



Contribution ID: 255

Type: **Talk (invited)**

## High throughput computing considerations for Public Health Genomics

*Monday, 3 December 2018 11:00 (20 minutes)*

Bioinformatics occupies the space between biology and computing and aims to answer questions in biology using analytical and computing methodology. At the South African National Bioinformatics Institute (SANBI), our research focus is on methods to store, retrieve and analyze genetic information that spans both communicable and non-communicable diseases. In the context of Public Health, that is a need to interrogate genetic information (DNA) from both hosts (human) and pathogens (bacterial or viruses) to understand susceptibility to diseases and ultimately to track infection trends in real time.

The ever-increasing volume of data being generated in the public domain places a strain on in-house computing resources. While the computing facility at SANBI-UWC is adequate for initial R&D, these resources are inadequate to complete projects timeously. The nature of the bioinformatics workflows that require CHPC resources can be grouped into (1) high throughput computing resources that are needed to describe the 1000s of genetic messages in a genome, versus (2) high performance computing resources that is needed to model a drug-protein interaction environment - these simulations require days/weeks of dedicated compute time.

We have leveraged the CHPC facility in the context of infectious disease research with a view to analyze genetic variation in bacterial genomes, and to identify drug targets in pathogen genomes. These analyses requires an environment that is able to support reproducible workflows and virtualization of the software environment. Examples of these use cases will be described.

### Presenter Biography

Alan Christoffels is the Director of the South African National Bioinformatics Institute and the DST/NRF Research in Bioinformatics and Public Health Genomics. After completing his studies in 2001, he moved to Singapore for a postdoc in the Sydney Brenner lab with a focus on genome evolution. During this time he contributed to the annotation and genome analysis of the Pufferfish genome and developed method for detecting genome duplication events. This methodology was later used in other international genome projects while he established his group in Singapore in 2004. In 2007 he returned to South Africa and established his group at the South African National Bioinformatics Institute based on the University of the Western Cape campus. He has driven a number genomics projects on the African continent. During the past 4 years his team has focused on infectious disease models including *M.tuberculosis*. The Christoffels laboratory ([christoffels.sanbi.ac.za](http://christoffels.sanbi.ac.za)) has been building tools ([combattb.sanbi.ac.za](http://combattb.sanbi.ac.za)) and analyzing data that leads to a better understanding of host-pathogen interactions. Central to these methods is a computational platform that relies on reproducible workflows and a graph database. The ultimate goal is to ensure that these methods are readily accessible by biomedical researchers with a view of improved disease surveillance.

**Primary author:** Prof. CHRISTOFFELS, Alan (University of the Western Cape)

**Presenter:** Prof. CHRISTOFFELS, Alan (University of the Western Cape)

**Session Classification:** HPC Applications

**Track Classification:** Bioinformatics and Biological Sciences