scientific data

Check for updates

OPEN

DATA DESCRIPTOR

A chromosomal level genome assembly of Nguni Sheep, *Ovis aries*

Lucky Tendani Nesengani[1]✉, Thendo Tshilate[1], Sinebongo Mdyogolo[1], Rae Smith[1], Tracy Masebe[1], Thomas Raphulu[2], Acclaim Moila[3], Abdoallah Sharaf[4,5], Sadik Muzemil[6], Isidore Houaga[1,7], Anne Muigai[8,9,10], A. Djikeng[1,11], Josiah Kuja[12], Sally Katee[11], Julian O. Osuji[13], Zahra Mungloo-Dilmohamud[14], Marietjie Botes[15], Skhumbuzo Mbizeni[1], Erich Jarvis[10], ThankGod Echezona Ebenezer[16] & Mapholi Ntanganedzeni[1]✉

Nguni sheep (*Ovis aries*) are indigenous to the Southern Africa region and common within the smallholder and poor resources farming systems. They are well adapted to different agroecological regions. However, limited genomic resources such as high-quality reference genomes have hindered our understanding of its adaptation and establishment of an effective breeding program. To address this, we assembled a chromosomal-level genome of Nguni sheep using a combination of PacBio HiFi reads and Omni-C reads. The genome size was estimated to be 2.9 Gb with a contig/scaffold N50 74 Mb and 99.6 Mb and a genome completeness of 96.1%, as estimated by the Benchmarking Universal Single-Copy Orthologs (BUSCO) program. The final genome encompassed a total of 25,926 protein-coding genes. The findings of this study provide a valuable genomic resource for understanding the adaptability of the Nguni sheep and the establishment of effective breeding programs.

Background & Summary

The Nguni sheep are indigenous to Southern Africa and comprise four ecotypes, namely Zulu, Pedi, Swazi, and Landim sheep[1,2]. Their origin can be traced to around 200 and 400 AD during the migration period of the Nguni people from central Africa to the Southern part of the continent[1,3–5]. The Nguni sheep are hardy, can walk long distances in search of forage and water and are naturally resistant to gastro-intestinal parasites and tick-borne diseases that are endemic in the region[6,7]. They therefore rarely fall sick or need vaccination[4,6]. They also have a strong foraging instinct making them well adapted to the harsh climate that is usually hot and humid[7]. Due to these traits, the Nguni sheep are therefore low maintenance, making them highly attractive to local smallholder and emerging farmers. Throughout Africa, sheep are important and valuable assets that provide meat, milk, wool, hide, and manure[8,9,46]. Sheep also play a critical cultural role in many communities by enhancing cultural practices and ceremonies due to the phenotypic appearance of multi-colored coat colors and patterns[7,10–13]. However, in recent times, the introduction of larger-bodied exotic sheep breeds has resulted in the Nguni sheep becoming a vulnerable species largely due to farmers' preference for the larger-bodied exotic breeds[14]. The Nguni sheep have small bodies when compared to the exotic breeds, giving the impression that they are a low-performance breed[15]. This perception resulted in indiscriminate crossbreeding practices that aimed to

[1]College of Agriculture and Environmental Sciences, UNISA Science Campus, Florida, Johannesburg, South Africa. [2]Limpopo Department of Agricultur, Polokwane, South Africa. [3]Inqaba Biotech, Pretoria, South Africa. [4]SequAna Core Facility, Department of Biology, University of Konstanz, Konstanz, Germany. [5]Genetic Department, Faculty of Agriculture, Ain Shams University, Cairo, Egypt. [6]Rothamsted Research, Harpended, United Kingdom. [7]The Roslin Institute and Centre for Tropical Livestock Genetics and Health (CTLGH), University of Edinburgh, Edinburgh, UK. [8]National Defence University-Kenya, Nakuru, Kenya. [9]Jomo Kenyatta University of Agriculture and Technology, Juja, Kenya. [10]Vertebrate Genome Lab, The Rockefeller University, New York, NY, USA. [11]International Livestock Research Institute, Nairobi, Kenya. [12]Bioinformatics Center, University of Copenhagen, Copenhagen, Denmark. [13]Department of Plant Science and Biotechnology, University of Port Harcourt, Port Harcourt, Nigeria. [14]Department of Digital Technologies, FoICDT, University of Mauritius, Mauritius, Mauritius. [15]School of Law, University of KwaZulu-Natal, Durban, South Africa. [16]African BioGenome Project (AfricaBP), Nairobi, Kenya. ✉e-mail: nesenlt@unisa.ac.za; maphon@unisa.ac.za

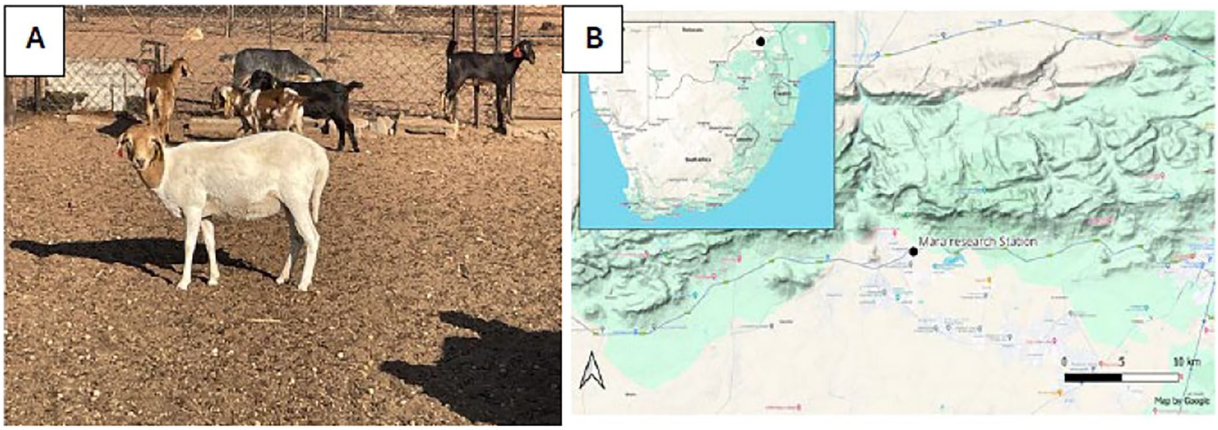

Fig. 1 A Nguni sheep (Brown Head and White body) at Mara Research Station, Limpopo Province of South Africa (**A**) and a map of Mara Research station located in Limpopo Province of South Africa (**B**).

improve their so-called inferior traits, proving to be a great threat to indigenous Nguni sheep. This has prompted the establishment of conservation programs to save these unique and irreplaceable genetic resources[14,16].

The Nguni sheep, like many African indigenous genetic resources[17], does not have a high-quality reference genome. Having a high-quality reference genome that is accessible to researchers is an essential genomic tool to establish efficient breeding programs and conservation strategies to guard against the extinction of indigenous genetic resources[18,19]. There are few genetic studies conducted on Nguni sheep. These studies assessed the genetic structure of Nguni sheep using microsatellite markers and Ovine 50 K chip and found that the Nguni sheep, despite the different ecotypes, the genetic admixtures were detected with Damara and Dorper sheep[2,3]. These findings are one of the pieces of evidence that we need to conserve the genetics of the Nguni sheep. Selepe *et al.*[14] also emphasized that the Nguni sheep is currently at a high risk of extinction as they reported the admixture of Nguni sheep with exotic breeds. Therefore, this breed has been reported as one of the important breeds within the South African government conservation program to avoid extinction[20–22].

Concurrently, in recent years, climate change has shown the need to have breeds that are well-adapted to changing climatic environments[23]. In this context, the development of effective breeding programs to improve production performance and efforts aimed at the conservation of indigenous breeds needs urgent attention. Such efforts require establishing high-quality reference genomes[19]. To understand the organism's genetic architecture requires a comprehensive evaluation of the genome and its function. The generation of an error-free, near-gapless reference genome is a step closer to realise such understanding. The majority of the indigenous species constitute a valuable genetic makeup that is under-studied. This is true in the African indigenous sheep. Indigenous African sheep genetic resources play a crucial role in the provision of food security and socio-economic components. The available sheep reference genome[24] may not fully explain the unique genome characteristics of the African indigenous breeds such as the Nguni sheep. This may be due to factors such as genetic diversity, evolutionary history, and the adaptation to environmental factors such as different climates in which the animals are exposed. Understanding the genetics of Nguni sheep is important to safeguard their genetic diversity, improve their production, and enhance their resilience to climate change and conservation. The African BioGenome Project (AfricaBP) is a coordinated Pan-African effort to sequence the genomes of 100,000 endemic and indigenous African species[17]. As part of AfricaBP's initiatives, we have produced a near error-free chromosome-scale complete genome of the indigenous Southern African Nguni sheep (*Ovis aries*). Here, a high-quality reference genome of a female Nguni sheep was generated using Pacific Biosciences (PacBio) HiFi and Dovetail genomics Omni-C technologies.

Methods

Ethics statement. The sample collection procedure including processing and handling of the animals used in this study was approved by the University of South Africa ethics committee (reference number: AREC-100818-024), Limpopo department of Agriculture 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). Furthermore, the ethics procedure was guided by the AfricaBP policy on ethics which emphasises ethical, legal, and social issues throughout the research activities. Finally, this work benefited from compliance consultations with the Department of Forestry, Fisheries and Environment (DFFE) as the Competent National Authority for biodiversity framework and digital sequence information in South Africa. Given that this work is solely for academic purposes with no commercialisation intentions either now or in the future, and the samples were sourced and the genome sequenced in South Africa, a bioprospecting permit is not required from DFFE.

Sample collection and sequencing. In this study, a pure breed of ewe Nguni sheep (Fig. 1A) was selected from Mara research station (Fig. 1B) herd for sampling in Limpopo province of South Africa (23° 05′S, 29° 25′E). Blood was collected from the jugular vein in EDTA tubes by the veterinarian and placed on dry ice, immediately transported to Inqaba Biotech laboratories in Pretoria, South Africa, and stored at −80 °C freezer until further processing. High molecular weight genomic DNA was extracted from 200 ul blood using Nanobind protocol for

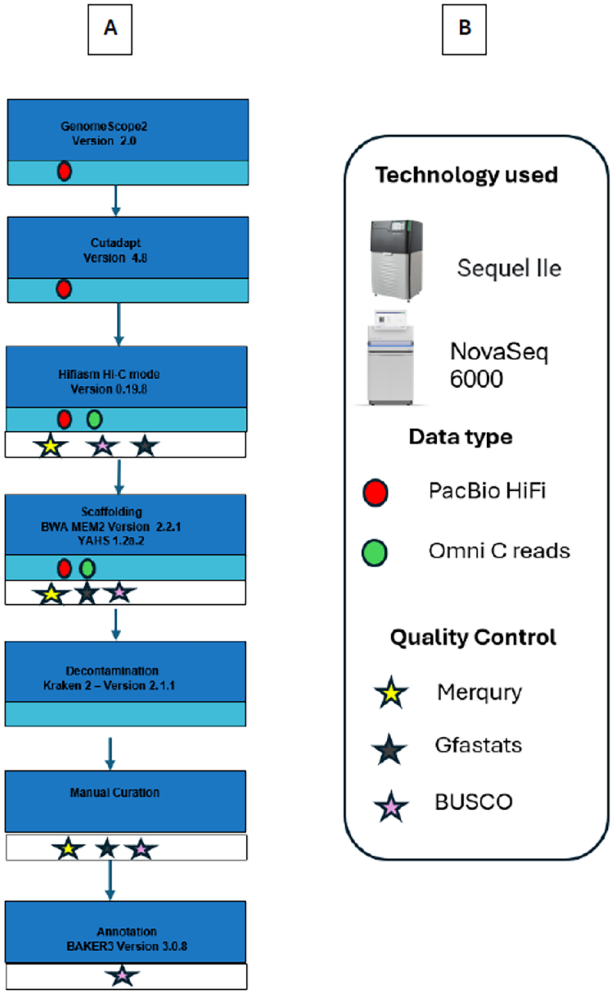

Fig. 2 The assembly workflow (**A**), Technology used (**B**).

whole blood high molecular weight (HMW) DNA extraction[25] to construct a sequencing library. The protocol was optimised for extraction from 200 ul of whole blood. Library sequencing on the PacBio Sequel IIe platform was done using SMRTbell® prep kit 3.0[26] following the manufacturer's instructions. Dovetail Omni-C library prep was performed from the same sample used for HiFi sequencing following the manufacturer's instructions. The resulting library was sequenced on NovaSeq 6000 instrument (Illumina). The total HiFi output data was 99 Gb at a coverage of 30 × while the output for Omni-C was 300 million read pairs.

Genome assembly. The genome assembly was conducted through a series of steps (Fig. 2) using the Vertebrate Genome Project (VGP) pipeline for genome assembly[27]. Firstly, the sequenced PacBio HiFi reads were profiled to assess genome, repeat contents and heterozygosity rate. We employed Cutadapt[28] on the HiFi reads to remove adaptors sequences then used meryl to generate k-mers (k = 21) count distribution and GenomeScope[29] to estimate genome for profiling the genome. This process was carried out using PacBio HiFi reads. Subsequent to this, the genome assembly was conducted using Hifiasm[30] in Hi-C mode, incorporating HiFi and Omni-C reads. This generates two haplotypes (hap 1 and hap 2) which are essentially equivalent. The assembly was further subjected to scaffolding with YaHS[31] then ran the decontamination using dual decontamination preparation workflow which make use of Kraken2[32] as shown in Fig. 3 below. The dual decontamination workflow generates scaffolds versus Omni-C reads contact map with both haplotypes for curations. PretextMap was used to manually curate the scaffolds and by orienting and correcting mis-assemblies using the genomic proximity signal of the Omni-C reads alignment against scaffolds. The genome assembly was submitted to the National Center for Biotechnology Information (NCBI) for validation purposes, to confirm that it had been thoroughly cleaned of any extraneous taxa sequences. The submitted genome was subsequently employed for downstream analyses. The final genome size of Nguni sheep is 2.9 Gb, with a heterozygosity rate of 0.236% (Fig. 3A). The genome was sequenced at a coverage of 30X. The quality of the assembly was further assessed using Merqury completeness, which was 99.6% for both assemblies (Table 1). The genome completeness was further assessed using BUSCO with a dataset lineage of cetartiodactyla_odb10, which had 96.1% complete genes, 0.8% fragmented, and 3.1% missing genes respectively, totalling 13,335 (Fig. 3B).

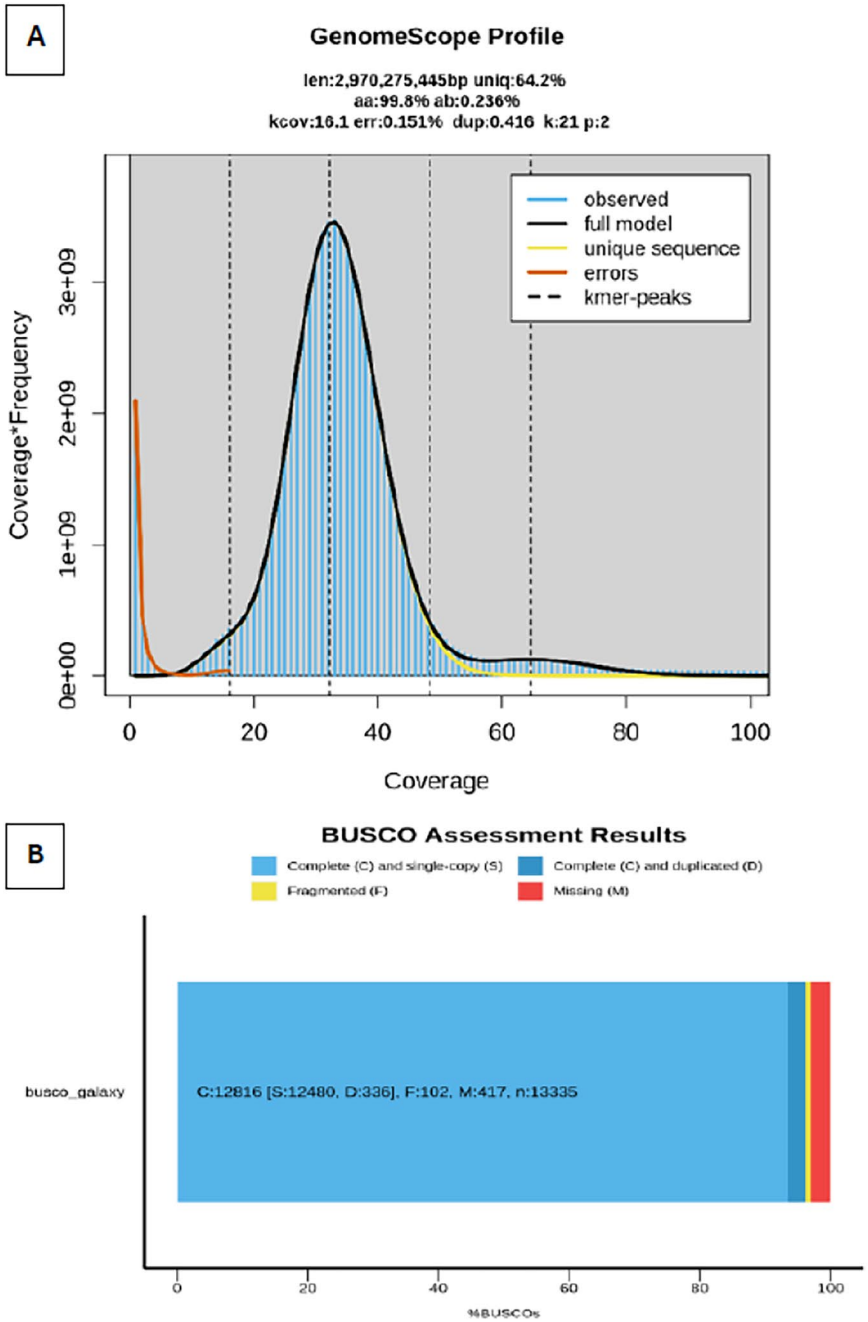

Fig. 3 (A) The linear plot generated by GenomeScope showing the observed k-mer profile. The coverage is shown on the x-axis and coverage frequency on the y-axis. (**B**). BUSCO plot of the Nguni Sheep. The BUSCO percentages of the complete (single – light blue), complete (duplicates - dark blue), fragmented (yellow) and Missing (red) are shown along the x-axis.

Assembly	Solid k-mers in assembly	Solid k-mers in reads	Completeness %
Assembly_01	1987975724	2098597483	94.7288
Assembly_02	1983305274	2098597483	94.5062
Both	2089729521	2098597483	99.5774

Table 1. Assembly quality results of merqury completeness.

Genome annotation. The preliminary phase in genome annotation entails the identification and masking of repeat elements in the Nguni sheep genome. This process utilizes RepeatMasker[33], a software tool that facilitates the identification and masking of repetitive elements within genomic sequences. The

Metrics	Nguni	Hu (GCA_040805955.1)	Tibetan (GCA_017524585.1)	Rambouillet (GCF_016772045.2)
Genome size	2.9 Gb	2.9 Gb	2.7 Gb	2.7 Gb
scaffolds	106	28	58	142
Total scaffold length	2.9 Gb	2.9 Gb	2.6 Gb	2.7 Gb
Scaffold N50	99.6 Mb	103.4 Mb	105.2 Mb	101 Mb
Scaffold L50	9	9	8	8
contigs	133	28	168	226
Contig N50	74 Mb	103.4 Mb	74.6 Mb	43 Mb
Contig L50	14	9	13	24
coverage	30.0X	52.2X	171.4X	55X
GC content	43.4%	43%	42%	42.0%

Table 2. Assembly stats for Nguni sheep as compared to the assembly stats for Hu, Tibetan, and Rambouillet Sheep.

characterization of Nguni_Sheep-specific repeats was facilitated through the implementation of repeat modelling with RepeatModeler v0.9, a software that incorporates tools such as RepeatScout v1.0.6[34], RECON v1.5.0[35], and TRF v4.09[36]. The identified repeat elements were then subjected to annotation using the "one_code_to_find_them_all" Perl script[37].

Protein coding genes for repeat masked assemblies were predicted using TIBERius v1.1.4[38]. The resulting predicted protein coding genes were subsequently used to build orthologous gene sets among the selected sheep breeds, Hu (GCA_040805955.1)[39], Tibetan (GCA_017524585.1)[40], and Rambouillet (GCF_016772045.2)[24] using Orthofinder[41], which clusters shared orthologous genes.

The assembly results are reported in Table 2, showed a scaffold N50 above 99.6 Mb and a contig N50 of 74 Mb. Our assembly showed a comparable contiguity with the latest sheep reference genomes, which include Ramboiuillet sheep genome from the International Sheep Genomics Consortium[24], as well as the Hu[39] and Tibetan[40] sheep genomes (Table 2). The variation among these assemblies is attributable to the sequencing coverage and the sequencing technology employed which likely influence the completeness and accuracy of the genome[42]. Additionally, the results provide the first comprehensive genome information for Nguni sheep, highlighting the genomic differences that may underpin phenotypic variations among breeds. Subsequent to scaffolding, the Nguni sheep genome assembly underwent a decontamination process to eliminate contaminants and further refine the genome. This refinement ensures the reliability of the assembly and enhances its utility as a resource for genomic and phenotypic studies.

Genome annotation of nguni sheep. The repeat annotation analysis of the Nguni genome revealed the presence of approximately 1435.61 Mb of repetitive sequences, constituting 47.39% of the total sequence length. The major categories of repetitive elements identified in the Nguni genome are summarized in Fig. 4, while a distribution of repetitive elements across the sheep genomes are presented in Table 3. The genome annotation parameters are reported in Table 4. Among the repetitive elements, 91.73% were annotated as known repeat motifs, while 8.27% were classified as uncharacterized elements unique to the Nguni Sheep. The most prevalent class of known repeats was the long interspersed nuclear elements (LINEs), which constituted 19.02% of the genome. While the number of repeat elements can influence genome size and certain levels of duplication within the genome may play a significant role in shaping the diverse gene repertoires in the organisms[43], this study has yet to determine the extent to which these genomic factors impact gene composition among sheep breeds.

The genome was annotated using a deep learning-based ab initio gene structure prediction tool, which generated gene predictions based on inherent structural evidence. This resulted in 25,926 predicted protein-coding genes. The BUSCO analysis, based on the cetartiodactyla_odb10 lineage dataset showed that 89.7% of the BUSCOs were complete, with 88.1% being single-copy and 1.6% duplicated. The annotation process also identified 1.2% fragmented BUSCOs and 9.1% missing BUSCOs, resulting in a total of 13,335 cetartiodactyla BUSCO groups that were searched.

Further experimental evidence may be necessary to provide a more robust demonstration of the distinct characteristics of the Nguni sheep. The primary objective of our study was to generate a high-quality reference genome. This information serves as the essential tool for studies aimed at improving the Nguni sheep's production and for conservation purposes. The generated genome reference will be made publicly available and accessible to the researchers for use in further research.

Data Records

The raw sequences were submitted to the NCBI and assigned accession number for SRA (SRR33210015) data of PRJNA1175667[44]. The final genome assembly for the Nguni sheep was submitted to the NCBI under the BioProject: PRJNA1175667, BioSample: SAMN44368269 and assigned the accession number JBLGTL000000000[45]. The annotation data of the genome is shared through figshare[46].

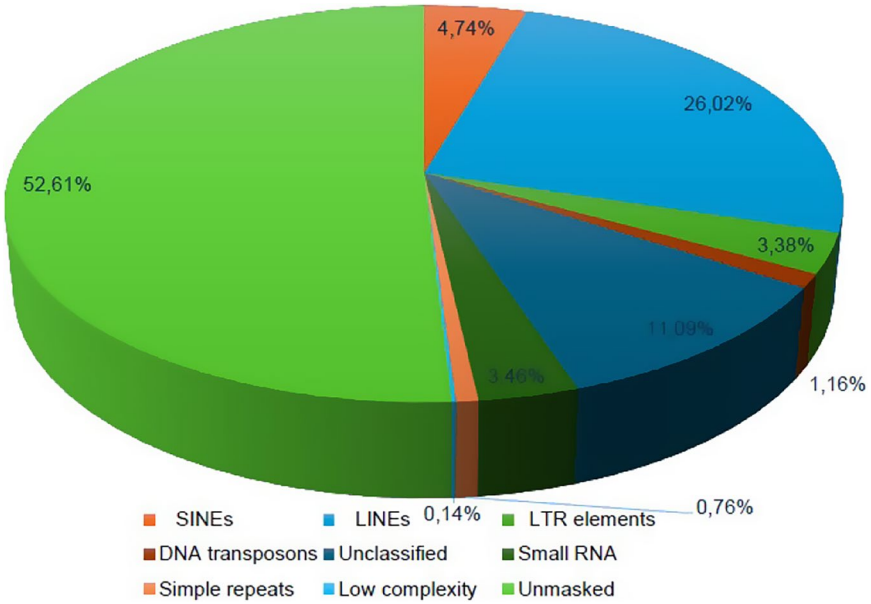

Fig. 4 Repetitive element distribution in percentage for the Nguni sheep genome. LINEs and SINEs.

Repeat elements	Number of elements	Length bp (%)
SINE	1033965	143681255 (4.74%)
LINE	1988994	788197598 (26.02%)
LTR element	404432	102487912 (3.38%)
DNA element	203332	35111612 (1.16%)
Small RNA	667574	104823114 (3.46%)
Simple repeat	530179	23022899 (0.76%)
Low complexity	35057	4280546 (0.14%)
Unclassified	86629	335995892 (11.09%)

Table 3. Summary of repeat elements identified in the genome of Nguni sheep.

Mean gene length (bp)	21926.2
Number of exons per gene	6.9627
Mean length (bp)	208.286
Total length (Mb)	1742.97
Genes with introns (%)	72.0975
Mean length (bp)	3433.12
Total length (Mb)	530.859
Average length (bp)	21926.2

Table 4. Genome annotation parameters of the Nguni sheep.

Technical Validation

The reference genome assembly of Nguni Sheep was supported by VGP workflows for genome assembly which were designed to reduce human error by employing workflows that produces near error-free high-quality reference genome. The assembly was done using Hifiasm on a Hi-C mode. The pipeline is comprised of different workflows that make use of methods such as BUSCO and mercury to assess the integrity of the genome.

Code availability

All the analyses done in the current study were processed by employing the VGP workflows that are publicly available in galaxy (https://galaxyproject.org/projects/vgp/workflows/). All the commands and pipelines were executed following the manual and protocols of the corresponding bioinformatics software. In all the workflows, unless mentioned and where necessary, we used the default parameters.

Received: 2 January 2025; Accepted: 2 July 2025;
Published online: 10 July 2025

References

1. Badenhorst, S. & Magoma, M. Sheep (Ovis aries) of Venda speakers during the second millennium AD in South Africa. *Int J Osteoarchaeol* **32**, 944–950 (2022).
2. Nxumalo, K. S., Grobler, P., Ehlers, K., Nesengani, L. T. & Mapholi, N. O. The genetic assessment of South African Nguni sheep breeds using the ovine 50 K chip. *Agriculture* **12**, 663 (2022).
3. Kunene, N. W. *et al*. Genetic diversity in four populations of Nguni (Zulu) sheep assessed by microsatellite analysis. *Ital J Anim Sci* **13**, 3083 (2014).
4. Nurse, G. T. Population movement around the northern Kalahari. *Afr Stud* **42**, 153–163 (1983).
5. Huffman, T. N. The archaeology of the Nguni past. *Southern African Humanities* **16**, 79–111 (2004).
6. Ramsay, K., Harris, L. & Kotzé, A. *Landrace Breeds: South Africa's Indigenous and Locally Developed Farm Animals* (1998).
7. Nyamukanza, C. C., Scogings, P. F., Mbatha, K. R. & Kunene, N. W. Forage–sheep relationships in communally managed moist thornveld in Zululand, KwaZulu-Natal, South Africa. *Afr J Range Forage Sci* **27**, 11–19 (2010).
8. Gorkhali, N. A., Han, J. L. & Ma, Y. H. Mitochondrial DNA variation in indigenous sheep (Ovis aries) breeds of Nepal (2015).
9. Pollott, G. & Wilson, R. T. Sheep and goats for diverse products and profits (2009).
10. Kunene, N. W. & Fossey, A. A survey on livestock production in some traditional areas of Northern Kwazulu Natal in South Africa. *Livest. Res. Rural Dev* **18**, 30–33 (2006).
11. van Zyl, E. A. & Dugmore, T. J. Imvu–The indigenous sheep of KwaZulu-Natal: A Zulu Heritage.
12. Kunene, N. W., Bezuidenhout, C. C., Nsahlai, I. V. & Nesamvuni, E. A. A review of some characteristics, socio-economic aspects and utilization of Zulu sheep: implications for conservation. *Trop Anim Health Prod* **43**, 1075–1079 (2011).
13. Fossey, A., Nesamvuni, E. A. & Kunene, N. Characterisation of Zulu (Nguni) sheep using linear body measurements and some environmental factors affecting these measurements. *S Afr J Anim Sci* **37**, 11–20 (2007).
14. Selepe, M. M., Ceccobelli, S., Lasagna, E. & Kunene, N. W. Genetic structure of South African Nguni (Zulu) sheep populations reveals admixture with exotic breeds. *PLoS One* **13**, e0196276 (2018).
15. Dzomba, E. F., Chimonyo, M., Snyman, M. A. & Muchadeyi, F. C. The genomic architecture of South African mutton, pelt, dual-purpose and nondescript sheep breeds relative to global sheep populations. *Anim Genet* **51**, 910–923 (2020).
16. Ribeiro, M. N. *et al*. Threatened goat breeds from the tropics: the impact of crossbreeding with foreign goats. *Sustainable Goat Production in Adverse Environments: Volume I: Welfare, Health and Breeding* 101–110 (2017).
17. Ebenezer, T. E. *et al*. Africa: sequence 100,000 species to safeguard biodiversity. *Nature* **603**, 388–392 (2022).
18. Paez, S. *et al*. Reference genomes for conservation. *Science (1979)* **377**, 364–366 (2022).
19. Brandies, P., Peel, E., Hogg, C. J. & Belov, K. The value of reference genomes in the conservation of threatened species. *Genes (Basel)* **10**, 846 (2019).
20. Köhler-Rollefson, I. Farm animal genetic resources: Safeguarding national assets for food security and trade (2004).
21. Mahlase, E. M. & Fakir, S. The use of indigenous animal genetic resources to promote sustainable rural livelihoods in South Africa. *Community-Based Management of Animal Genetic Resources* **127** (2003).
22. Mapiye, C., Chikwanha, O. C., Chimonyo, M. & Dzama, K. Strategies for sustainable use of indigenous cattle genetic resources in Southern. *Africa. Diversity (Basel)* **11**, 214 (2019).
23. Rojas-Downing, M. M., Nejadhashemi, A. P., Harrigan, T. & Woznicki, S. A. Climate change and livestock: Impacts, adaptation, and mitigation. *Clim Risk Manag* **16**, 145–163 (2017).
24. Davenport, K. M. *et al*. An improved ovine reference genome assembly to facilitate in-depth functional annotation of the sheep genome. *Gigascience* **11**, giab096 (2022).
25. Mitchell, M. W. *et al*. High molecular weight DNA extraction and long-read next-generation sequencing of human genomic reference standards (2020).
26. Ekholm, J. *et al*. Streamlined SMRTbell® library generation using addition-only, single tube strategy for all library types reduces time to results. *J Biomol Tech* **30**, S38–S38 (2019).
27. Larivière, D. *et al*. Scalable, accessible and reproducible reference genome assembly and evaluation in Galaxy. *Nat Biotechnol* **42**, 367–370 (2024).
28. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal* **17**, 10–12 (2011).
29. Vurture, G. W. *et al*. GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* **33**, 2202–2204 (2017).
30. Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods* **18**, 170–175 (2021).
31. Zhou, C., McCarthy, S. A. & Durbin, R. YaHS: yet another Hi-C scaffolding tool. *Bioinformatics* **39**, btac808 (2023).
32. Wood, D. E. & Salzberg, S. L. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol* **15**, R46 (2014).
33. Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinformatics* **25**, 4–10 (2009).
34. Price, A. L., Jones, N. C. & Pevzner, P. A. De novo identification of repeat families in large genomes. *Bioinformatics* **21**, i351–i358 (2005).
35. Bao, Z. & Eddy, S. R. Automated de novo identification of repeat sequence families in sequenced genomes. *Genome Res* **12**, 1269–1276 (2002).
36. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* **27**, 573–580 (1999).
37. Bailly-Bechet, M., Haudry, A. & Lerat, E. "One code to find them all": a perl tool to conveniently parse RepeatMasker output files. *Mob DNA* **5**, 1–15 (2014).
38. Gabriel, L., Becker, F., Hoff, K. J. & Stanke, M. Tiberius: end-to-end deep learning with an HMM for gene prediction. *Bioinformatics* **40**, btae685 (2024).
39. Luo, L.-Y. *et al*. Telomere-to-telomere sheep genome assembly reveals new variants associated with wool fineness trait. *bioRxiv* 2024–2027 (2024).
40. Liang, X. *et al*. Genomic structural variation contributes to evolved changes in gene expression in high-altitude Tibetan sheep. *Proceedings of the National Academy of Sciences* **121**, e2322291121 (2024).
41. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol* **20**, 1–14 (2019).
42. Tshilate, T. S., Ishengoma, E. & Rhode, C. A first annotated genome sequence for Haliotis midae with genomic insights into abalone evolution and traits of economic importance. *Mar Genomics* **70**, 101044 (2023).
43. Meagher, T. R. & Vassiliadis, C. Phenotypic impacts of repetitive DNA in flowering plants. *New Phytologist* **168**, 71–80 (2005).
44. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRR33210015 (2025).
45. Nesengani,L.T. and Mapholi,N. Ovis aries breed Pedi, whole genome shotgun sequencing project. Genbank. https://identifiers.org/ncbi/insdc:JBLGTL000000000.1 (2025).
46. Nesengani, L.T. and Mapholi, N. Nguni Sheep Annotation Data. e https://doi.org/10.6084/m9.figshare.28078811 (2025).

Acknowledgements
This project is supported by University of South Africa, Africa-nuanced sustainable development goals research support programme. The samples for this project were provided by Mara research station, Limpopo province of South Africa and the technical support was provided by the AfricanBP. A special thanks to Mapholi lab for providing the resources and managing the project through grant of African-Nuanced Sustainable Development Goals Research Support Programme (ASDG-RSP) awarded by the University of South Africa to Mapholi lab. Finally, we would like to thank Mr Thomas Nkhumeleni Mbedzi from the Department of Forestry, Fisheries and Environment for the consultation engagement on this work.

Author contributions
L.T.N.- Literature review, sampling, analyses, diagrams, manuscript drafting, validation and manuscript review; T.T.- analyses, validation and manuscript review; S.M. - sampling and manuscript review; R.S.- diagrams, validation and manuscript review; T.M.- validation and manuscript review; T.P.- sampling; A.M.- review; A.S.- analyses, validation and manuscript review; S.M.- validation and manuscript review, I.H.- validation and manuscript review; AM- literature review, validation and manuscript review; A.D.- manuscript review; JK- review; A.M.- sequencing; S.K.- compliance; J.O.- review; Z.D.- review; Z.M.D.- validation and manuscript review; M.B.- Compliance; S.K.- sampling; E.J.: Validation and review the manuscript; T.E.E.- literature review, validation and manuscript review; N.M.: connived idea, literature review, secured funding validation and manuscript review.

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to L.T.N. or M.N.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.