Molotsi, A., Tshilate, T., Nesengani, L., Mdyogolo, S., Smith, R., Hlongwane, N., Hadebe, S., Masebe, T., Mafokwane, T., Mapholi, N. (2026) Detection of structural variants overlapping genes with putative functions in Southern White Rhino (*Ceratotherium simum*) in efforts to its conservation in South Africa, *Proceedings 13th World Congress on Genetics Applied to Livestock Production* 23971: https://doi.org/10.31274/wcgalp.23971 ©2026 Molotsi, et al. CC BY-NC-ND

# Detection of structural variants overlapping genes with putative functions in Southern White Rhino (*Ceratotherium simum*) in efforts to its conservation in South Africa

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

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

## Introduction

Southern White Rhinos (SWR) are known for their distinctive two horns. They were on the brink of extinction and listed on the IUCN Red List of Threatened Species in 2020 as Near Threatened due to poaching (Emslie, 2020). This has led to increased conservation efforts, and in-situ breeding programs have been perpetuated in Southern Africa. Besides poaching, there are diseases that threaten the conservation of the Southern White Rhino. Among these diseases is *Mycobacterium tuberculosis* (Miller *et al.,* 2018) and various skin diseases caused by nematodes (Bauwens *et al.,* 1996; King'ori *et al.,* 2024). Structural variants are known to be used as markers in diseases in humans and other mammals (Chen *et al.,* 2024). Structural variants are segments that are larger than 50 base pairs that include insertions, deletions, duplications, inversions, and translocations (Chen *et al.,* 2024). Currently, both a draft genome and a chromosomal-level genome are available for the Southern White Rhino, which was sequenced using Illumina short reads and PacBio HiFi reads, respectively. However, there are limited studies available on the structural variants within genes known to be involved in putative functions to adaptation and disease resistance traits in the SWR in South Africa. Therefore, this study aimed to identify structural variants in the Southern White Rhino through the alignment of a recently sequenced SWR to the chromosomal-level SWR genome (Oskolov *et al.,* 2025) and elucidate the structural variants and SNPs found overlapping genes regulating disease resistance traits.

## Materials & Methods

**Ethics statement.** The procedures for animal handling, sample collection, and all research-related activities were approved by the University of South Africa, Animal Research Ethics Committee (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), reference number: 12/11/1/1/23 (6508 AC).

**Sampling.** Blood samples were obtained from one female SW Rhino from the South African National Biodiversity Institute (SANBI) biobanking unit in Pretoria, South Africa. High molecular weight DNA was extracted using the Monarch extraction kit using 200 µL of blood as input. The SMRTBell® prep kit 3.0 was used to prepare the sequencing library. Sequencing was done using PacBio Revio, yielding a total of 13.3 Gb of data with a genome coverage of 5.7X.

**Genome assembly.** The genome assembly was carried out through the Vertebrate Genome Project (VGP) workflows on the Galaxy platform (Abueg *et al.,* 2024). These workflows start with utilising the genomeScope2 (Cheng *et al.,* 2022) for genome characterization and

estimating genome size, which resulted in an estimated genome size of 2,48 Gb for the Southern White Rhino from k-mer =17 analysis (Figure 1a). This was followed by the genome assembly using HiFiasm (Cheng *et al.,* 2021). BUSCO was run using the laurasiatheria_odb10 database, which includes 12,234 genes from 52 species (Figure 1b). The assembly was then subjected to purging to remove haplotypic duplication and contig overlaps in the draft assembly using Purge_dups (v1.2 .6) (Guan, 2019). Thereafter, the genome was evaluated using the gfastats v 1.3.11 tool (Formenti *et al.,* 2022).

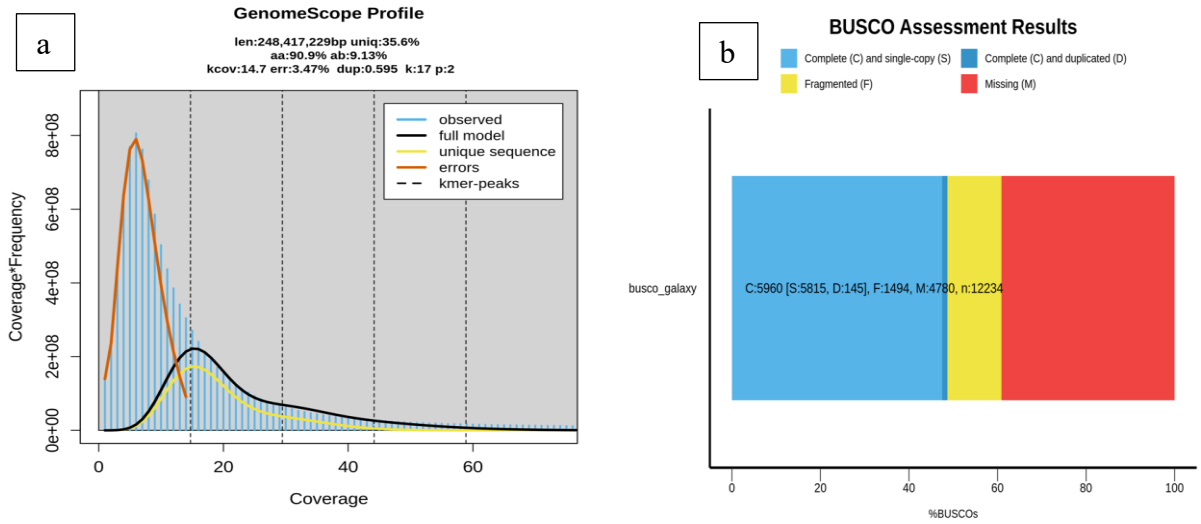


**Figure 1 a)** Kmer analysis to estimate genome size and heterozygosity of the genome. **b)** The BUSCO summary image presenting the complete and single-copy genes, complete and duplicated genes, fragmented genes, and missing genes.

**Alignment and structural variants identification.** Due to the low coverage obtained from the HiFi sequencing, we aligned the raw reads of our sample to the chromosomal-level genome of the ASM2365373v1(GCA_023653735.1) (Oskolov *et al.,* 2025) using minimap Version 2.28 (Li, 2018). This was followed by sorting and the removal of duplicates using SAMtools Version 1.22 (Danecek *et al.,* 2021), whereafter structural variants were detected using Sniffles 2. The annotation file obtained from CerSimSim1.0 reference genome and structural variant output file from Sniffles were converted to BED format. Bedtools was used to intersect regions where structural variants overlap with genes. SNP annotation was done using snpEff v5.3 and EnrichR was utilized for the gene ontology analysis.

## Results & Discussion

The assembly statistics prior to purging indicated a genome size of 2.3 Gb and the scaffold N50 of 67 Kb. The number of scaffolds was 45,369, with an average scaffold length of 50,792.93 Kb and a GC content of 39,82%.

After purging, the number of scaffolds decreased to 39,464, and the scaffold N50 is 70 Kb, and the scaffold L50 is 8,690 (Table 1). Compared to the chromosomal genome of both the Southern White Rhino (Oskolov *et al.,* 2025) and Northern White Rhino (Ruggeri *et al.,* 2026), the SWR genome used in this study is fragmented and will benefit from increasing the sequencing depth to improve its genome quality.

However, we were able to annotate SNPs and detect structural variants for the SWR, comparing it to the chromosomal-level assembly (Oskolov *et al.,* 2025). The SNP annotation yielded a total of 1,154,334 variants, including 1,279 high-impact SNPs, i.e. SNPs that can lead to major

structural changes of proteins. The number of structural variants detected using Sniffles2 was 7,531, including 114 deletions, 79 insertions, and 1 inversion on chromosome 41.

**Table 1.** Summary statistics of the genome assembly of the Southern White Rhino compared to previously published assemblied of the Southern White Rhino (Oskolov *et al*., 2025) and the Northern White Rhino (Ruggeri *et al.,* 2026).

| Assembly statistics | SWRf | SWRm | NWRm |
|---|---|---|---|
| Genome size | 2.3 Gb | 2.9 Gb | 2.5 Gb |
| Number of chromosomes | - | 42 | 41 |
| Number of scaffolds | 39,464 | 306 | 390 |
| Scaffold N50 | 70 Kb | 67.6 Mb | 65.9 Mb |
| Scaffold L50 | 8,690 | 16 | 14 |
| Contig N50 | 70 Kb | 50.4 Mb | 16 Mb |
| GC content | 39.82% | 41.5% | 41% |
| Genome coverage | 5.7X | 28X | 75X |
| Assembly level | Contig level | Chromosomal | Chromosomal |

SWRf - Southern White Rhino (female) in current study (after purging); SWRm - Southern White Rhino (male), GCA_023653735.1 after Oskolov *et al.* (2025); NWRf - Northern White Rhino (female), GCA_021442165.1 after Ruggeri *et al.* (2026)

Genes overlapping structural variants on chromosome 1 were LOC101398896, encoding for transmembrane TPR repeat-containing protein 2 and LOC101394096, encoding for myotubularin-related protein 13 isoform X2. On chromosome 2 the insertion of 542bp was overlapping with LOC101391778, encoding for forkhead box protein P2. High-impact SNPs were found to be significantly associated with GRB10 gene, known to regulate hematopoietic stem cell self-renewal and regeneration after an injury (Yan *et al.,* 2022), as well as P2RX7, which plays a role in immune cells that acts against inflammation.

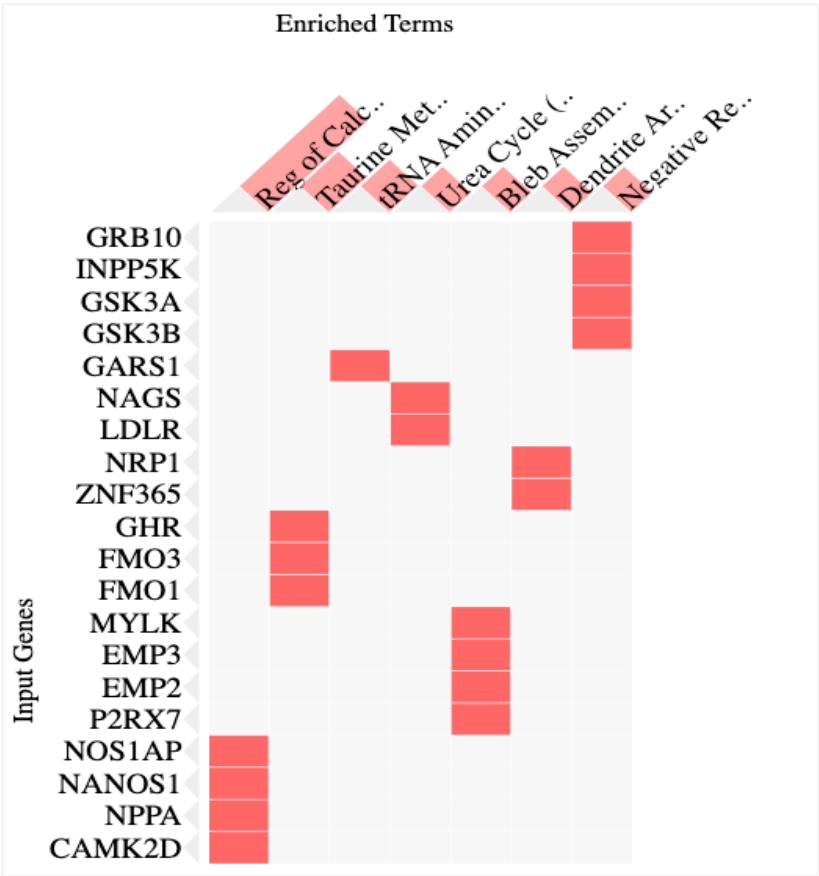

**Figure 2.** Enriched gene ontology terms for the Southern White Rhino, indicating the biological process influenced by high impact SNPs and the genes associated with them.

## Conclusions

The study revealed structural variants and SNPs present within the genome of the Southern White Rhino. These variants overlap with genes that play a role in putative functions in the Southern White Rhino. Therefore, further studies including phenotypic data on specific disease incidences in SWR are needed to validate these findings.


## Acknowledgements

We would like to thank Kim Labuschagne from SANBI for providing the sample for sequencing. As well as UNISA for providing the funding for the project from the catalytic niche area. We also appreciate the technicians at Inqaba, Terry-Lynn Costello and Acclaim Moila for assisting with the laboratory work for this project.


## References


Abueg, L.A.L., Afgan, E., Allart, O., *et al*. (2024) Nucleic Acids Res. 52(W1),W83-W94 (2024). https://doi:10.1093/nar/gkae410

Chen Y., Khan M.Z., Wang X., *et al*. (2024) Front. Vet. Sci. 11:1416220. https://doi:10.3389/fvets.2024.1416220

Cheng, H., Jarvis, E.D., Fedrigo, O., *et al*.(2022) Nat Biotechnol. 40(9),1332-1335 (2022). https://doi:10.1038/s41587-022-01261-x

Cheng H., Concepcion G.T., Feng X., *et al*. (2021) Nat Methods. 18:170-175. https://doi:10.1038/s41592-020-01056-5

Danecek P, Bonfield JK, Liddle J, *et al*. (2021) Gigascience. 10(2). https://doi:10.1093/gigascience/giab008

Emslie R. (2020) Ceratotherium simum. The IUCN red list of threatened species 2020. Available at: https://www.iucnredlist.org/species/4185/45813880 Visited on: 06-02-2026

Formenti, G., Abueg L., Brajuka A., *et al*. (2022) Bioinfor. 38(17),4214-4216. https://doi.org/10.1093/bioinformatics/btac460

Gribouval, L., Sourdaine, P., Lareyre, JJ. *et al*. (2018) Sci Rep **8**, 6942. https://doi.org/10.1038/s41598-018-24643-1

Guan D. (2019) Purge_dups. Preprint posted online 2019

King'ori E., Waiguchu G., Ruoro M., *et al*. (2024) Transbound Emerg Dis. 10;2024:2655970. https://doi: 10.1155/2024/2655970

Li H. Minimap2: Bioinformatics. 2018;34(18):3094-3100. https://doi:10.1093/bioinformatics/bty191

Miller M.A., Buss P., Parsons S.D.C., *et al*. (2018) Emerg Infect Dis. 24:2373-2375. https://doi: 10.3201/eid2412.180293

Ruggeri E., Prunier J., Sirard M.A., *et al*. (2026) PLoS One.21:e0340594. https://doi:10.1371/journal.pone.0340594

Oskolkov N., Jin C., Clinton S.L., *et al*. (2025) Gigascience. 14:giaf108. https://doi:10.1093/gigascience/giaf108

Yan X, Himburg HA, Pohl K, *et al*. Cell Rep. 2016 Nov 1;17(6):1584-1594. https://doi:10.1016/j.celrep.2016.10.025