Contribution ID: **395** Type: **Talk**

Reference Genome Assembly, Pangenome Construction, and Population Analysis of the Spotted Hyena (Crocuta Crocuta) from the Kruger National Park

Monday, 1 December 2025 12:00 (20 minutes)

The spotted hyena (Crocuta crocuta) is a highly social carnivore with a complex behavioural and ecological functions, making it an important model for studying genetic diversity, adaptation, and evolution. However, previous draft genomes for C. crocuta have been incomplete and derived from captive individuals, limiting insights into natural genetic variation. Here, we present a high-quality de novo genome assembly and the first pangenome of wild spotted hyenas sampled from the Kruger National Park, South Africa, alongside population-level analysis.

Using Oxford Nanopore Technologies (ONT) long-read sequencing, we assembled a 2.39 Gb reference genome with a scaffold N50 of 19.6 Mb and >98% completeness. We further performed short-read resequencing at 10-32X depth per individual, revealing >4 million single nucleotide variations and and ~1 million insertions and deletions per individual. To capture genomic variation beyond a single reference, we constructed a draft pangenome using Progressive Genome Graph Builder (PGGB). The resulting pangenome comprises ~2.47 Gb, with 35.2 million nodes, 48.4 million edges, and 159,060 paths, incorporating sequences from all individuals. Its graph structure revealed substantial topological differences, which may correspond to biologically relevant variations.

The breadth of these analyses required extensive use of the CHPC's computing resources. Long-read genome assembly and polishing were executed on high-memory nodes to accommodate the error-correction and scaffolding steps. Repeat and gene annotation pipelines (RepeatModeler, BRAKER3) as well as variant discovery with GATK and BCFtools were parallelised to accelerate execution. Pangenome graph construction was particularly computationally intensive, requiring large-scale parallelisation and significant memory and storage capacity to manage multi-genome alignments and graph building.

This study provides the most contiguous wild-derived genome to date for the species, the first draft pangenome for C. crocuta, and establishes a foundation for future conservation and comparative genomics. Importantly, it demonstrates the critical role of HPC resources in enabling large-scale bioinformatics pipelines - from genome assembly to pangenome construction and population-level analysis - in non-model organisms.

Presenting Author

Ansia van Coller

Email

ansia.vancoller@mrc.ac.za

Student or Postdoc?

No. Not a student nor Postdoc.

Institute

South African Medical Research Council

Registered for the conference?

Yes

CHPC User

Yes

CHPC Research Programme

CBBI1195

Workshop Duration

Primary author: Dr VAN COLLER, Ansia (South African Medical Research Council)

Co-authors: Dr GLANZMANN, Brigitte (SAMRC/BGI and Stellenbosch University); Dr CARSTENS, Nadia (SAMRC Genomics Platform); Prof. KINNEAR, Craig (SAMRC Genomics Platform); Ms COLE, Victoria (SAMRC Genomics Platform); Dr KERR, Tanya (Stellenbosch University); Prof. GOOSEN, Wynand (Stellenbosch University); Prof. MILLER, Michele (Stellenbosch University)

Presenter: Dr VAN COLLER, Ansia (South African Medical Research Council)

Session Classification: HPC Applications

Track Classification: Bioinformatics and Biological Sciences