Hlongwane, N., Smith, R., Mdyogolo, S., Nesengani, L., Molotsi, A., Tshilate, T., S., Masebe, T., Mapholi, N. (2026) Identification and characterization of structural variants in the Windsnyer and Kolbroek indigenous pigs using long read sequencing, *Proceedings 13th World Congress on Genetics Applied to Livestock Production* 25477: https://doi.org/10.31274/wcgalp.25477 ©2026 Hlongwane, et al. CC BY-NC-ND

Identification and characterization of structural variants in the Windsnyer and Kolbroek indigenous pigs using long read sequencing

N.L. Hlongwane[1*], R.M. Smith[2], S. Mdyogolo[2], L.T. Nesengani[1], A.H. Molotsi[1], T.S. Tshilate[1], S.A. Hadebe[1], T.M. Masebe[2], and N.O. Mapholi[1]

[1]University of South Africa, Johannesburg, South Africa; [2]University of South Africa, Johannesburg, South Africa; Corresponding author: hlongn2@unisa.ac.za

Introduction

Pigs provide animal protein and contribute to food security across the world. In South Africa, smallholder farmers use indigenous breeds that are adapted to the region to minimise the costs of farming. In contrast, intensive commercial production prefers exotic breeds, selected for feed efficiency and growth rates (Halimani et al., 2012; Madzimure et al., 2013). The presence of indigenous and commercial farming systems makes the conservation and genetic characterisation of indigenous pigs essential to maintain food system resilience (Halimani et al., 2010). The Kolbroek and Windsnyer are common South African indigenous pig breeds with distinct phenotypic and production characteristics compared to exotic breeds. Indigenous pigs are characterised by slow growth rates, however, are hardy and well adapted to local forage and extensive scavenging systems. A study by Moyo et al. (2018) noted a distinct skin histology and thermoregulatory adaptations linked to environmental stress in Windsnyer. Another study used a genome-wide approach and revealed that Kolbroek and Windsnyer breeds represent distinct genetic clusters within-population diversity in Southern Africa (Hlongwane et al., 2020). Despite this knowledge, detailed characterization of structural genomic variation in these significant indigenous breeds remains poorly understood. Structural variants (SVs) are particularly understudied, although they can have significant influence on traits of economic importance. While SNP studies have shown population structure and admixture patterns in pigs from South Africa, they cannot capture SV diversity effectively. This study applied PacBio HiFi long-read sequencing technology to generate genome assemblies for Kolbroek and Windsnyer. Structural variants (SVs) were identified and characterised through comparison with the *Sus scrofa* 11.1 reference and commercial pig genomes. These findings aim to support conservation efforts and develop breeding strategies suitable for local production systems.

Materials & Methods

Sample collection and whole genome sequencing. Ethical approval for the study was obtained from the University of South Africa (reference number: AREC-100818-024) and the Department of Agriculture Land Reform and Rural Development (DALRRD) under section 20 of the Animal Diseases Act 1984 (Act 35 of 84) (ref no 12/11/1/1/23 (6508 AC)). Blood samples were collected from a purebred Kolbroek (female) and from a Windsnyer trio (boar, sow, and female progeny). Genomic DNA was extracted using high molecular weight (HMW) Nanobind® protocol. The DNA was sheared to a target fragment size of 10-25 kb and libraries were prepared using PacBio SMRTbell 3.0 and sequenced on the PacBio Revio platform. A Dovetail Omni-C library was prepared from the same sample following the manufacturer's instructions and sequenced on the Illumina NovaSeq 6000 platform.

Mapping and variant identification. High-quality clean reads were obtained by removing adaptors and low-quality sequences. Genome assemblies for Windsnyer and Kolbroek were generated using the Vertebrate Genome Project (VGP) workflows on the Galaxy platform, incorporating long-read assembly and Omni-C based scaffolding. The *Sscrofa* 11.1 reference and additional pig genomes (Duroc, Large White and Landrace) were downloaded from NCBI for comparative analyses, with emphasis on Windsnyer and Kolbroek. All assemblies were assessed for quality using QUAST (v5.3.0) and BUSCO (v6.0.0). Minimap2 (v2.24) was used to align Kolbroek (KOL), Windsnyer (WIN), Duroc (DUR), Large White (LWT) and Landrace (LAN) assemblies to the *Sscrofa* 11.1 reference genome, and alignments were converted to sorted BAM files using SAMtools (v1.16.1). Structural variants (SVs) were identified using svim-asm (v1.0.3) with parameters *-min_sv_size 50*, *--max_sv_size 200000*, and *--min_mapq 10*. Prior to merging, VCF files were filtered to include only autosomes (1-18) and the X chromosome. To reduce biases, sequencing depth and SV calling variability, a binary presence/absence matrix of SVs across breeds was constructed for Principal Component Analysis (PCA) and downstream analyses. SVs were classified as private (present in one breed only), shared (present in ≥2 breeds), or core (present across all breeds). Gene annotations were extracted from the *Sus scrofa* GFF file by selecting entries annotated as "gene" and converted to BEDTools to identify SV gene overlaps. Quantitative trait loci (QTLs) were obtained from Animal QTLdb 57 (*Sscrofa* 11.1) and standardised to the reference genome. Genomic data interval analyses were performed using BEDTools, Unix command-line tools, while statistical analyses and visualization were conducted in R version 4.5.3.

Results & Discussion

The availability of high-quality indigenous pig breed genome assemblies provided a foundation for a thorough characterisation of structural diversity at the breed level. The assembled pig genomes were within the expected genome size (2.5-2.6 Gb), and showed strong contiguity, (scaffold N50: 36.9 Mb - 142 Mb), with BUSCO completeness scores of 98.9% - 99.4%, low duplication (0.3-0.4%) and minimal missing genes (≤0.8%), indicating suitability for downstream analyses. Insertions (53.3-59.2%) and deletions (40.3-46.2%) dominated the SV landscape across all breeds (Table 1), consistent with patterns observed in pig pangenome studies (Zong et al., 2023; Yang et al., 2024). Fisher's exact test showed that deletions were significantly more enriched in genic regions than insertions ($p < 2.2 \times 10^{-16}$, OR = 3.18, 95% CI: 3.12-3.25), indicating greater potential for gene disruption. SV count varied across breeds with Windsnyer and Kolbroek having the highest SV densities (18.02 and 17.69 SVs/Mb), while Duroc showed fewer, reflecting differential genomic variability (Figure 1A). Duroc and Windsnyer showed higher SV based divergence, whereas Landrace and Large White clustered closely (Figure 1B-C), suggesting shared ancestry consistent with findings by Groenen et al. (2012) and Hlongwane et al. (2020). Although most SVs were shared across breeds (41.7%), private SVs were observed particularly in Windsnyer. Small SVs (<1 kb) accounted for 89-96% of variants across breeds (Figure 1D), indicating that SV diversity is largely dominated by short indels. Together, these patterns reflect underlying demographic histories and selection pressures shaping breed differentiation. However, sample size limitation and sequencing depth bias across breeds may influence SV detection sensitivity and should be considered when interpreting absolute SV counts.

Jaccard distance analysis identified pronounced genomic stratification among breeds (Figure 2A). Duroc and Windsnyer were the most divergent (0.815 and 0.761), whereas Landrace and Large White showed the closest relationship (0.689). Hierarchical structuring supported this structure with high fidelity (cophenetic correlation = 0.973), separating Landrace, Large White

and Kolbroek from Duroc and Windsnyer. Kolbroek's central position, sharing 3,937 SVs with Windsnyer (Figure 2B) is consistent with historical connectivity among South African pig populations (Hlongwane et al., 2020). The highest structural divergence detected in the Windsnyer agrees with previous reports of its unique genetic background (Hlongwane et al., 2024). A total of 293 SVs overlapped with 1,962 genes. Notable genes were associated with reproduction (*TEX14*, *RBFOX1*), growth and metabolism (*PRDM16*, *PIK3R4*), muscle and carcass traits (*NAV2*, *LAMB4*) and immunity (*ITKN*, *CUBN*). Although a limited sample size may constrain population-level inference, permutation testing confirmed enrichment of SV gene overlaps beyond random expectation (empirical $p = 0.028$), supporting non-random distribution of structural variants in functionally relevant genomic regions.

Conclusions

South African indigenous pigs have distinct structural variant patterns that help explain differences in key characteristics such as adaptation to extensive and low-input production systems. The observed genomic diversity represents a valuable resource to understand their unique traits and genetic variation that has been substantially reduced in commercial breeds under intensive selection. Therefore, conservation of this variation is essential to prevent genetic erosion and to promote the development of resilient breeding programs adapted to local conditions.

Acknowledgements

Funding for this research was provided by the University of South Africa, African-Nuanced Sustainable Development Goals Research Support Programme (ASDG-RSP).

References

Groenen, M.A., Archibald, A.L., Uenishi, H., Tuggle, C.K., Takeuchi, Y. (2012). Nature, 491(7424): 393-398. https://doi.org/10.1038/nature11622

Halimani, T.E., Muchadeyi, F.C., Chimonyo, M., and Dzama, K. (2010). Ecological Economics, 69(5): 944-951. https://doi.org/10.1016/j.ecolecon.2010.01.005

Halimani, T.E., Muchadeyi, F.C., Chimonyo, M., and Dzama, K. (2012). South African Journal of Animal Science, 42(5), 507-510. https://doi.org/10.4314/sajas.v42i5.13

Hlongwane, N.L., Hadebe, K., Soma, P., Dzomba, E. F., and Muchadeyi, F. C. (2020). Frontiers in Genetics, 11, 344. https://doi.org/10.3389/fgene.2020.00344

Hlongwane, N.L., Dzomba, E. F., Hadebe, K., van der Nest, M. A., Pierneef, R. et al. (2024). Animals, 14(2), 236. https://doi.org/10.3390/ani14020236

Madzimure, J., Chimonyo, M., Zander, K.K., and Dzama, K. (2012). Tropical animal health and production, 45(1), 135-142. http://dx.doi.org/10.1007/s11250-014-0555-z

Moyo, D., Gomes, M., and Erlwanger, K.H. (2018). Journal of the South African Veterinary Association, 89(1), 1-10. https://hdl.handle.net/10520/EJC-11925d98eb

Yang, L., Yin, H., Bai, L., Yao, W., Tao, T. et al. (2024). Genomes Biology, 25(1), 116. https://doi.org/10.1186/s13059-024-03253-3

Zong, W., Wang, J., Zhao, R., Niu, N., Su, Y. et al. (2023). Journal of Animal Science and Biotechnology, 14(1), 136. https://doi.org/10.1186/s40104-023-00929-x