



Contribution ID: 248

Type: **Talk (invited)**

Next-generation DNA sequencing: why HPC is becoming a vital component of biological research

Tuesday, 4 December 2018 11:20 (20 minutes)

During the past decade, massive parallel DNA sequencing technologies have completely changed the way in which genetic data are generated and analysed. Instead of sequencing a few hundred nucleotides and focusing on a handful of genes, it is now possible to generate data from entire genomes. The human genome contains about 3 billion nucleotides, and even this is tiny compared to some plant genomes. The resulting datasets are so large that it is impossible to assemble and analyse them without HPC. In this presentation, I demonstrate some of the HPC pipelines my lab uses to assemble genomic data and reconstruct evolutionary relationships between species. These data sets are small compared to what we have planned for the near future, highlighting the necessity to invest more heavily in HPC resources in South Africa and help the country's biologists become internationally competitive. I also focus on the present reluctance of many traditional geneticists to adopt the new technology, and suggest that this bottleneck needs to be closed by finding common ground between biologists, computer scientists and other stakeholders. In the near future, we need to jointly train bioinformatics students who operate at the interface of the different disciplines.

Presenter Biography

Primary author: Prof. TESKE, Peter (University of Johannesburg)

Presenter: Prof. TESKE, Peter (University of Johannesburg)

Session Classification: HPC Applications

Track Classification: Bioinformatics and Biological Sciences