess the genetic makeup of Ankamali pig variety. This helps in identifying and conserving distinct genetic variations within Ankamali pigs for conservation purposes, ensuring that maximum genetic diversity is maintained at minimal expense.

Advancements in genetics and genomic technologies have led to the discovery of more genes and genomes (Enard *et al*., 2014; Frantz *et al*., 2015). Additionally, as sequencing costs have decreased, powerful tools are now accessible for studying these genetic elements. In the past decade, the adoption of high throughput SNP genotyping and whole-genome sequencing has offered detailed insights into genome-wide variants in various livestock species (Rubin *et al*., 2012; Zhao *et al*., 2018). The aim of this study was to identify the variants from whole genome sequence data of Ankamali pigs using variant calling techniques.

MATERIALS AND METHODS

Experimental animals and whole-genome sequencing

Blood samples were collected from 12 Ankamali pigs (6 females and 6 males) reared under uniform management conditions and fed with pelleted pig feed at the Centre for Pig

Production and Research, Kerala Veterinary and Animal Sciences University (KVASU), Mannuthy,. DNA samples were extracted from all pigs using the Qiagen DNeasy Tissue kit (Qiagen, Düsseldorf, Germany), and the integrity and purity of the DNA were verified by agarose gel electrophoresis and A260/280 ratio. The DNA samples from the 12 Ankamali pigs were pooled together at equimolar concentrations and subjected to high-throughput sequencing by outsourcing (Clevergene, Bengaluru). The whole genome sequencing library was prepared using the QIAseq FX DNA Library Kit for Illumina, followed by end-repair, A-tailing, ligation of pair-ended adapters, size-selection for sequencing and amplification. Finally, the amplified library was analysed and loaded onto the Illumina NovaSeq 6000 platform for cluster generation, followed

by paired-end sequencing.

Reads alignment, variant calling and annotation

To ensure high-quality data, a Fastq quality check was conducted to analyse parameters such as base quality score distribution, average base content per read and GC distribution in the raw sequence data. Reads that passed the quality check were aligned to the Sus scrofa reference genome (Suss11.1, GenBank assembly GCF_000003025.6) using BWA- MEM (Li and Durbin, 2009). The resulting bam file was sorted using Picard SortSam and processed to remove duplicates using Picard MarkDuplicates (http://picard.sourceforge.net). Subsequently, the GATK HaplotypeCaller algorithm (McKenna *et al*., 2010) was applied to call variants in the Ankamali pig genome. The identified variations were classified into types such as single nucleotide variants (SNVs), single nucleotide polymorphisms (SNPs), insertions (INS), deletions (DEL), among others. The count of transitions and transversions was determined based on the total number of SNPs identified. Variants were further categorised by functional classes including missense, nonsense and silent mutations. Annotation of these variants was performed using SnpEff v5.0.e. The identified variants were then compared with gene and exon boundaries, protein coding regions, non-coding regions and untranslated regions (UTRs) to determine their genomic locations in Ankamali pigs. Functional annotation was conducted specifically within the coding regions of the Ankamali genome to establish their genomic positions, mutation types, alternate bases, altered codons and resulting amino acid changes.

RESULTS AND DISCUSSION

To gain insight into the genomic characterisation, we conducted whole-genome sequencing on 12 Ankamali pigs, marking the first comprehensive analysis of genome-wide variants in this breed. The whole genome sequencing of Ankamali pigs yielded 205.66 Gbs of raw sequence data as paired-end 150 bp sequences, generating approximately 1.37 billion reads (1,371,054,804 reads). About 95.47% of R1 reads and 94.27% of R2 reads were obtained using paired-end sequencing, with a base quality score (Q) greater than 30. Following quality control, which retained 1,360,094,278 reads, these were mapped to the Sus scrofa 11.1 reference genome sequences (Suss11.1, GenBank assembly GCF_000003025.6), with 99.77 per cent (1,357,023,508 reads) aligning accurately to the reference genome assembly. Post-alignment, the reads were sorted and any duplicate reads were identified and removed from the alignment dataset. In

this analysis, a total of approximately 26.6 million variants were identified. These variants included 21,303,641 (80.7 per cent) single nucleotide polymorphisms (SNPs), 3,056,981 (11.4 per cent) insertions and 2,243,967 (8.4 per cent) deletions, among others. The distribution of variant types is detailed in Table 1. Michelle *et al*. (2023) reported that, n Ankamali pigs, the total genome length obtained exceeded 2.5 billion bases, with an average of one genetic variant detected for every 94 bases sequenced. Within the Ankamali pig genome, 69.58 per cent of identified SNPs were transitions, while 30.41 per cent were transversions. The observed Ts/Tv ratio for SNPs was 2.29, indicative of accuracy consistent with recommendations for whole-genome sequencing SNP analysis (DePristo *et al*., 2011). Similar studies in cattle have been referenced (Stothard *et al*., 2011; Choi *et al*., 2014; Choi *et al*., 2015).

Annotation of 26,604,589 single nucleotide variants (SNVs) in the genome of Ankamali pigs revealed that 11,966,610 variants (45.09 per cent) were located in intronic regions, with 10,056,432 variants (37.79 per cent) found in intergenic regions. The remaining variants included 2,139,505 (8.04 per cent) upstream gene variants, 875,490 (3.29 per cent) downstream gene variants, 231,530 (0.87 per cent) 3-prime UTR variants, 101,748 (0.38 per cent) 5-prime UTR variants and 73,475 (0.28 per cent) intragenic variants. This genomic region distribution in Ankamali pigs is documented in Table 2. Only a small portion of the genome, approximately two percent, codes for proteins. These findings are consistent with several whole-genome sequencing studies demonstrating that more than half of SNVs in cattle and pigs are located in non- coding regions (Mei *et al*., 2018; Zhang *et al*., 2020). Bartonicek *et al*. (2017) noted that non-coding regions such as intergenic and intron regions play critical roles in various cellular processes, including gene expression, transcriptional control and gene splicing.

The identified variations were annotated to evaluate their impact on corresponding amino acid codons. This annotation process identified specific altered codons and aminoacids resulting from each variation, categorising mutations into types such as silent, missense and nonsense mutations with their respective counts. Table 3 presents the classification of variant numbers by functional class in the Ankamali pig genome. Among all SNPs located in exon regions, 66.19 per cent were classified as silent mutations, 33.52 per cent as missense mutations and 0.29 per cent as nonsense mutations. The missense to silent mutation ratio of 0.50, comparable to findings in Nero Siciliano pigs (D'Alessandro *et al*., 2019), serves as a metric for evolutionary rates according

to Hu and Banzhaf (2008). This ratio indicates purifying selection in Ankamali pigs, suggesting the removal of harmful mutations and a slowed rate ofamino acid changes over evolutionary time.

Variants can significantly impact protein structure, with the nature and scope ofvariation playing crucial roles. Over 150 million variant effects were identified and classified into three categories: impact, region and type. According to the impact classification (Table 4), high-impact variants were relatively rare (0.01 per cent), while low-impact (0.26 per cent) and moderate-impact (0.10 per cent) variants were more prevalent in Ankamali pigs. Over 99% of these effects were categorised as modifiers. In terms of type classification (Table 5) within the Ankamali pig genome, nine high-impact variants, seven modifier impact variants, six moderate-impact variants and five low-impact variants were identified. High-impact variants can significantly disrupt protein function, potentially causing truncation or loss of function, whereas low-impact variants are assumed to have minimal effect on protein behaviour. Modifier impacts affect non-coding genes, where prediction is challenging or unsupported by evidence. Based on the region classification (Table 6) in the Ankamali pig genome, over 40% of effects were found in intronic and transcript regions. These classifications provide insights into how variants influence different aspects of genetic function and protein structure in Ankamali pigs.

CONCLUSION

The current study has generated the complete genome sequence of Ankamali pigs, a native variety from Kerala. The study highlights a higher genomic variability in Ankamali pigs compared to the reference genome, underscoring their distinct phenotypic and genetic characteristics. The study has identified potential variants that could be pivotal for future research on the advantageous traits of Ankamali pigs. These findings are crucial for developing enhanced conservation strategies aimed at preserving and utilising the unique genetic resources of this variety effectively.

ACKNOWLEDGMENTS

The authors acknowledge Kerala Veterinary and Animal Sciences University (KVASU) and All India Coordinated Research Project (AICRP) on Pig for providing the facilities for the conduct of research. We thank Clever gene Labs Pvt Ltd, Bengaluru, Karnataka for the sequencing and bioinformatics services

REFERENCES

Bartonicek, N., Clark, M.B., Quek, X.C., Torpy, J.R., Pritchard, A.L., Maag, J.L.V., Gloss, B.S., Crawford, J.,

Taft, R.J., Hayward, N.K. and Montgomery, G.W. 2017. Intergenic disease-associated regions are abundant in novel transcripts. *Genom. Biol*. **18**:1-16.

Behl, R., Sheoran, N., Behl, J. and Vijh, R.K. 2006. Genetic analysis of Ankamali pigs of India using microsatellite markers and their comparison with other domesticated Indian pig types. *J. Anim. Breed. Genet.* **123**(2):131-135.

Choi, J.W., Choi, B.H., Lee, S.H., Lee, S.S., Kim, H.C., Yu, D., Chung, W.H., Lee, K.T.,

Chai, H.H., Cho, Y.M. and Lim, D. 2015. Whole-genome resequencing analysis of Hanwoo and Yanbian cattle to identify genome wide SNPs and signatures of selection.*Mol. cells*. **38**: 466-473.

Choi, J.W., Liao, X., Stothard, P., Chung, W.H., Jeon, H.J., Miller, S.P., Choi, S.Y., Lee, J.K., Yang, B., Lee, K.T. and Han, K.J. 2014. Whole-genome analyses of Korean native and Holstein cattle breeds by massively parallel sequencing. *PloS one*. **9**:101127.

D'Alessandro, E., Sapienza, I., Giosa, D., Giuffrè, L. and Zumbo, A. 2019. In silico analysis of meat quality candidate genes among Nero Siciliano, and Italian heavy pigs genomes. *Large Anim. Rev*. **25**: 137-140.

DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., Del Angel, G., Rivas, M.A., Hanna, M. and McKenna, A. 2011. A framework for variation discovery and genotyping using next generation DNA sequencing data. *Nat. Genet*. **43**: 491-498.

Enard, D., Messer, P.W. and Petrov, D.A. 2014. Genome-wide signals of positive selection in human evolution. *Genom. Res*. **24**(6):885-895.

Frantz, L.A., Schraiber, J.G., Madsen, O., Megens, H.J., Cagan, A., Bosse, M., Paudel, Y., Crooijmans, R.P., Larson, G. and Groenen, M.A. 2015. Evidence of long-term gene flow and selection during domestication from analyses of Eurasian wild and domestic pig genomes. *Nat. Genet*. **47**(10):1141-1148.

Gaur G.K., Chhabra A.K. and Paul S. 1997. Growth intensity of indigenous pigs from birth to slaughter age. *Indian J. Anim. Sci*. **67**:344–346.

Hu, T. and Banzhaf, W. 2008. Nonsynonymous to Synonymous Substitution Ratio: Measurement for Rate of Evolution in Evolutionary Computation. In International

conference on parallel problem solving from nature. 448-457.

Li, H. and Durbin, R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*. **25**(14): 1754-1760.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M. and DePristo, M.A. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genom*. *Res*. **20**:1297-1303.

Mei, C., Wang, H., Liao, Q., Wang, L., Cheng, G., Wang, H., Zhao, C., Zhao, S., Song, J., Guang, X. and Liu, G.E. 2018. Genetic architecture and selection of Chinese cattle revealed by whole genome resequencing. *Mol. Biol. Evol*. **35**: 688-699.

Michelle, E.R., Manoj, M., Rojan, P.M., Tina, S., Aravindakshan, T.V., Usha, A.P. and Unnikrishnan, M.P. 2023. Identification of genetic variants by whole genome sequencing in Ankamali pigs of Kerala. J. *Vet. Anim. Sci*. **54**:524-531

Rubin, C.J., Megens, H.J., Barrio, A.M., Maqbool, K., Sayyab, S., Schwochow, D., Wang, C., Carlborg, Ö., Jern, P., Jørgensen, C.B. and Archibald, A.L. 2012. Strong signatures of selection in the domestic pig genome. *Proceed. Nat. Acad. Sci.* **109**:19529-19536.

Stothard, P., Choi, J.W., Basu, U., Sumner-Thomson, J.M., Meng, Y., Liao, X. and Moore,

S.S. 2011. Whole genome resequencing of black Angus and Holstein cattle for SNP and CNV discovery. *BMC. Genom*. **12**:1-14.

Zhang, W., Yang, M., Zhou, M., Wang, Y., Wu, X., Zhang, X., Ding, Y., Zhao, G., Yin, Z. and Wang, C. 2020. Identification of signatures of selection by whole-genome resequencing of a Chinese native pig. *Front. Genet*. **11**: 566255.

Zhao, P., Yu, Y., Feng, W., Du, H., Yu, J., Kang, H., Zheng, X., Wang, Z., Liu, G.E., Ernst,

C.W. and Ran, X. 2018. Evidence of evolutionary history and selective sweeps in the genome of Meishan pig reveals its genetic and phenotypic characterization. *Gigascience*. **7**(5):giy058.