Indian Journal of Animal Sciences **96** (5): 386–390, May 2026/Article
https://doi.org/10.56093/ijans.v96i5.172641

Comparative analysis of genome-wide variants in Ankamali pigs of Kerala with exotic pig breeds

MICHELLE ELEZABETH ROY[1], M MANOJ✉[2], TINA SADAN[3], P M ROJAN[4], T V ARAVINDAKSHAN[5], A P USHA[6] and M P UNNIKRISHNAN[2]

Kerala Veterinary and Animal Sciences University, Mannuthy Thrissur 680 001 Kerala India

Received: 16 October 2025; Accepted: 10 July 2026

ABSTRACT

The Ankamali pig is a domesticated indigenous variety from Kerala. Conservation of such native pig varieties is vital for future animal production due to their potential as sources of genetic diversity. The present study was undertaken with the objective of identifying genome-wide genetic variants in Ankamali pigs using next generation sequencing (NGS) technology and their genome-wide comparison with exotic pig breeds. Genomic DNA was extracted from 12 pigs, pooled after quality checks and subjected to whole genome sequencing (WGS). The sequencing produced 1.36 billion quality-checked reads, with 1.357 billion successfully mapped to the reference genome. Variant calling analysis revealed 26.6 million single nucleotide variants, including 21.3 million single nucleotide polymorphisms and 5.3 million indels. Annotation of these SNVs in the Ankamali pig genome revealed that 45.09% of variants were located in intronic regions, while 37.79% variants were present in intergenic regions. When comparing Ankamali pigs to Duroc pigs, fewer SNPs were found in the Duroc breed. A noticeable difference in the number of variants were found between Ankamali pigs and the Danish Large White pig. Whole genome comparisons among Ankamali pigs, Wannan Black pigs, and Anquing Six-End White pigs with the Sus scrofa reference genome 11.1 showed comparable total numbers of SNVs. The proportion of variants detected in intronic, upstream, downstream and 5′UTR region were different in Ankamali pigs, Wannan black and Anquing six-end white pig. The higher genomic variability in Ankamali pigs compared to the reference genome highlighted their unique phenotypic and genetic traits, underscoring the importance of conservation efforts.

Keywords: Ankamali pig, Single nucleotide variants, Variant calling, Whole genome sequencing,

In India, the pig population primarily consists of many non-descript pigs that do not conform to specific breed standards and are mostly raised under scavenging conditions. Only a small percentage of recognized indigenous breeds have distinct characteristics and production types. Indigenous breeds are increasingly valued for their potential in sustainable or organic food systems and are considered to hold significant genetic variation within livestock species. Therefore, they are crucial genetic resources that must be preserved for future food security (FAO, 2007). Ankamali pig, the only domesticated pig variety of Kerala is distributed mainly in the Ankamali block of Ernakulam District and among some backyard pig farmers in Trichur District. These pigs are also maintained at organized farms at Kerala Veterinary and Animal Sciences University, Trichur. Ankamali pigs are known for their disease resistance, heat and humidity tolerance, and suited to specific agro-climatic conditions of Kerala. Despite these advantageous traits, their long-term survival is uncertain because of inbreeding. It is a concern for indigenous breeds with small populations, which can lead to a loss of genetic diversity, affecting survival rates, reproductive efficiency, and adaptability to environmental changes (Frankham and Ralls,1998, Herrero-Medrano *et al*. 2013)

Genetic characterization of livestock breeds using genetic marker technology is essential for improving breeding practices and guiding biodiversity conservation efforts. Recently, whole-genome sequencing has become a cost-effective method for evaluating genomic variation across populations. Unlike commercially available SNP chips, whole-genome sequencing offers a comprehensive and unbiased approach to assessing genetic diversity and identifying genome-wide variants in pigs (Esteve-Codina *et al*. 2011; Groenen *et al*. 2012). Next-generation sequencing (NGS) also facilitates direct examination of polymorphisms in coding regions, which can influence selective processes. Studies such as genome-wide association studies (GWAS) are increasingly used to explore

Present address: [1]M.V.Sc Scholar, Department of AGB, CVAS, KVASU, Mannuthy, Thrissur, Kerala, India. [2]Associate Professor, CPPR, KVASU, Mannuthy, Thrissur, Kerala, India. [3]Ph.D Scholar, [4]Assistant Professor, [5]Senior Professor and Head, Department of AGB, CVAS, KVASU, Mannuthy, Thrissur, Kerala, India.[6]Senior Professor and Head, CPPR, KVASU, Mannuthy, Thrissur, Kerala, India. ✉Corresponding author email: manojm@kvasu.ac.in

the genetic mechanisms underlying complex traits in pigs at the genome level (Zhang *et al.* 2014; Xiong *et al.* 2015). A key requirement for these studies is the identification of genome-wide genetic markers. Structural variants (SVs) are significant contributors to genetic diversity, and it includes insertions/deletions (InDels), inversions, translocations and copy number variations (CNVs) (Alkan *et al.* 2011, Tattini *et al.* 2015). Additionally, analyzing genetic variation across genomes is valuable for understanding genetic diversity and population structure (Ai *et al.* 2013).

Therefore, the objective of this study was to identify genome-wide genetic variants in Ankamali pigs using next generation sequencing (NGS) technology and their genome-wide comparison with exotic pig breeds using the publicly available, whole-genome sequencing datasets. This study can provide novel insights into the genetic diversity of Ankamali pigs and their potential roles in enhancing the breeding and management of Ankamali pigs.

MATERIALS AND METHODS

Experimental animals and whole-genome sequencing: Blood samples were collected from 12 Ankamali pigs (6 females and 6 males), kept under consistent management conditions and fed pelleted pig feed at Centre for Pig Production and Research, Kerala Veterinary and Animal Sciences University (KVASU), Mannuthy, for high-throughput sequencing. DNA was extracted from these samples using the Qiagen DNeasy Tissue kit (Qiagen, Düsseldorf, Germany), with DNA integrity and purity checked by agarose gel electrophoresis and A260/280 ratios. The DNA from all pigs was pooled at equimolar concentrations. The whole-genome sequencing library was then prepared with the QIAseq FX DNA Library Kit for Illumina, including steps for end-repair, A-tailing, ligation of paired-end adapters, size-selection, and amplification. The final amplified library was analyzed and loaded onto the Illumina NovaSeq 6000 platform by outsourcing at Clever gene, Bengalur for cluster generation and paired-end sequencing.

Bioinformatic analysis: To ensure high-quality data, a Fastq quality check was performed to assess parameters such as base quality score distribution, average base content per read, and GC distribution in the raw sequence data. Raw reads that passed the quality check were aligned to the pig reference genome obtained from NCBI (GCF_000003025.6_Sscrofa_11.1_genomic). Alignment was carried out using Burrows-Wheeler Aligner (BWA) software (version 0.7.17) (Abuín *et al.* 2015) with default settings. The resulting alignments were converted from SAM format to BAM format using SAM tools (Li *et al.* 2009). The aligned reads were sorted with Picard tool (version 1.100). Duplicate reads were identified and removed using the Picard 'MarkDuplicates' command, and only uniquely mapped reads were retained for further analysis.

Variant calling analysis: Variants were called using GATK's 'HaplotypeCaller' with default parameters (McKenna *et al.* 2010). Genomic DNA from 12 animals was pooled at equimolar concentrations prior to library preparation and HaplotypeCaller was run by specifying the pooled sample ploidy (24) to account for the pooled-sequencing design. The analysis was intended to identify variants segregating within the pooled population rather than to infer individual genotypes. Following variant calling, raw variants were subjected to hard filtering using GATK VariantFiltration. Variants with low confidence were removed using standard quality metrics (QUAL) Quality by Depth (QD), Mapping Quality (MQ), Fisher Strand Bias (FS), Mapping Quality Rank Sum (MQRankSum), Read Position Rank Sum (ReadPosRankSum) and depth of coverage (DP)), retaining only high-quality variants for downstream analyses. Variant Quality Score Recalibration (VQSR) was not applied because an appropriate high-confidence training set was unavailable for the pooled sequencing data. The identified variants were categorized into single nucleotide variants (SNVs), single nucleotide polymorphisms (SNPs), insertions (INS), deletions (DEL), and other types. Variant annotation was performed with SnpEff v5.0.e (Cingolani, 2012). These variants were then compared against gene and exon boundaries, protein-coding regions, non-coding regions, and untranslated regions (UTRs) to determine their positions on the Ankamali pig genome. Functional annotations of variants in the coding regions included identifying the variant's position, mutation type, alternate base, changed codon and altered amino acids. Additionally, the genome-wide variants of Ankamali pigs were compared with those from two European pig breeds (Duroc and Danish Large White) and two Asian pig breeds (Wannan Black Pig and Anquing Six-End White Pig) using publicly available whole-genome sequencing datasets.

RESULTS AND DISCUSSION

The whole genome sequencing of Ankamali pigs generated 205.66 giga bases of raw data in paired-end 150 bp sequences, producing around 1.37 billion reads (1,371,054,804). After quality control, which retained 1,360,094,278 reads, these were aligned to the Sus scrofa 11.1 reference genome. Of these, 99.77% (1,357,023,508 reads) were accurately mapped to the reference genome. Following alignment, duplicate reads were identified and removed from the dataset. Variant calling analysis revealed 26,604,589 single nucleotide variants (SNVs), including 21,303,641 (80.1%) single nucleotide polymorphisms (SNPs), 3,056,981 (11.5%) insertions, and 2,243,967 (8.4%) deletions, among others. Among the 21,303,641 SNPs identified, 69.59% were transitions and 30.41% were transversions, resulting in an overall transition/transversion (Ts/Tv) ratio of 2.29, consistent with a high-quality mammalian whole-genome variant dataset. Annotation of these SNVs in the Ankamali pig genome showed that 11,966,610 variants (45.09%) were situated in intronic regions, while 10,056,432 variants (37.79%) were found in intergenic regions. The remaining variants were distributed

as follows: 2,139,505 (8.04%) in upstream gene regions, 875,490 (3.29%) in downstream gene regions, 231,530 (0.87%) in 3-prime UTRs, 101,748 (0.38%) in 5-prime UTRs, and 73,475 (0.28%) in intragenic regions. Among the annotated coding variants, 634 stop-gain and 147 stop-loss variants were identified. As these variants are predicted to introduce premature termination codons or alter normal translation termination, they represented potentially high-impact mutations that might influence gene function. These variants provided valuable candidates for future functional genomic and association studies in Ankamali pigs.

The numbers of different classes of variants identified in Ankamali pigs were compared with the available reports of genomic variability of exotic pig breeds (Duroc, Danish Large White, Wannan Black Pig and Anquing Six-End White Pig) and the results are presented in Table 1. Ankamali pigs showed a higher number of variants relative to the Sus scrofa 11.1 reference genome than the reported values for the European breeds (Duroc and Danish Large White), whereas the total variant counts were comparable to those reported for the Asian breeds (Wannan Black Pig and Anqing Six-End White Pig).

Comparison of Ankamali pigs with Duroc pig: WGS comparison of Ankamali pigs and Duroc pig breed with *Sus scrofa* reference genome 11.1 revealed that Ankamali pigs had over 21 million (21,303,641) SNPs, while Duroc breed had only less than 8 million (7,918,935) SNPs. Similarly, Zhang *et al*. (2020) used the same reference sequence for comparing the whole genome of Duroc pigs. In Ankamali pigs, 11,996,610 variants (45.09%) were located in intronic regions, compared to 3,011,754 variants (38.03%) in Duroc pigs. The proportion of variants in intergenic regions was 47.21% for Ankamali pigs and 59.24% for Duroc pigs. Ankamali pigs had more variants in the 5′UTR (101,748) and 3′UTR (231,530) regions compared to Duroc pigs, which had 11,141 and 70,233 variants, respectively. Upstream and downstream variants in Ankamali pigs were 2,139,505 and 875,490, respectively, while in Duroc pigs these numbers were 41,584 and 47,606. Additionally, Ankamali pigs had more synonymous (88,081) and non-synonymous (57,736) variants than Duroc pigs, which had 27,575 synonymous and 16,006 non-synonymous variants. Stop gain and stop loss variants were also more numerous in Ankamali pigs compared to Duroc pigs. Because the Sus scrofa 11.1 reference genome was assembled from a Duroc individual, a lower number of variants was expected for Duroc due to reduced genetic distance from the reference. Therefore, the observed differences in variant counts between Ankamali and Duroc should not be interpreted as evidence of greater genomic diversity in Ankamali pigs but rather as differences relative to the reference genome.

Comparison of Ankamali pigs with Danish large white pigs (DLW): In the whole genome comparison with DLW, Ankamali pigs had 21.3 million SNPs, while DLW pigs had 10.63 million SNPs (Zhang *et al*. 2020). The number of SNPs in Danish Large White was similar to that found in Duroc pigs. Ankamali pigs had a higher number of intronic variants, totaling 11,996,610 compared to 4,112,329 in DLW pigs. Additionally, Ankamali pigs had over ten million variants in intergenic regions, while

Table 1. Comparison of genomic variability of Ankamali pigs and exotic pig breeds

Sl. No.	Variant Type	Ankamali pig	Duroc	Danish Large White	Anquing-six-end white pig	Wannan black pig
1	Average mapping (%)	99.77	99.67	99.44	-	84.70
2	High-quality bases (Gb)	205.66	817.35	1603.98	469.0	501.52
3	Single nucleotide variants (SNVs)	26,604,589	-	-	26,401,669	21,316,754
4	Single nucleotide polymorphisms (SNPs)	21,303,641	7,918,935	10,635,485	-	16,249,548
5	Insertions (INS)	3,056,981	-	-	-	2,898,582
6	Deletions (DEL)	2,243,967	-	-	-	2,168,624
7	Intron	11,996,610	3,011,754	4,112,329	9,772,851	7,892,061
8	Intergenic regions	10,056,435	4,691,012	6,200,521	-	12,722,683
9	Upstream	2,139,505	41,584	62,281	149,936	121,953
10	Downstream	875,490	47,606	67,917	153,338	-
11	5′ Untranslated region(5′UTR)	101,748	11,141	19,118	55,002	44,897
12	3′ Untranslated region (3′ UTR)	231,530	70,233	99,978	244,339	198,836
13	Stop loss	634	186	58	-	-
14	Stop gain	147	38	278	-	-
15	Synonymous	88,081	27,575	45,246	140,436	-
16	Non – synonymous	57,736	16,006	24,919	72,779	-

DLW pigs had more than six million. The upstream regions in Ankamali pigs contained over two million variants, whereas DLW had only 62,281 variants. For UTR regions, Ankamali pigs had 101,748 variants in the 5′UTR and 231,530 in the 3′UTR, compared to 19,118 and 99,978 variants in DLW pigs, respectively. Furthermore, DLW pigs had fewer synonymous (45,246) and non-synonymous variants (24,919) than Ankamali pigs, which had 88,051 synonymous and 57,736 non-synonymous variants. The observed differences in variant counts between Ankamali and DLW pigs are likely to be influenced by reference bias, as both studies used the Duroc-derived Sus scrofa 11.1 reference genome. DLW, being a European breed, is likely to be more genetically related to the reference genome than Ankamali pigs, potentially explaining the discrepancies.

Comparison of Ankamali pigs with Wannan black pig: Ankamali pigs and Wannan Black pigs (a Chinese native breed) had a total of over 26.6 (26,604,589) and 21.3 million (21,316,754) single nucleotide variants (SNVs), respectively (Zhang *et al*. 2020). In Wannan Black pigs, 76.23% of these variants were single nucleotide polymorphisms (SNPs), which is lower than the 80.08% observed in Ankamali pigs. Wannan Black pigs had approximately 13.59% insertions (2,898,582) and 10.17% deletions (2,168,624), while Ankamali pigs had 11.49% insertions (3,056,981) and 8.43% deletions (2,243,967). Ankamali pigs had 45.09% of variants in intronic regions, which is higher compared to 37.02% in Wannan Black pigs. Conversely, the proportion of variants in intergenic regions was lower in Ankamali pigs (37.80%) compared to Wannan Black pigs (47.82%). For UTR regions, Ankamali pigs had 101,748 variants in the 5′UTR (0.30%) and 231,530 in the 3′UTR (0.87%). In Wannan Black pigs, 5′UTR variants accounted for 0.21% (44,897) and 3′UTR variants for 0.93% (198,836). Overall, the number of variants in Ankamali pigs was similar to that in Wannan Black pigs. The similar variant counts observed for the two Asian breeds are consistent with both being compared against the same Duroc-derived reference genome. However, these observations should be interpreted as reference-relative differences rather than direct estimates of genomic diversity.

Comparison of Ankamali pigs with Anquing six-end white pig: In a whole genome comparison with Anquing six-end white pig, Ankamali pigs and Anquing Six-End White pigs had a total of 26.6 and 26.4 million SNVs, respectively (Zhang *et al.* 2020). Ankamali pigs had 11,996,610 variants (45.09%) in intronic regions, whereas Anquing Six-End White pigs had 9,772,851 variants (37.01%) in these regions. For UTR regions, Ankamali pigs had 101,748 variants (0.37%) in the 5′UTR and 231,530 variants (0.87%) in the 3′UTR. In contrast, Anquing Six-End White pigs had 55,002 variants (0.20%) in the 5′UTR, which was fewer than in Ankamali pigs, and 244,339 variants (0.92%) in the 3′UTR, which was slightly higher than in Ankamali pigs. Regarding synonymous and non-synonymous variants, Ankamali pigs had 88,081 (0.43%) synonymous and 57,736 (0.27%) non-synonymous variants. Anquing Six-End White pigs had more synonymous (140,432) and non-synonymous variants (72,779) compared to Ankamali pigs. Overall, the total number of variants identified in Ankamali and Anqing Six-End White pigs were similar. Because both breeds were compared with the Duroc-derived Sus scrofa 11.1 reference genome, the observed similarity is likely to reflect comparable genetic distance from the reference**.** However, the proportions of variants in specific regions, such as intronic, upstream, downstream, and 5′UTR, differed between the two breeds**.**

Thus, the present study generated the first whole-genome variant catalogue for Ankamali pigs, a native pig population of Kerala. The identified variants provided a valuable genomic resource for future studies on population genetics, functional variation and trait-associated loci. Comparisons with previously published datasets indicated differences in the number and distribution of variants relative to the Duroc-derived Sus scrofa 11.1 reference genome. However, these differences should be interpreted cautiously because reference genome ancestry may influence variant discovery across breeds. The variant catalogue generated in this study will facilitate future investigations of the genetic basis of economically important traits and support conservation and breeding programmes for Ankamali pigs. Although these pigs grow slower than many exotic breeds, their unique adaptability, resulting from this genetic variation, underscore the need for conservation efforts. The potential genetic variants identified in this study could be valuable for future research into the positive traits of Ankamali pigs and will aid in developing improved conservation strategies.

ACKNOWLEDGMENTS

The authors acknowledge Kerala Veterinary and Animal Sciences University and All India Coordinated Research Project on Pig for providing the facilities for the conduct of research.

REFERENCES

Ai H, Huang L and Ren J. 2013. Genetic diversity, linkage disequilibrium and selection signatures in Chinese and Western pigs revealed by genome-wide SNP markers. *PLoS One* **8**: e56001

Abuín J M, Pichel J C, Pena T F and Amigo J. 2015. BigBWA: approaching the Burrows–Wheeler aligner to Big Data technologies. *Bioinformatics* **31**(24): 4003–4005.

Alkan C, Coe B P and Eichler E E. 2011. Genome structural variation discovery and genotyping. *Nature Reviews Genetics* **12**(5): 363–376.

Cingolani P. 2012. Variant annotation and functional prediction: SnpEff. In Variant calling: methods and protocols, New York, NY: Springer US.

Esteve-Codina A, Kofler R, Himmelbauer H, Ferretti L, Vivancos A P, Groenen M A, Folch J M, Rodríguez M C and Pérez-Enciso M. 2011. Partial short-read sequencing of a highly *inbred* Iberian pig and genomics inference thereof. *Heredity* **107**(3): 256–264.

FAO. 2007. Global plan of action for animal genetic resources

and the Interlaken Declaration. FAO Commission on Genetic Resources for Food and Agriculture, FAO: Rome, Italy.

Frankham R and Ralls K. 1998. Inbreeding leads to extinction. *Nature* **392**(6675): 441–442.

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 and Li S. 2012. Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* **491**(7424): 393–398.

Herrero-Medrano J M, Megens H J, Crooijmans R P, Abellaneda J M and Ramis G.2013. Farm-by-farm analysis of microsatellite, mt DNA and SNP genotype data reveals inbreeding and crossbreeding as threats to the survival of a native S panish pig breed. *Animal genetics* **44**(3): 259–266.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G and Durbin R. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**(16): 2078–2079.

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. *Genome Research* **20**(9): 1297–1303.

Tattini L, D'Aurizio R and Magi A. 2015. Detection of genomic structural variants from next-generation sequencing data. *Frontiers in Bioengineering and Biotechnology* **3**: 92.

Xiong X, Liu X, Zhou L, Yang J, Yang B, Ma H, Xie X, Huang Y, Fang S, Xiao S and Ren J. 2015 Genome-wide association analysis reveals genetic loci and candidate genes for meat quality traits in Chinese Laiwu pigs. *Mammalian Genome* **26**: 181–190.

Zhang L C, Li N, Liu X, Liang J, Yan H, Zhao K B, Pu L, Shi H B, Zhang Y B, Wang L G and Wang L X. 2014. A genome-wide association study of limb bone length using a Large White 9 Minzhu intercross population. *Genetics Selection Evolution* **46**: 56.

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. *Fronter in Genetics* **11**: 566255