Tshilate, T., Nesengani, L., Smith, R., Mdyogolo, S., Molotsi, A., Hadebe, S., Hlongwane, N., Masebe, T., Mapholi, N. (2026) Functional landscape of repetitive elements in the Nguni sheep genome, *Proceedings 13th World Congress on Genetics Applied to Livestock Production*: 24090 https://doi.org/10.31274/wcgalp.24090 ©2026 *Tshilate, et al.* CC BY-NC-ND

Functional landscape of repetitive elements in the Nguni sheep genome

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

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

Introduction

Nguni sheep (*Ovis aries*) are indigenous to the Southern Africa region and are generally known to be well adapted to the tropical climatic conditions. They have excellent mothering ability and resistance to the diseases endemic to the region. (Lu et al., 2024; Nesengani et al., 2025). Despite their importance, limited efforts have been made to generate genetic resources for understand Nguni's genetic mechanisa. To address this gap, we recently published a high-quality chromosomal-level genome assembly for the Nguni sheep using PacBio HiFi and Omni-C sequencing data. This reference genome (2.9 Gb, 96.1% BUSCO completeness) showed that repetitive sequences make up 47.39% of its content, with Long Interspersed Nuclear Elements (LINEs) being the most prevalent class at 19.02% (Figure 1) (Nesengani et al., 2025). This finding is significant, as repetitive elements are no longer considered "junk" DNA but are now understood to be major drivers of genomic evolution, variation, and the regulation of gene expression (Liao et al., 2023). Moreover, they may comprise over two-thirds of mammalian genomes and are recognized to play critical roles in a spectrum of biological processes, from normal cellular homeostasis to disease susceptibility (de Koning et al., 2011; Liao et al., 2023; Snowbarger et al., 2024).

Despite this growing recognition, the specific role of transposable elements and other repeat elements in shaping the adaptive traits of indigenous breeds like the Nguni sheep remains profoundly underexplored. This research addresses this gap by leveraging a new genome assembly to systematically investigate the landscape of repetitive elements in the Nguni sheep genome and their functional interplay with protein-coding genes. By identifying TE-gene overlaps and analysing their functional annotations, this study aims to illuminate the potential molecular mechanisms such as insertional mutagenesis, epigenetic regulation, or contribution of novel regulatory sequences that may underpin the breed's renowned resilience. This foundational analysis provides insight into the genetic architecture of local adaptation but also establishes a crucial resource for future comparative genomic studies and informed conservation breeding programs.

Materials & Methods

Data and genome used. The primary genomic resource for this analysis was the Nguni sheep reference genome (GenBank Assembly accession: JBLGTL000000000, BioProject: PRJNA1175667) (Nesengani et al., 2025). The genome assembly file (in FASTA format) and its corresponding structural annotation file (in GTF format), containing coordinates of protein-coding genes, were downloaded from Figshare (https://doi.org/10.6084/m9.figshare.28078811). To ensure a sensitive and comprehensive identification of repetitive sequences, a custom repeat library was constructed. The standard Dfam/RepBase library for *Ovis aries* was supplemented with de novo repeat families predicted from the Nguni genome using RepeatModeler v2.0.4 (Bailly-Bechet et al., 2014; Tarailo-Graovac and Chen, 2009).

This hybrid approach captures both conserved repeats and lineage- or breed-specific repetitive elements that may not be fully represented in curated databases.

Identification and classification of repetitive elements. The repetitive landscape was characterized using RepeatMasker v4.1.5 (Tarailo-Graovac and Chen, 2009), a standard tool for screening DNA sequences against libraries of repetitive elements. The software was run with the custom repeat library in -lib mode. The -s (slow) search option was selected for higher sensitivity, and the -xsmall option was used to generate soft-masked output where repetitive bases are converted to lowercase, preserving the original sequence. The primary output file from RepeatMasker was parsed to generate a comprehensive summary of repetitive element content. This includes the total genomic proportion, and the breakdown by major classes (e.g., LINEs, SINEs, LTR retrotransposons, DNA transposons) and families (e.g., L1, Bov-B, ERV1). A BED file detailing the genomic coordinates of each annotated repetitive element was created for downstream intersectional analysis.

Genomic distribution and functional enrichment analysis. To investigate the spatial relationship between repetitive elements and functional genes, the coordinates of masked repeats were intersected with the coordinates of protein-coding genes from the assembly's Gtf annotation file using BEDTools v2.30.0. Specifically, the intersect Bed function was used to identify Repeats located within exonic, intronic, or promoter regions (defined as sequences 2 kb upstream of the transcription start site) and Repeats located within 10 kb upstream or downstream of gene boundaries. The list of genes found to overlap with or be proximal to repetitive elements was subjected to functional enrichment analysis. Gene ontology (GO) terms and Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathways were assigned using EggNOG-mapper v2.1.12 (Cantalapiedra et al., 2021) via the eggNOG 5.0 database.

Results & Discussion

Repetitive element analysis reveals that the Nguni sheep genome hold extensive repetitive sequences (47.39%) (Figure 1a), with retrotransposons (LINEs and SINEs) being most prevalence class associated with protein-coding genes. A total of 22,755 genes (87.7% of all annotated genes) were found to contain or be proximal to repetitive elements. Functional annotation identified 15 genes within repeat-associated regions, all of which (100%) were in these regions. Fisher's exact test showed significant enrichment of immune-related genes among repeat-associated genes ($p = 0.003$). Key genes identified included *DOCK2* (immune cell migration), *ITPR2* and *CADPS* (calcium signalling), and TENM2 and CNTN4 (neuronal development). These genes are central to calcium signalling, leukocyte migration and autophagy pathways biological processes critical for environmental resilience. (Liang et al., 2024; Snowbarger et al., 2024; Wang et al., 2024).

The disproportionate presence of genic retrotransposons aligns with emerging models where TEs act as dynamic genomic modulators (Ferrari et al., 2021; Zhao et al., 2023). However, this study has limitations: genome-wide GO enrichment was not possible due to limited annotation coverage and we did not compare with commercial breeds to assess breed-specificity. Additionally, the absence of epigenetic or expression data precludes definitive claims about regulatory causality (Ferrari et al., 2021; Kooverjee et al., 2023; Liang et al., 2024). Future research employing comparative population genomics, epigenomic sequencing, and functional validation is essential to confirm the mechanistic role of these TE-gene interactions and determine whether these elements are drivers of adaptation in this climate-resilient breed.

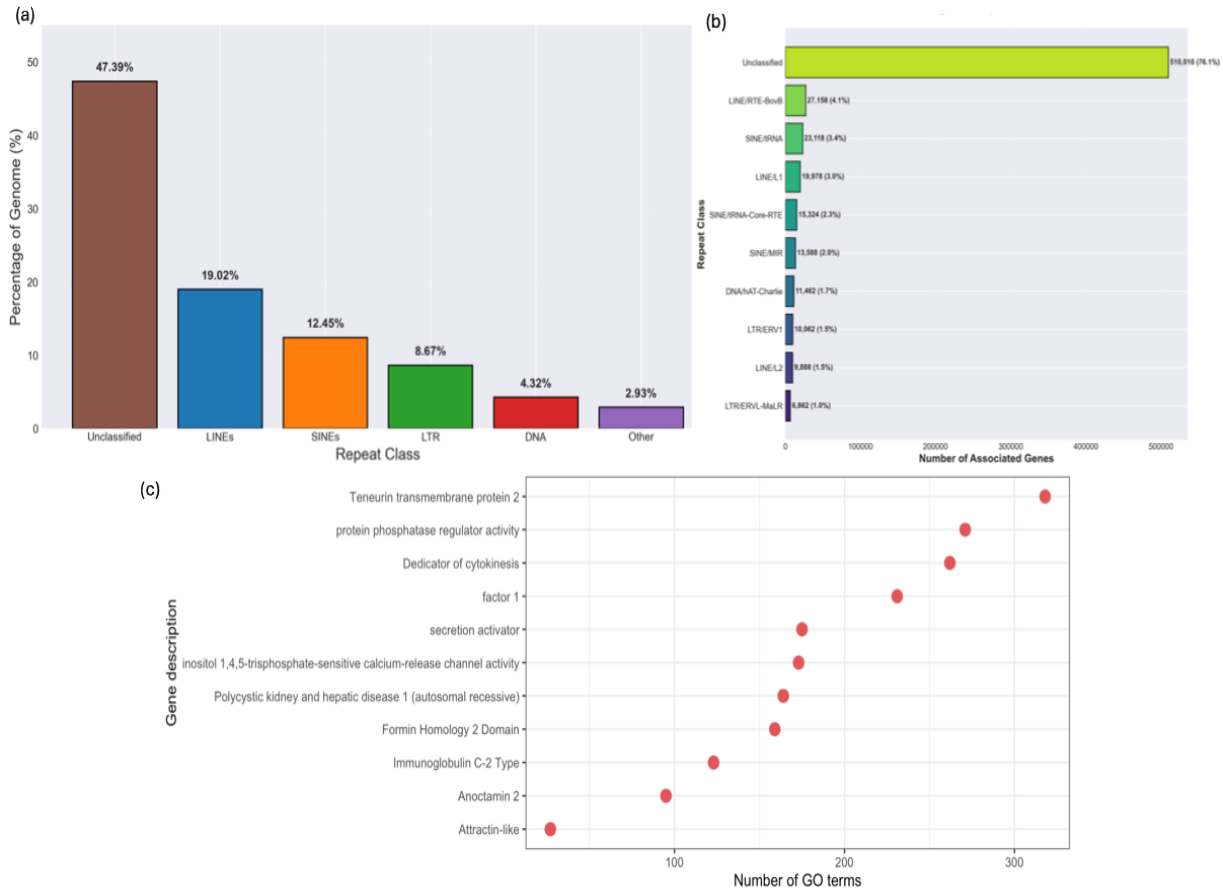

Figure 1. Repeat landscape and functional gene annotation of the Nguni sheep genome. (a) represent the proportion of the genome occupied by major repeat classes indicated by bars. (b) Number of genes overlapping each major repeat class. (c) Dot plot of the genes with the highest number of Gene Ontology (GO) terms. The y-axis represents gene descriptions, and the x-axis shows the number of associated GO terms.

Conclusions

This study is the first to map and analyse repetitive elements in the Nguni sheep genome. The results connected specific relationship between transposable elements and key genes which are involved in natural development, immunity and calcium signalling to the breed. These findings suggest repetitive sequences are not just genomic "junk" but may actively modulate genetic networks that improve resilience in individuals. This reframes indigenous breeds as an important reservoir of adaptive variation, where repetitive DNA provides new clues for breeding climate resilient livestock.

Acknowledgements

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

References

Bailly-Bechet, M., Haudry, A., Lerat, E., 2014. Mob. DNA 5, 1–15. https://doi.org/10.1186/1759-8753-5-13

Cantalapiedra, C.P., Hernandez-Plaza, A., Letunic, I., Bork, P., Huerta-Cepas, J., 2021Mol. Biol. Evol. 38, 5825–5829. https://doi.org/10.1093/MOLBEV/MSAB293

de Koning, A.P.J., Gu, W., Castoe, T.A., Batzer, M.A., Pollock, D.D., 2011. PLoS Genet. 7, e1002384. https://doi.org/10.1371/JOURNAL.PGEN.1002384

Ferrari, R., Grandi, N., Tramontano, E., Dieci, G., 2021. Retrotransposons as Drivers of Mammalian Brain Evolution.11, 376. https://doi.org/10.3390/LIFE11050376
Kooverjee, B.B., Soma, P., van der Nest, M.A., Scholtz, M.M., Neser, F.W.C., 2023. Animals (Basel). 13. https://doi.org/10.3390/ANI13152513
Liang, X., Duan, Q., Li, B., Wang, Y., Bu, Y., Zhang, Y., Kuang, Z., Mao, L., An, X., Wang, H., Yang, X., Wan, N., Feng, Z., Shen, W., Miao, W., Chen, J., Liu, S., Storz, J.F., Liu, J., Nevo, E., Li, K., 2024. Proc. Natl. Acad. Sci. U. S. A. 121, e2322291121. https://doi.org/10.1073/PNAS.2322291121;WGROUP:STRING:PUBLICATION
Liao, X., Zhu, W., Zhou, J., Li, H., Xu, X., Zhang, B., Gao, X., 2023. Commun. Biol. 6, 954-. https://doi.org/10.1038/S42003-023-05322-Y;SUBJMETA
Lu, Z., Yuan, C., An, X., Chen, Z., Guo, T., Liu, J., 2024. Scientific Data 2024 11:1 11, 711-. https://doi.org/10.1038/s41597-024-03564-x
Nesengani, L.T., Tshilate, T., Mdyogolo, S., Smith, R., Masebe, T., Raphulu, T., Moila, A., Sharaf, A., Muzemil, S., Houaga, I., Muigai, A., Djikeng, A., Kuja, J., Katee, S., Osuji, J.O., Mungloo-Dilmohamud, Z., Botes, M., Mbizeni, S., Jarvis, E., Ebenezer, T.G.E., Ntanganedzeni, M., 2025. Scientific Data 2025 12:1 12, 1193-. https://doi.org/10.1038/s41597-025-05514-7
Snowbarger, J., Koganti, P., Spruck, C., 2024. Biomolecules 14, 1250. https://doi.org/10.3390/BIOM14101250
Tarailo-Graovac, M., Chen, N., 2009. Curr. Protoc. Bioinformatics Chapter 4. https://doi.org/10.1002/0471250953.BI0410S25
Wang, L., Taylor, T., Rathnakumar, K., Khyzha, N., Liang, M., Alizada, A., Campitelli, L.F., Pour, S.E., Patel, Z.M., Antounians, L., Tobias, I.C., Hughes, T., Roy, S., Mitchell, J.A., Fish, J.E., Wilson, M.D., 2024. bioRxiv 2022.10.25.513724. https://doi.org/10.1101/2022.10.25.513724
Zhao, P., Peng, C., Fang, L., Wang, Z., Liu, G.E., 2023. Genet. Sel. Evol. 55, 50. https://doi.org/10.1186/S12711-023-00821-2