



Contribution ID: 41

Type: **Student Micro-talk**

Assembly of the rooibos genome using long and short read sequencing data

In South Africa, more than 3000 plant species are used in traditional medicine. Medicinally active compounds have been identified from diverse endemic South African plant species, but so far close to nothing is known about the genomic background of these plants. Plant genome sequencing can provide information on the genes and biosynthetic pathways that are relevant for plant growth (e.g. stress tolerance) and for production of medicinally active compounds. The introduction of short-read sequencing technologies (such as Illumina) revolutionized plant genome sequencing, owing to their reduced costs, high accuracy rates and throughput. Assembling plant genomes however remains challenging due to high levels of repetitiveness, particularly for highly heterozygous and/or polyploid plants. Repeat regions are generally not spanned by the short reads, leading to fragmented or incomplete draft-assemblies. The rapidly evolving Nanopore sequencing technology has made it possible to obtain longer read lengths (>10kbp), which simplifies and improves genome assemblies.

Our research team has sequenced the genome and diverse transcriptomes of rooibos (*Aspalathus linearis*) using Illumina sequencing technologies. The best assembly of the short read data, achieved using MaSuRCA, yielded an N50 of 10kbp. We have also established a protocol for Nanopore sequencing of rooibos, generating up to 0.5 TB of sequencing data per MinION run. This data is currently being assembled at CHPC. Here, we are presenting on the assembly of the rooibos genome using different long read (Flye, canu, Shasta) and hybrid (MaSuRCA, HASLR) genome assembly programs, as well as diverse polishing tools such as racon, medaka, nanopolish and pilon. A first assembly of the MinION data using Flye improved the N50 to 77kbp; the largest contig is ~1Mbp long!

Student?

Yes

Supervisor name

Dr Uljana Hesse

Supervisor email

uhesse@uwc.ac.za

Primary authors: Ms MGWATYU, Yamkela (Department of Biotechnology, University of the Western Cape, Robert Sobukwe Road, Bellville 7535); Dr HESSE, Uljana (Department of Biotechnology, University of the Western Cape, Robert Sobukwe Road, Bellville 7535)

Presenter: Ms MGWATYU, Yamkela (Department of Biotechnology, University of the Western Cape, Robert Sobukwe Road, Bellville 7535)

Session Classification: Micro-talks

Track Classification: Bioinformatics and Biological Sciences