SASBi

Newsletter

Inside the Issue

South African Society for Bioinformatics

UNIVERSITY *of the* **WESTERN CAPE**

SANBI — South African National Bioinformatics Institute

SOUTH AFRICAN SOCIETY FOR BIOINFORMATICS

SASBi is a national association, conscious of its African and international context as it seeks to promote the discipline of bioinformatics in all its facets, locally, regionally and internationally. It is committed to excellence in the learning, teaching and research of the discipline and will strive to explore the applications of bioinformatics and computational biology in addressing the needs of the South African society. The website for SASBi is currently being revamped but more information on SASBi can be obtained at this address: http://sasbi.weebly.com/ . If you are interested in becoming a member please apply online at the link provided.

SOUTH AFRICAN NATIONAL BIOINFORMATICS INSTITUTE

SANBI was founded in 1996 by Winston Hide, a South African computational biologist who returned from the USA. The Institute has been developed within the University of the Western Cape faculty of Natural Sciences. It is legally part of the UWC. In 2000 SANBI submitted a successful proposal for Unit funding from the SA Medical Research Council and established Africa's first UWC/SANBI Bioinformatics capacity development unit.

Our vision is to to become a centre of global, African and South African excellence, achieving the highest levels in biomedical research and education. SANBI aims to conduct cutting edge bioinformatics and computational biology research relevant to South African, African and global populations. To this end, the development of human resources in bioinformatics and computational biology is critically important. SANBI not only adds value to the academic programme of the University of the Western Cape but also focuses on the generation of relevant biomedical research in the interests of the African continent.

Our computational infrastructure is deployed in three ways: a High Performance Cluster, a research cloud hosting a virtual machine and a storage network. We have 232 CPU cores with 1952 GB of RAM on our HPC cluster, OpenStack cloud providing 160 virtual CPU cores, 574GB of RAM and 6.7TB of disk space while our SANReN internet connectivity allows for 10Gb/s internet speed.

Capacity development at SANBI includes: Bioinformatics undergraduate module to 3rd year Biotechnology students, Bioinformatics project included in the Biotechnology Honours programme and a one year internship programme as part of the DST/NRF Research Chair programme. SANBI has for a number of years taken the leading role in training postgraduate students from UWC, UCT, SU and CPUT in Bioinformatics as part of an Introduction to Bioinformatics Course held every year at SANBI for five weeks.

The total number of registered postgraduate students at SANBI is 27 of which 5 is Post doctorate, 11 is PhD and 11 Msc. We have 5 Academic staff, 7 Technical and two admin staff.

Research highlights at SANBI: 1) **B3 Africa/Baobab LIMS**: The Baobab Laboratory Information Management System is a web based application for automating part of the biobank operation. This software development is part of a large programme to develop a network of biobanks in Africa. 2) **Exatype**: In 2017 the use of Exatype HIV drug resistance pipeline was rolled out as part of a spinoff company called Hyrax Biosciences. Diagnostic laboratories in the USA is using Exatype for routine HIV drug resistance testing. 3) **COMBAT-TB**: In 2014 SANBI initiated the COMBAT-TB project in participation with Stellenbosch University, UCT and University of Kwazulu Natal. The project includes a set of workflows for M.tuberculosis data analysis and a graph database for M.tuberculosis genome annotation. 4) **Exome sequencing for clinical research:** The rapid update of high throughput sequencing in clinical research globally requires analytical methods to prioritized functionally relevant genetic variants. One of our projects focuses specifically on analysing large datasets of next generation sequencing reads with a view to identify disease-associated SNPs. Our method published in 2017 has now been applied to a few collaborative projects with the University of Cape Town and Stellenbosch and the resulting publications demonstrate the impact of our computational approach. For example, recently we used exome sequencing to identify a

SANBI staff members from left: Dr Gordon Harkins, Margaret Kumalo, Peter van Heusden, Fungiwe Mpithi, Prof Junaid Gamieldien, Ferial Mullins, Prof Simon Travers, Dr Ruben Cloete, Prof Alan Christoffels, Eugene De Beste, Dr Ravnit Grewal.

BIOINFORMATICS EDUCATIONAL SUMMIT

"My son is in computers", my father used to say. "In computers" was a catch all term of the time that described everything from sales to repairs to programming to systems design. Similarly bioinformatics is growing into a field whose practitioners include clinicians, software engineers, life science researchers and more. Given this diversity of roles, what does a bioinformatics curriculum look like? This is a question that the International Society for Computational Biology (ISCB)'s Education Committee (EC) has been puzzling over for some years.

In 2018 members of the ISCB EC published a resource that described core competencies and model personas of bioinformatics practitioners. The model personas were further subdivided into three types: bioinformatics researchers, users and engineers. In May of this year ISCB, H3ABioNet, GOBLET and ELIXIR organised a four day Bioinformatics Education Summit in Cape Town, South Africa to refine this competencies and personas framework and model its application to bioinformatics curricula and short courses. The revised framework is due for publication in July 2019, in time for the ISMB meeting in Basel, Switzerland.

Existing or new curricula and courses can be mapped to the competencies framework, noting for each course module which competencies the module addresses (for example general biology, or statistical research methods), and whether this is a core focus of the module. Bloom's taxonomy, a hierarchical model that describes the level of complexity of learning objectives, can be used to further describe the level at which certain competencies are addressed. For example, a bioinformatics engineer might need to be able to evaluate the construction of software systems, whereas a life science researcher might only need to be able to apply such software systems.

In addition to helping educators design and refine their courses, the competencies framework will be part of a planned future ISCB programme to endorse curricula and courses. Such endorsements could contribute to the international development of bioinformatics education by making qualifications valid across institutions and countries.

Moving on from the work on the bioinformatics core competencies framework, the Bioinformatics Education Summit considered the development of an online Train-the-Trainer course and the GOBLET bioinformatics trainer resources portal. This portal will contain content on skills required for training, how to develop training materials, how to organise and deliver training, how to assess trainees, evaluate a course and finally how to move towards endorsing and accrediting a course. In addition to the higher education oriented training resources there is a collection of resources on introducing high school learners to bioinformatics.

Finally the Bioschemas markup specifications were introduced. Bioschemas is a development from the schemas.org project that allows websites and online resources to be annotated in a way that computers can understand. In the training context there are draft schemas that describe training materials, courses and course instances. Once created and exported as JSON-LD scripts these can be embedded in web pages. The ELIXIR TeSS web crawler can then read these descriptions and automatically add information on materials and courses to its search portal. While the schemas are still in draft, and the tools for creating Bioschemas JSON-LD are still in their infancy,

Members that attended the Bioinformatics Educational Summit at UCT.

Science Bytes: Focussing on Tuberculosis articles

Next-generation genome annotation: we still struggle to get it right. Steven L. Salzberg, Genome Biology (2019) 20:92

DOI: https://doi.org/10.1186/s13059-019-1715-2

Cryptic Resistance Mutations Associated With Misdiagnoses of Multidrug-Resistant Tuberculosis. Irving Cancino-Muñoz et al. The Journal of Infectious Diseases. Undetected Multidrug-Resistant Tuberculosis • jid 2019:XX (XX XXXX) • 1

DOI: 10.1093/infdis/jiz104

Deciphering Within-Host Microevolution of *Mycobacterium tuberculosis* through Whole-Genome Sequencing: the Phenotypic Impact and Way Forward. S. D. Ley, M. de Vos, A. Van Rie, R. M. Warren. Microbiology and molecular biology reviews

DOI: 10.1128/MMBR.00062-18

Structure based identification of novel inhibitors against ATP synthase of Mycobacterium tuberculosis: A combined in silico and in vitro study. International Journal of Biological Macromolecules 135 (2019) 582–590. Mohd Shahbaaz a, Ruben Cloete a, Melanie Grobbelaar b, Samantha Sampson b, Alan Christoffels a.

https://doi.org/10.1016/j.ijbiomac.2019.05.108

Bioinformatics in the media

How we used computers to figure out drugs that can beat drug-resistant TB. Published January 14, 2019 3.16pm SAST

https://theconversation.com/how-we-used-computers-to-figure-out-drugs-that-can-beat-drug-resistant-tb-108775

Save the Date — Upcoming events

16 -18 September 2019: Infectious Disease 'Omics. London School of Hygiene & Tropical Medicine, UK. Specifically, this course offers hands-on experience of processing sequencing data to construct genomes, identifying genomic variants and applying downstream methods, such as phylogenetics. Further, the course covers transcriptomic and proteomic analysis in human and pathogen settings. High profile examples, including malaria, TB and MRSA, will be used to illustrate the concepts, and there is a strong emphasis on how to implement the methods in practice, with the majority of sessions computer-based. https://www.lshtm.ac.uk/study/courses/short-courses/infectious-disease-omics

6-8 November 2019: Epigenomics of Common Diseases Conference, Wellcome Genome Campus, UK. We are pleased to announce the 8th Epigenomics of Common Diseases conference and invite scientists working on experimental and computational epigenomics and epigenetics to showcase the latest developments in the field and discuss epigenomic approaches in common human diseases. Registration deadline 08 October 2019. **Abstract submission deadline. 10 September 2019.** https://coursesandconferences.wellcomegenomecampus.org/our-events/epigenomics-common-diseases-2019/

1-7 December 2019: Next Generation Sequencing Bioinformatics: Computational course. Wellcome Genome Campus, UK. Application and Bursary deadline 12 September 2019.

https://coursesandconferences.wellcomegenomecampus.org/our-events/ngsbio-2018/

Student Profiles @ SANBI, UWC

Dr Dominique Anderson

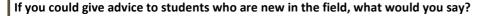Post doctoral fellow: Bioinformatics (SANBI, UWC)

What is your current project:

Baobab LIMS, biobanking Laboratory Information Management System

Name: Dominique Anderson

Fill us in on your career up to this point:

I did my PhD in Metagenomics of extreme Psychrophiles (cold loving microbes) at IMBM at the University of the Western Cape. I was so lucky that my project had a 'bit of everything' and it was so exciting. Following obtaining my PhD, I did a post-doc and then decided to venture into industry. I obtained a Technology Innovation Agency Intern fellowship and was placed at Kapa Biosystems (now Roche Sequencing Solutions). I specialized in recombinant protein expression and purification and this led to me moving to a Biotech Start-up. I returned back to the Academic sphere with an interest in biobanking and informatic tools for biocollections, particularly in low- and middle-income countries. Now, I am an analyst for Baobab LIMS, a software tool built in-house at the South African National Bioinformatics Institute and I am so grateful and proud to be part of an incredible and talented team.

If you could give advice to students who are new in the field, what would you say?

The field may be very intimidating at first, but everything takes time and a bit of practice. Don't be scared to ask for help, have passion for what you do and understand that sometimes you may need to put in some extra time. This field is exciting and there are so many discoveries to be made. The trick is to keep asking questions and seeking opportunity to answer those questions, stay curious, and enjoy learning, every single day no matter how steep the curve may be. Nothing worthwhile is ever easy.

What are your hobbies/activities you do in your free time?
I love to paint, and I spend a fair amount of time gardening. There is something magical about dirt caking your hands and having the patience to wait for something to grow.

If you were an animal – what would you be and why?
Most likely an elephant – somewhat chilled, but can really move when needed, very loyal, protective and determined.

Allison Stander

PhD: Bioinformatics (SANBI, UWC)

What is your current project:
De novo assembly of the draft genome of *aspalathus linearis* (rooibos).

Fill us in on your career up to this point:
BSc: Human Life Sciences, majoring in genetics and physiology. And an honours in genetics. For my honours project, I used quantitative reverse transcription PCR (RT-qPCR) to analyse transcript expression in grapefruit infected with citrus tristeza virus and citrus dwarfing viroid. I attended both my undergrad and honours degrees at Stellenbosch University

If you could give advice to students who are new in the field, what would you say?
Google is your biggest friend. Many of the problems you will face, many others have come across before. People are usually very willing to help and answer questions on science forums.

What are your hobbies/activities you do in your free time?
Origami, watching series, exploring new places, and wine tasting,

If you could change one thing about yourself, what would it be?
Be less easily distracted.

Describe what you were like at age 10?
Very energetic, happy, curious, and social.

Name: Allison Stander

Dr Mohammed Shabaaz

Post Doctoral fellow: Bioinformatics (SANBI, UWC)

What is your current project:

Drug design and discovery against the drug resistant strains of *Mycobacterium tuberculosis*

Fill us in on your career up to this point:

I am working as a post-doctoral fellow under the supervision of Prof Alan Christoffels in South African National Bioinformatics Institute (SANBI), University of Western Cape. I completed my PhD in Chemistry from Durban University of Technology with specialization in Computational Chemistry and Bioinformatics. I published 25 research articles in peer reviewed international journals and two book chapters.

If you could give advice to students who are new in the field, what would you say?

Drug designing against tuberculosis is a very challenging task and require multiple approaches in order to formulate an efficient therapeutic molecules.

Name: Mohammed Shabaaz

Editorial Team— *This issue brought to you by*

Dr Ruben Cloete,

Mr Peter van Heusden

Prof Alan Christoffels