SANBI
South African National
Bioinformatics Institute

ANNUAL REPORT 2017

FROM HOPE TO ACTION THROUGH KNOWLEDGE

RESPICE PROSPICE
UNIVERSITY *of the*
WESTERN CAPE

Contents

About SANBI

WHO WE ARE

Bioinformatics is a specialist discipline straddling the fields of biology, mathematics and computer science and is integral to modern biological research. The South African National Bioinformatics Institute (SANBI) is situated at the University of the Western Cape in Cape Town (UWC). Our primary focus is the development and implementation of computational methodologies which allow researchers to accelerate their genomics data analyses. SANBI aims to heighten awareness of bioinformatics in South Africa and to assist the country in making optimal use of this technology. As the leading bioinformatics entity in Africa, we continue to foster local and regional collaborations on health-related topics that cover both communicable and non-communicable diseases.

SANBI provides a focus for biological research located in Africa and as such, is dedicated to:

- the development of online specialised resources for genomics and genome informatics;
- capacity development in genomics and bioinformatics in Africa; and
- the development and implementation of genome annotation methods.

OUR VISION

To become a center of excellence, achieving the highest level in research and education in the global, African and South African context.

OUR MISSION

- To conduct cutting edge bioinformatics and computational biology research relevant to South African, African and global populations.
- To develop human resources in bioinformatics and computational biology by educating and mentoring scientists.
- To increase awareness of and access to bioinformatics and computational biology resources.

OUR GOALS

- To generate and publish high quality, relevant bioinformatics research.
- To train and graduate competent and productive researchers.
- To add value to the academic programme of UWC.
- To enhance other research fields through collaborative projects.
- To establish sources of renewable funding to pursue the mission of the institute.

POLICY MANDATES

National Strategic Plan for HIV/AIDS, STIs and TB (2017 - 2022)

The vision and mission of SANBI align with the National Strategic Plan (NSP) 2017 – 2022 that outlines how the country will respond to the prevention and treatment of HIV and AIDS, TB and STIs. Specifically the NSP aims to "…strengthen strategic research activities to create validated evidence for innovation, improved efficiency and enhanced impact…".

The Department of Science and Technology's 10-Year Innovation Plan (2008 - 2018)

One of the five Grand Challenge areas specified in this Plan is the "Farmer to Pharma" value chain to strengthen the bioeconomy. SANBI's genomics programme, which straddles both communicable and non-communicable diseases, aligns clearly with this Grand Challenge.

The SAMRC Act (Act 58 of 1991)

As an extramural unit of the SAMRC, SANBI falls under the legislative mandates of the SAMRC. At Section 3, this Act states that the Legislative Mandate of the SAMRC is:
"…through research, development and technology transfer, to promote the improvement of the health and quality of life of the population of the Republic, and to perform such functions as may be assigned to the SAMRC by or under this Act…".

Director's Message

SANBI prides itself as a research-led entity embedded within the University of the Western Cape and more generally in higher education in South Africa. Topics of sustainability and impact have been uppermost in our minds during 2017 and culminated in a new strategic plan for the next five years. Through this process our staff have renewed their commitment to a national and pan-African research agenda where we see ourselves playing an integral role in contributing to the development of a new generation of African computational biology graduates.

In a climate of shrinking financial resources, we have celebrated successful funding applications that included securing funds from a BRICS funded platform, nGAP programme (for young academics), DST/NRF Research Chair renewal and the SANBI MRC Bioinformatics Unit funding. These financial resources underpin our various collaborative projects that allow us to expose our students to inter-disciplinary research.

Expanding our footprint in Africa continued during 2017 as we engaged in bioinformatics training and research programmes in Ghana, Kenya, Sierra Leone and Uganda. Our PhD graduates who have returned to faculty positions in Kenya and Ghana have certainly strengthened our partnerships in these countries.

We continue to promote bioinformatics through active involvement in the African Society for Bioinformatics and Computational Biology (ASBCB) and the Global Emerging Pathogens Treatment Campaign. As of December 2017, we have launched the Improper Scientist, an official magazine of the ASBCB and a medium through which we continue to promote computational biology activities across Africa.

Our social media presence is increasing steadily and we encourage interested individuals to follow our active Twitter accounts namely **@pvanheus** (Peter van Heusden), **@alangchris** (Alan Christoffels), **@BaobabLIMS** (SANBI biobank software), **@DomAnderson83** (Dominique Anderson), **@78034_3507** (Thoba Lose) and **@SANBI_SA** (South African National Bioinformatics Institute).

I acknowledge with thanks the positive contributions from SANBI staff and students alike and I look forward to another year of tangible impact through our research.

Professor Alan Christoffels PhD, M.ASSAf
Director & DST/NRF Research Chair in Bioinformatics and Public Health Genomics
SAMRC Bioinformatics Unit
South African National Bioinformatics Institute
University of the Western Cape

Year in Review

SANBI's presence in the global research community has grown year on year. Recent SANBI alumni have taken up active roles in the private sector and academic spheres. Current active collaborations span across the USA, Europe, Africa and Australia. A noteworthy collaborative project such as the B3Africa consortium has resulted in SANBI leading the development of biobank software (Baobab LIMS) and the training of prospective users in the Ivory Coast, Rwanda, The Gambia and Uganda. During 2017, groups in SA, Russia and China secured a grant to analyse *M.tuberculosis* genomic data across these countries.

RESEARCH HIGHLIGHTS

The impact on the international scientific research community can be seen specifically in the three projects detailed below.

B3Africa/Baobab LIMS

The B3Africa project aims to provide resource-limited biobanks with a set of easy to install open source biobank/bioinformatics tools. The development and implementation of the Baobab LIMS as one of these tools, is led by SANBI in collaboration with the National Health Laboratory Services (NHLS)-Stellenbosch University Biobank (NSB). Baobab LIMS was demonstrated to international audiences throughout 2017.

The SANBI team (Alan Christoffels, Peter van Heusden, Hocine Bendou, Eugene de Beste and Jamie Southgate) was invited by the Global Emerging Pathogens Treatment Consortium (GET) to run a workshop on data management design and LIMS usage in Freetown, Sierra Leone, from 18 - 20 July.

Participants at the Data Management workshop in Freetown, Sierra Leone, July 2017.

The Baobab LIMS was widely showcased at the Global Biobank Week in Stockholm, Sweden, from 12 - 15 September 2017. At the conference, Dominique Anderson delivered the oral presentation entitled "Bioresources Management: Open-Source LIMS including Baobab LIMS".

At the 11th International Conference on Cancer in Africa, organised by the the African Organisation for Research and Training in Cancer (AORTIC), B3Africa partners had a booth and met with African researchers from 7 - 10 November in Kigali, Rwanda. There were approximately 900 participants in total.

Hocine Bendou with Dominique Meunier at the 11th International Conference on Cancer in Africa in Rwanda.

The second B3Africa face-to-face training was held in collaboration with the Biobank and Cohort Building Network (for low-middle income countries) symposium "From Biobank Infrastructure to Research: How BCNet Member Biobanks and Cohorts Are Contributing to Address Public Health Concerns" from 27 November to 1 December in Lyon, France. SANBI was well represented amongst the 58 participants from the 30 organisations that were in attendance. Two talks and training sessions were delivered by Alan Christoffels and Dominique Anderson who were supported by Anja Bedeker, Zipho Mashologu and Campbell Rae.

Exatype

2017 saw significant impact in the use of Exatype HIV drug resistance pipeline through the spinoff company Hyrax Biosciences. One of the largest diagnostic laboratories in the USA is now using Exatype for its routine HIV drug resistance testing. Additionally, the Kenyan Medical Research Institute (KEMRI), the Clinton Health Access Initiative in Kenya and an industry partner have been facilitating the rollout of routine HIV drug resistance testing in Kenya.

Exatype was also expanded to support TB drug susceptibility testing. In parallel with this a complete TB resistance scoring algorithm was developed. In order to undertake validation of this TB solution, partnerships were established with researchers at Stellenbosch University as well as the Critical Path to TB Drug Regimens and partners in the US Center for Diseases Control (CDC).

COMBAT-TB

The COMBAT TB project which was initiated by SANBI in 2014, in participation with Stellenbosch University, UCT and University of KwaZulu-Natal, made significant progress in 2017. The project has developed two main components: a set of workflows for *M. tuberculosis* data analysis that operate on the Galaxy platform and the COMBAT TB Explorer, a graph database of *M. tuberculosis* genome annotation.

Thoba Lose presented this work and its utility in variant characterisation at the Neo4j Life Sciences and Healthcare workshop in Berlin in June 2017.

The combined system was also presented at the Galaxy Community Conference (GCC) in Montpellier, France in June. The COMBAT TB technical team (Thoba Lose, Zipho Mashologu and Peter van Heusden) also participated in the GCC Hackathon, joining the Galaxy developer community in adding code to the Galaxy project.

AWARDS

Alan Christoffels received the UWC Outstanding Alumnus Award in November 2017.

INSTITUTIONAL FUNDING

- The New Generation of Academics Programme (nGAP), funded by the Department of Higher Education and Training, approved the appointment of Ruben Cloete as Lecturer.
- SANBI received a SAMRC Mid-Career Scientist Programme award, which is a research initiative aimed at supporting burgeoning scientists in areas of strategic interest to both the National Department of Health and the SAMRC. This will further expand our profile in strategic applications of informatics and genomics to health R&D, under the mentorship of Junaid Gamieldien.

RESEARCH PROJECTS FUNDING

- The DST/NRF Research Chair in Bioinformatics and Public Health Genomics was renewed for a further 5-year cycle starting in 2018.
- BRICS funding was awarded for TB research which spans a collaboration across China, Russia and South Africa (Alan Christoffels together with Dr Martie van der Walt from the SAMRC TB Platform).
- Gordon Harkins together with his collaborators on the endogenous viral elements (EVE) study received substantial funding from the French National Funding Agency (ANR).

RESEARCH OUTPUTS

A total of 18 peer-reviewed publications appeared in a variety of high quality scientific journals in the past year. A 7-year project culminated in the publishing of the landmark paper on the SA Human Genome Programme while the H3AfricaBioNet project published 6 articles in the Special Section on H3Africa in *Biopreservation and Biobanking*.

CAPACITY DEVELOPMENT

Of the 34 registered postgraduate students, 13 are MSc and 14 are PhD students while 7 are Postdoctoral fellows. 1 MSc and 2 PhD student graduated during the 2017 academic year.

COLLABORATIONS

Alan Christoffels visited Dr John Nkengasong who is the Director of the Africa Center for Diseases Control and Prevention (ACDC) in Addis Ababa, Ethiopia to explore partnerships between SANBI and the Africa CDC.

Alan Christoffels visited the African Malaria Network Developing Excellence in Leadership and Genetics Training for Malaria Elimination in sub-Saharan Africa (DELGEME) Programme in Bamako, Mali.

Alan Christoffels with Dr Djimde and his DELGEME team.

Staff

As a research institute, the Director of SANBI reports through the faculty of Natural Sciences to the University of the Western Cape. SANBI staff are made up of a diverse group of research scientists, technical and administrative staff who all contribute to a dynamic productive working environment. Academic staff are tasked with research, securing funding, student graduations and capacity development. The computing infrastructure is maintained by a team of technical staff while administrators ensure the smooth running of daily operations.

ACADEMIC

ALAN CHRISTOFFELS, PROF
Director & DST/NRF Research Chair in Bioinformatics & Public Health Genomics
DST/NRF Research Chairs Programme
SAMRC Bioinformatics Unit

JUNAID GAMIELDIEN, PROF
Associate Professor
UWC

GORDON HARKINS, DR
Senior Lecturer
UWC

ULJANA HESSE, DR
Senior Lecturer
UWC

SIMON TRAVERS, PROF
Associate Professor
UWC

** Funding source in italics*

TECHNICAL

HOCINE BENDOU
Software Developer
NIH (H3Africa)

QUINTON COERT
Software Developer
EU (B3Africa)

EUGENE DE BESTE
Systems Administrator
SA Medical Research Council

PHILLIP LABUSCHAGNE
Senior Software Developer
HVTN funding from the Bill and Melinda Gates Foundation

THOBA LOSE
Software Developer
SA Medical Research Council

BARUCH LUBINSKY
Software Developer
DST-HIV funding

ZIPHO MASHOLOGU
Software Developer
SA Medical Research Council

CAMPBELL RAE
Web Developer
DST/NRF Research Chairs Programme

PETER VAN HEUSDEN
Senior Systems Developer
UWC

ADMINISTRATION

MARGARET KUMALO
Secretary
DST/NRF Research Chairs Programme

FUNGIWE MPITHI
Receptionist
SA Medical Research Council

FERIAL MULLINS
Finance Administrator
UWC

MARYAM SALIE
Student Administrator
SA Medical Research Council

JUNITA WILLIAMS
Administrator part-time
DST/NRF Research Chairs Programme

Capacity Development

SANBI offers training courses that are recognised as being of the highest calibre. Students are encouraged to participate in workshops and conferences relevant to their research projects or visit collaborators from South Africa or abroad.

Students are expected to publish their work in peer-reviewed journals and to present at laboratory meetings and to wider audiences. The support provided by SANBI exposes students to internationally competitive research environments.

UNDERGRADUATE TRAINING PROGRAMME

Bioinformatics Module (BTN 315)

The UWC undergraduate Bioinformatics Module is taught to approximately 97 third year students by SANBI faculty. Students attended lectures, practical modules and tutorials on the following topics:
Databases and their Applications in Disease Genomics; Introduction to Molecular Evolution and Phylogenetics; Comparative Genomics; and Alignments and Database Searching.

Honours Progamme

Although SANBI does not have an Honours programme, students who attain a pass rate of >60% can include a bioinformatics component to their Honours research project.

Internship Programme

As part of the DST/NRF Research Chair Programme, SANBI provides a 1-year internship programme to students who graduated with a BSc degree. In 2017, student Ayodeji Adebiyi worked on protein-protein interactions and Peter Abiodun worked on information security in the lab of Alan Christoffels.

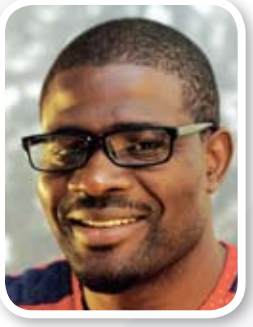
POSTGRADUATE TRAINING PROGRAMME

Masters Programme

SANBI offers a research MSc in Bioinformatics. Candidates with an Honours (BSc) degree or equivalent in a related scientific subject area may apply. The MSc degree is usually completed within two years.

LYNLEY ABDOLL
Alan Christoffels

TOLUWALEKE ADEMUYIWA
Alan Christoffels



OLABODE AJAYI
Alan Christoffels

EUGENE DE BESTE
Alan Christoffels

FANECHKA ESTERHUYSEN
Junaid Gamieldien

ERESHIA GABIER
Simon Travers

WARREN JACOBUS
Alan Christoffels

CLINT MERCUUR
Simon Travers

YAMKELA MGWATYU
Uljana Hesse

BEN ILUNGA MUTEBA
Uljana Hesse

MUHAMMAD SAEED NATHA
Junaid Gamieldien

JAMIE SOUTHGATE
Alan Christoffels

PETER VAN HEUSDEN
Alan Christoffels

**Supervisor in italics*

Doctoral Programme

The PhD programme at SANBI is well-established and alumni of this programme are now working all over South Africa and at bioinformatics research sites around the world. Candidates should be in possession of a Masters degree in Bioinformatics or in a related scientific field subject area such as Computer Science, Mathematics, Biochemistry, Engineering. The PhD degree must be completed within five years although most students aim to complete the degree in three years.

IBRAHIM AHMED
Alan Christoffels

HOCINE BENDOU
Alan Christoffels

TRACEY CALVERT-JOSHUA
Alan Christoffels

SARAH DERAEDT
Alan Christoffels

SOULEYMANE DIALLO
Alan Christoffels

MMAKAMOHELO DIREKO
Alan Christoffels

ROUX-CIL FERREIRA
Simon Travers

PHILIP LABUSCHAGNE
Simon Travers

BRIDGET LANGA
Junaid Gamieldien

ANATI NKAULE
Alan Christoffels

CATHERINE ROSSOUW
Alan Christoffels

EMILY STANDER
Uljana Hesse

EMIL TANOV
Gordon Harkins

LARRY VAN VUUREN
Nicki Tiffin

Postdoctoral Programme

Postdoctoral fellows are admitted to the research programme after consultation with a potential SANBI supervisor. Fellowships may last for a maximum of three years.

IBRAHIM AHMED
Alan Christoffels

DOMINIQUE ANDERSON
Alan Christoffels

RUBEN CLOETE
Alan Christoffels

SAMUEL EGIEYEH
Alan Christoffels

JOHANN JOSEF EICHER
Simon Travers

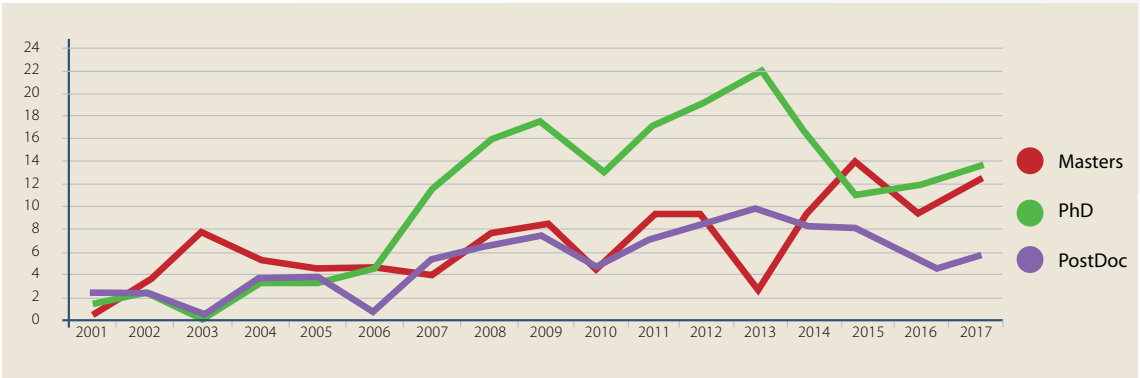
RAJAN SHARMA
Alan Christoffels

MOHD SHAHBAAZ
Alan Christoffels

2017 SANBI Postgraduate Registration

In 2017 the student cohort comprised of 7 Postdoctoral fellows, 13 Doctoral and 14 Masters students.

SANBI Postgraduate Registrations Trend 2001 – 2017:



2017 SANBI Graduations

SANBI has great pleasure in announcing the following graduates for 2017:

IBRAHIM AHMED

PhD

Alan Christoffels

Thesis Title:

Computational prediction of host-pathogen protein interactions.

LARRY VAN VUUREN

PhD

Nicki Triffin

Thesis Title:

Identification of coding variants associated with familial systematic lupus erythematosus through whole exome sequencing.

Workshops and Courses

Numerous capacity development efforts are organised or hosted by SANBI faculty throughout the year in an effort to further develop human resources in the field of bioinformatics and computational biology.

NATIONAL INTRODUCTION TO BIOINFORMATICS COURSE

This annual course runs for approximately 5 weeks from February at SANBI. Introductory modules and hands-on practical modules are covered by lecturers from SANBI and other national universities in the areas of Programming; Molecular Biology and Biological Databases; Statistics using Python and Python Programming; Sequence Analysis; Genomics and Next Generation Sequencing. 20 participants from UWC, UCT, University of Stellenbosch and Cape Peninsula University of Technology (CPUT) were in attendance.

H3AFRICABIONET

SANBI continued to provide training as part of the requirement to support H3Africa research projects in the form of the Workshop on predicting the effects of mutations/nsSNPs on protein structure using Homology Modeling, Gibbs free energy calculation webservers (SDM and mCSM) and Interaction analysis. This workshop was held from 28 – 30 August for a wide range postgraduate audience in Molecular Biology, Chemistry, Pharmacology and related fields. The 22 participants were from the SAMRC, UWC, UCT, University of Stellenbosch and University of the Witwatersrand.

Attendees of the H3AfricaBioNet workshop in August.

B3AFRICA CONSORTIUM

SANBI participated at consortium meetings and training throughout 2017 as detailed below.

Hocine Bendou at the exhibition stand to promote the B3Africa project and Baobab LIMS at the AORTIC meeting in Kigali, Rwanda.

African participants at the BCNet Symposium in Lyon, France.

B3Africa Consortium AGM, in Uppsala Sweden.

African delegation at the Baobab LIMS training workshop at the BCNet Symposium at IARC in Lyon, France.

B3Africa Consortium Working Group in Uppsala, Sweden.

TEACHING

Name	Institution	Date	Course Details
Junaid Gamieldien, Gordon Harkins, Uljana Hesse, Simon Travers	UWC	May 2017	Introduction to Bioinformatics, taught to approximately 54 3rd year Biotechnology Students.
Gordon Harkins	Cape Peninsula University of Technology, Bellville	June 2017	Introductory Statistics for Researchers. Invited to teach a series of lectures to Btech students from the Department of Nature Conservation and Marine Science at CPUT.
Gordon Harkins	CIRAD, Reunion Island	September 2017	Invited to deliver a lecture course and workshop on "the use of methods to estimate the date of integration of endogenous viral elements within plant and animal host genomes."

SEMINARS PRESENTED BY SANBI

Name	Institution	Date	Seminar Details
Alan Christoffels	Woldia University, Ethiopia	April 2017	Presented a seminar entitled: "Is Biology an example of big data science?" to the Deans of Woldia University.
Junaid Gamieldien	British Columbia Cancer Agency, Vancouver, Canada	November 2017	Prioritizing results from high-throughput omics experiments using knowledge graphs and phenotype semantic modeling.

SEMINARS PRESENTED BY VISITING ACADEMICS

Name	Institution	Date	Seminar Details
Dr Dmitry Maslov	Department of Genetics and Biotechnology, Vavilov Institute of General Genetics at the Russian Academy of Sciences	May 2017	"Developing anti-tuberculosis drugs: current approaches and new challenges".
Dr Quinn Wills	Head of Genomics, Novo Nordisk Research Centre, Oxford	May 2017	His talk was aimed at functional genomicists, computational biologists, and liver/diabetes researchers. He provided an overview of the realities being faced as part of his goal to integrate massively parallel gene editing and single-cell methods to model liver metabolism and identify type 2 diabetes drug targets.
Dr Reinhard Hiller	Managing Director, CPGR	July 2017	An informative talk about biotechnology, genomics and omics in general in the South African context.
Prof Akin Abayomi	Head of Department of Haematopathology Tygerberg Academic Hospital	October 2017	"Biosecurity in the 21st century and relevance to Africa".

From Left to Right: Alan Christoffels, Lesibana Malinga (SAMRC), Dmitriy Maslov (Vavilov Institute of General Genetics, Russia), Martie van der Walt (SAMRC) and Valeriy Danilenko (Vavilov Institute).

Prof Akin Abayomi presenting his seminar at SANBI.

Computational Resources

SANBI's IT team supports the work of the institute through software development and by crafting our in-house research cloud, data storage and High Performance Computing (HPC) facilities.

SANBI's core server infrastructure is deployed in three ways: a HPC cluster, a research cloud hosting virtual machines and a storage network. The HPC and cloud are supported by a set of Dell blade servers with a total of 232 CPU cores. In addition to the blade servers a R710 rack mounted server, with 12 CPU cores and 24 GB of RAM, is used as a storage controller. The storage network has two parts, a Dell EqualLogic storage array that hosts 30 TB of SATA disk based storage and a Ceph storage cluster running on Supermicro servers that provide 183 TB of storage. The servers and the storage array are interconnected with 10 Gb Ethernet.

Internet connectivity is provided by UWC and SANBI is connected to SANReN, the South African Research and Communication network that provides 10Gb/s connections between South African universities and research centers.

SANBI HPC CLUSTER

The cluster hardware configuration was not expanded in 2017 and remains on par with the configuration from 2016. We have 232 CPU cores and 1952 GB of RAM on our Dell cluster.

Server Type	Server Count	RAM	CPU cores
M710HD	6	64 GB - 128 GB	12
M910HD	1	512 GB	16
M620	6	96 GB	12
M610X	2	32 GB	12
M630	3	96 GB - 128 GB	16

At the end of 2017 the Sun Grid Engine software that has managed the cluster since its inception was retired. It will be replaced in 2018 by the Slurm (Simple Linux Utility for Resource Management) cluster scheduler. This will allow for better support for parallel computing using MPI (the Message Passing Interface) while keeping SANBI up to date with trends in HPC workload management.

VIRTUAL MACHINE INFRASTRUCTURE

The SANBI VM cloud that our previous systems administrator, Long Yi, has implemented is still currently in use and has had no significant changes made to it. At the end of 2017, SANBI hired a new systems administrator, Eugene de Beste, who looks forward to replacing our VM cloud with a new system in 2018.

SANBI currently hosts 115 virtual machines using 160 virtual CPU cores, 574 GB of RAM and 6.7 TB of disk space, spread across 9 virtual machine hosts. These virtual machines include those used to host key SANBI infrastructure as well as web applications presenting SANBI research to the outside world.

NETWORKING AND INTERNET

The external networking was moved from the previous 30 Mbit/s Neotel business ADSL line to the UWC's internet backbone. This move effectively increases our internet capacity by approximately 20x. SANBI now also has direct access to the 10 Gbit/s SANReN backbone.

STORAGE AND BACKUP

We continued to improve the Ceph storage array that had been planned and implemented since 2014. The capacity of the array has been upgraded from 2016's 72 TB of storage to a pool of 118 TB in 2017. The beginning of this year brought yet another upgrade to the total storage pool of the cluster, which now sits at a total of 183 TB. The scalability of Ceph continues to prove that the switch from a more traditional storage solution has been worth it.

SOFTWARE DEVELOPMENT

The highlights of software development at SANBI in 2017 have been the development and release of the Baobab LIMS and continued work on the COMBAT-TB suite of *M. tuberculosis* research and analysis resources.

BAOBAB LIMS

The Baobab Laboratory Information Management System (LIMS) is a web based application for automating part of a biobank operation. These include detailed inventory management, automated reporting, tracking the movement of biological specimens between facilities and instrument integration. This software development is part of a large programme to develop a network of biobanks in Africa. This network will greatly assist large scale genomic projects requiring a significant amount of heterogeneous biospecimens from diverse backgrounds.

The development and implementation of the Baobab LIMS is led by SANBI in collaboration with the National Health Laboratory Services (NHLS)-Stellenbosch University Biobank (NSB). The Baobab development team acquired a new member at the end of 2017, as Quinton Coert has joined Hocine Bendou in developing the software.

COMBAT TB

The COMBAT TB project was initiated in 2014 and made significant progress in 2017. The project has developed two main components: a set of workflows for *M. tuberculosis* data analysis that operate on the Galaxy platform and the COMBAT TB Explorer, a graph database of *M. tuberculosis* genome annotation.

Underlying the COMBAT TB Explorer database is a Neo4j graph database collecting a wide variety of data about the pathogen's genome and associated knowledge, including protein annotation and genome variants associated with drug resistance phenotypes. Thoba Lose presented this work and its utility in variant characterisation at the Neo4j Life Sciences and Healthcare workshop in Berlin in June 2017.

Neo4j Lifesciences and Healthcare Workshop participants in Berlin. Thoba Lose from SANBI (far right).

The combined system was presented at the Galaxy Community Conference (GCC) in Montpellier, France in June 2017. The COMBAT TB technical team (Thoba Lose, Ziphozakhe Mashologu and Peter van Heusden) also participated in the GCC Hackathon, joining the Galaxy developer community in adding code to the Galaxy project.

MENTORSHIP

In the past SANBI has assisted in the mentorship of students to enter the annual Centre for High Performance Computing national Student Cluster Competition. This involvement has resulted in students from UWC winning the local competition in 2013 and the International Supercomputing Conference Student Cluster Competition in 2014, a first for the university. In 2016, MSc students from SANBI mentored yet another team which achieved first place in the local competition and second place, internationally, in 2017.

At the end of 2017 one of the students from winning team of the 2014 international competition was brought on as the new systems and cloud administrator for SANBI. This highlights the the long term benefit of our mentoring activities.

BIOINFORMATICS AND ASTRONOMY PARTNERSHIP

On 8 August SANBI technical staff hosted a meeting with the Institute for Data Intensive Astronomy and IT staff from UWC and UCT to discuss their cloud computing plans with regards to the African Research Cloud and possible future collaborations.

SANBI in the Media

A press release was issued on the launch of Baobab LIMS by the SAMRC:

SAMRC software provides affordable access to ethical biobanking activities

9 June 2017 Cape Town. The South African Medical Research Council (SAMRC) has developed an open-source laboratory information management system (LIMS) software to give researchers affordable access to human biobanking activities.

"Commercial LIMS alternatives are currently available but do not cater to resource-limited settings where researchers cannot afford licensing fees," says Professor Alan Christoffels, Director at the SAMRC's Bioinformatics Research Unit based at the University of the Western Cape. "This new software is available for free download allowing more researchers across the globe to have access to an essential tool that will facilitate biospecimen management."

See: http://www.mrc.ac.za/Media/2017/21press2017.htm

Work on Ruben Cloete's drug resistant TB study enjoyed coverage on a SAfm evening radio show with presenter Karen Key and a news article appeared on the Healthcare News section of bizcommunity.com.

Researchers use genome data to find MDR TB-fighting compounds

24 August 2017 By its very nature, multidrug-resistant tuberculosis (MDR TB), including extensively drug-resistant tuberculosis (XDR TB), is becoming increasingly difficult to treat, so researchers are using genome data to develop compounds that will combat the disease.

According to Dr Ruben Cloete of the South African National Bioinformatics Institute (Sanbi) at the University of the Western Cape, 10% of infected individuals have the latent form of mycobacterium tuberculosis, which could become active later in life. This makes the discovery of novel compounds very important to reduce the critical reservoir of potential cases of TB.

See: http://www.bizcommunity.com/Article/196/148/166479.html

In October, the CPGR announced its use of Baobab LIMS.

South Africa's Centre for Proteomic & Genomic Research (CPGR) to implement Baobab LIMS to boost lab quality control and operational performance

18 October 2017 The Centre for Proteomic and Genomic Research (CPGR) today announced plans to implement the Baobab LIMS (Laboratory Information Management System), an open-source platform developed with a view to make state-of-the-art tracking of samples, data and lab reagents feasible in resource-constrained settings.

Baobab LIMS was funded by the European Union Horizon2020 programme as part of an integrated approach to building IT infrastructure to bridge biobanking activities between Europe and Africa (www.b3africa.org).

The project, published in the Biobanking and Biopreservation Journal in April 2017, is a culmination of 5 years of work with multinational academic partners in South Africa, Kenya, Nigeria, Uganda, Austria, France and Sweden. The South African effort was led by The South African National Bioinformatics Institute (SANBI) at the University of the Western Cape (UWC).

See: http://www.cpgr.org.za/south-africas-centre-for-proteomic-genomic-research-cpgr-to-implement-baobab-lims-to-boost-lab-quality-control-and-operational-performance/

Research Outputs

SANBI's profile of excellence is reflected in its ability to publish high impact scientific articles in internationally recognised publications. Details of the institute's outputs are described in the tables that follow.

JOURNAL PUBLICATIONS

Noteworthy papers for 2017 include:

- The H3Africa project had 6 articles featured in the April 2017 issue of *Biopreservation and Biobanking* in the Special Section on H3Africa.
- In December 2017 the landmark paper on the South African Human Genome Program was finally published after 7 years, with contribution from the labs of Alan Christoffels and Junaid Gamieldien.

#	Publication (SANBI contributors)	Published Date	Impact Factor
1.	*A practical guide to filtering and prioritizing genetic variants.* **Mahjoubeh Jalali, Junaid Gamieldien.** January 2017 · BioTechniques 01/2017; 62(1) DOI:10.2144/000114492	January 2017	2.030
2.	*Virome Assembly and Annotation: A Surprise in the Namib Desert.* **Uljana Hesse, Peter van Heusden**, Bronwyn Michelle Kirby, Israel Olonade, Marla Tuffin. January 2017 Frontiers in Microbiology 8 DOI: 10.3389/fmicb.2017.00013	January 2017	4.076
3.	*Metagenomics-based discovery and molecular characterization of novel geminiviruses from two Mediterranean-climate biodiversity hotspots.* Galzi Serge, Bernardo Pauline, Fernandez Emmanuel, Julian Charlotte, Ferdinand Romain, **Harkins Gordon William**, Peterschmitt Michel, Martin Darren Patrick, Filloux Denis, Roumagnac Philippe. 16th Meeting of Plant Virology. 15 - 19 January 2017. Aussois, France	January 2017	
4.	*The role of Kenya in the trans-African spread of maize streak virus strain A.* Daniel Pande, **Eugene Madzokere**, Penelope Hartnady, Simona Kraberger, James Hadfield, Karyna Rosario, Anja Jäschke, Adérito L Monjane, Betty E Owor, Mathews M Dida, Dionne N Shepherd, Darren P Martin, Arvind Varsani, **Gordon W Harkins**. Feb 2017 · Virus Research 232:69-76 DOI: 10.1016/j.virusres.2017.02.005	February 2017	2.628
5.	*Accessing Biospecimens from the H3Africa Consortium.* Christine M. Beiswanger, Alash'le Abimiku, Nadia Carstens, **Alan Christoffels**, Jantina de Vries, Audrey Duncanson, Morne du Plessis, Maria Giovanni, Katherine Littler, Nicola Mulder, Jennifer Troyer, Louise Wideroff and H3Africa Data and Biospecimen Sharing Organizing Committee. April 2017 · Biopreservation and Biobanking 04/2017; 15(2). DOI:10.1089/bio.2017.0008	April 2017	1.698

#	Publication (SANBI contributors)	Published Date	Impact Factor
6.	*Baobab Laboratory Information Management System: Development of an Open-Source Laboratory Information Management System for Biobanking.* **Hocine Bendou, Lunga Sizani**, Tim Reid, Carmen Swanepoel, **Toluwaleke Ademuyiwa**, Roxana Merino-Martinez, Heimo Meuller, Akin Abayomi and **Alan Christoffels**. Biopreservation and Biobanking Volume 15, Number 2, 2017 DOI: 10.1089/bio.2017.0014	April 2017	1.698
7.	*Selecting a Laboratory Information Management System for Biorepositories in Low- and Middle-Income Countries: The H3Africa Experience and Lessons Learned.* Kyobe Samuel, Musinguzi Henry, Lwanga Newton, Kezimbira Dafala, Kigozi Edgar, Katabazi Fred Ashaba, Wayengera Misaki, Joloba Moses Lutaakome, Abayomi Emmanuel Akin, Swanepoel Carmen, Abimiku Alash'le, Croxton Talishiea, Ozumba Petronilla, Thankgod Anazodo, **Christoffels Alan**, van Zyl Lizelle, Mayne Elizabeth Sarah, Kader Mukthar, Swartz Garth, and H3Africa Biorepository PI Working Group. Biopreservation and Biobanking. April 2017, 15(2): 111-115. https://doi.org/10.1089/bio.2017.0006	April 2017	1.698
8.	*Delineating Transcriptomic Profiles in PTSD: An RNAseq Investigation.* Sian Hemmings, Laetitia Dicks, **Mahjoubeh Jalali, Junaid Gamieldien**, Soraya Seedat. May 2017 Biological Psychiatry 81 (10), S222-S223 DOI:10.1016/j.biopsych.2017.02.1158	May 2017	11.412
9.	*Structural and functional effects of nucleotide variation on the human TB drug metabolizing enzyme arylamine N-acetyltransferase 1.* **Ruben Cloete, Wisdom A. Akurugu**, Cedric J. Werely, Paul D. van Helden, **Alan Christoffels**. Journal of Molecular Graphics and Modelling 75 (2017) 330–339	June 2017	1.754
10.	*Exome sequencing identifies targets in the treatment-resistant ophthalmoplegic subphenotype of myasthenia gravis.* Melissa Nel, **Mahjoubeh Jalali Sefid Dashti, Junaid Gamieldien**, Jeannine M. Heckmann. Neuromuscular Disorders 27 (2017) 816–825	June 2017	2.969
11.	*The role of microRNAs in the therapeutic action of D-cycloserine in a post-traumatic stress disorder animal model: an exploratory study.* Malan-Müller, Stefanie; Fairbairn, Lorren; Hart, Stephanie; Daniels, Willie M.U.; **Jalali Sefid Dashti, Mahjoubeh**; Kidd, Martin; Seedat, Soraya; **Gamieldien, Junaid**; Hemmings, Sîan M.J. Psychiatric Genetics 2017 Aug;27(4):139-151. doi: 10.1097/YPG.0000000000000176.	August 2017	1.557
12.	*The Transcription Factor 7-Like 2–Peroxisome Proliferator-Activated Receptor Gamma Coactivator-1 Alpha Axis Connects Mitochondrial Biogenesis and Metabolic Shift with Stem Cell Commitment to Hepatic Differentiation.* Ana'is Wanet, Marino Caruso, **Jean-Baka Domelevo Entfellner**, Mehdi Najar, Antoine Fattaccioli, Catherine Demazy, Jonathan Evraerts, Hoda El-Kehdy, Guillaume Pourcher, Etienne Sokal, Thierry Arnould, **Nicki Tiffin**, Mustapha Najimi, Patricia Renard. Stem Cells 2017 AlphaMed Press 1066-5099/2017/$30.00/0 http://dx.doi.org/10.1002/stem.2688	August 2017	5.599

#	Publication (SANBI contributors)	Published Date	Impact Factor
13.	*Exploring the potential of T7 bacteriophage protein Gp2 as a novel inhibitor of mycobacterial RNA polymerase.* J du Plessis, **R.Cloete**, L.Burchell, P.Sarkar, R.M.Warren, **A.Christoffels**, S Wigneshweraraj, S.L.Sampson. Tuberculosis, Volume 106, September 2017 Pages 82 – 90.	September 2017	2.873
14.	*The interaction of polymorphisms in extracellular matrix genes and underlying miRNA motifs that modulate susceptibility to anterior cruciate ligament rupture.* Kyle Willard, Sasha Mannion, **Colleen J. Saunders**, Malcolm Collins, Alison V. September. Journal of Science and Medicine in Sport DOI: http://dx.doi.org/10.1016/j.jsams.2017.08.017	September 2017	3.857
15.	*Designing a course model for distance-based online bioinformatics training in Africa: The H3ABioNet experience.* Kim T. Gurwitz , Shaun Aron , Sumir Panji, Suresh Maslamoney, Pedro L. Fernandes, David P. Judge, Amel Ghouila, **Jean-Baka Domelevo Entfellner**, Fatma Z. Guerfali, **Colleen Saunders**, Ahmed Mansour Alzohairy, Samson P. Salifu, Rehab Ahmed, **Ruben Cloete**, Jonathan Kayondo, Deogratius Ssemwanga, Nicola Mulder, H3ABioNet Consortium's Education Training and Working Group as members of the H3Africa Consortium. October 2017. PLoS Comput Biol 13(10): e1005715. https://doi.org/10.1371/journal.pcbi.1005715	October 2017	4.829
16.	*Geometagenomics illuminates the impact of agriculture on the distribution and prevalence of plant viruses at the ecosystem scale.* Pauline Bernardo, Tristan Charles-Dominique, Mohamed Barakat, Philippe Ortet, Emmanuel Fernandez, Denis Filloux, Penelope Hartnady, Tony A Rebelo, Stephen R Cousins, François Mesleard, Damien Cohez, Nicole Yavercovski, Arvind Varsani, **Gordon W Harkins**, Michel Peterschmitt, Carolyn M Malmstrom, Darren P Martin and Philippe Roumagnac. October 2017. The ISME Journal (2017), 1–12	October 2017	9.664
17.	*Whole-genome sequencing for an enhanced understanding of genetic variation among South Africans.* Ananyo Choudhury, Michèle Ramsay, Scott Hazelhurst, Shaun Aron, Soraya Bardien, Gerrit Botha, Emile R. Chimusa, **Alan Christoffels, Junaid Gamieldien, Mahjoubeh J. Sefid-Dashti**, Fourie Joubert, Ayton Meintjes, Nicola Mulder, Raj Ramesar, Jasper Rees, Kathrine Scholtz, Dhriti Sengupta, Himla Soodyall, Philip Venter, Louise Warnich, Michael S. Pepper. December 2017. Nature Communications. DOI: 10.1038/s41467-017-00663-9	December 2017	12.124
18.	*An integrated and comparative approach towards identification, characterization and functional annotation of candidate genes for drought tolerance in sorghum (Sorghum bicolor (L.) Moench).* **Abdi Woldesemayat, Adugna, van Heusden, Peter**, Ndimba, Bongani, **Christoffels, Alan**. December 2017. BMC Genetics. 18. DOI 10.1186/s12863-017-0584-5	December 2017	2.266

CONFERENCE PARTICIPATION

Presenter	Month	Conference Details	Title	Oral/Poster
Alan Christoffels	February	International Centre of Insect Physiology and Ecology, Kenya	COMBAT-TB: developing analytical tools to fight Tuberculosis.	Oral
Alan Christoffels	February	International Livestock Research Institute, Kenya	COMBAT-TB: developing analytical tools to fight Tuberculosis.	Oral
Emily Stander	May	Research Day at Rooibos Ltd Clanwilliam	Rooibos genomics.	Oral
Yamkela Mgwatyu	May	Research Day at Rooibos Ltd, Clanwilliam	Thin Layer Chromatography method to visualize aspalathin.	Oral
Thoba Lose	June	Neo4j Life Sciences and Healthcare Workshop, Berlin, Germany.	Prioritizing SNPs using the Neo4j Galaxy Interactive Environment.	Oral
Peter van Heusden	June	Galaxy Community Conference 2017, Montpellier, France	COMBAT TB: an integrated environment for Tuberculosis data analysis and surveillance.	Oral
Emily Stander	August	UWC Research Open Day	Simple and fast detection of aspalathin in rooibos plant and tea extracts.	Poster
Yamkela Mgwatyu	August	UWC Research Open Day	Genome Size Estimation of Rooibos.	Poster
Eugene de Beste	August	UWC Research Open Day	Enabling the processing of bioinformatic workflows where data is located through the use of cloud and software container technologies.	Oral
Roux-Cil Ferreira	August	UWC Research Open Day	Using molecular dynamics to illustrate the changes in the glycan shields of two HIV-1 envelope trimmers after the loss of a glycan.	Oral
Alan Christoffels	August	3rd African Conference on Emerging Infectious Diseases and Biosecurity. Accra, Ghana.	Africa's infrastructure readiness to handle data analytics.	Oral
Dominique Anderson	September	Global Biobank Week, Sweden	Bioresource management; open-source LIMS including Baobab LIMS.	Oral

Presenter	Month	Conference Details	Title	Oral/Poster
Ruben Cloete	October	ISCB-Africa ASBCB Conference on Bioinformatics 2017, Entebbe, Uganda	Structural and functional effects of nucleotide variation on the human drug metabolizing enzyme arylamine N-acetyltransferase.	Oral
Eugene de Beste	October	ISCB-Africa ASBCB Conference on Bioinformatics 2017, Entebbe, Uganda	Enabling the processing of bioinformatic workflows where data is located through the use of cloud and container technologies.	Oral
Eugene de Beste	October	ISCB-Africa ASBCB Conference on Bioinformatics 2017, Entebbe, Uganda	Moving Code to Data.	Poster
Jamie Southgate	October	ISCB-Africa ASBCB Conference on Bioinformatics 2017, Entebbe, Uganda	Establishing the framework for an African Genome Archive.	Poster
Peter Van Heusden	October	ISCB-Africa ASBCB Conference on Bioinformatics 2017, Entebbe, Uganda	COMBAT TB, an integrated environment for TB sequence data storage, analysis and visualisation.	Oral
Mohd Shahbaaz	October	ISCB-Africa ASBCB Conference on Bioinformatics 2017, Entebbe, Uganda	Structural insights into the inhibitory mechanism of Mycobacterium tuberculosis cytosine monophosphate kinase: An in silico study.	Poster
Samuel Egieyeh	October	ISCB-Africa ASBCB Conference on Bioinformatics 2017, Entebbe, Uganda	Chemoinformatics and in-silico drug development workflow: Enhancing "hits" to "drug candidates" optimization in drug discovery.	Oral
Anati Nkaule	October	ISCB-Africa ASBCB Conference on Bioinformatics 2017, Entebbe, Uganda	Designing new kinase inhibitor derivatives as therapeutics against Mycobacterium tuberculosis: A docking and molecular dynamics simulation study.	Poster

JOURNAL EDITING AND REVIEWS

PI Name	Journal
Alan Christoffels	**Editorial Board Member:** Data Journal **Reviewer:** BMC Structural Biology, Experimental and Molecular Pathology, Global Health Epidemiology Genomics.
Junaid Gamieldien	**Reviewer:** Nature Scientific Reports, PLoSONE, Molecular Genetics and Genomic Medicine, Clinical Genetics, Computer Methods and Programs in Biomedicine, BMC Bioinformatics
Gordon Harkins	**Reviewer:** Diversity and Distributions
Uljana Hesse	**Reviewer:** Environmental Microbiology

EXPERT PANEL OR COMMITTEE MEMBERSHIP

PI Name	Journal
Alan Christoffels	**Governing Council and Co-Chair:** ISCB Africa ASBCB Conference on Bioinformatics 2017, Entebbe, Uganda.
Simon Travers	**Expert Panel Member:** Critical Path to TB Drug Regimens Workshop, Washington DC, USA.

Research Laboratories

RESEARCH LABORATORY OF ALAN CHRISTOFFELS

Members of the Christoffels lab.

HIGHLIGHTS OF THE LAB FOR 2017

During 2017 our efforts were directed at strengthening our analytical capability to respond to communicable diseases. This work, originally funded in 2013 has now matured to the point where we have made significant advances in (1) Analytical tools to support biospecimen information management particulary with reference to centralised biobanks and (2) a platform to analyse genomics sequence data in real-time.

1. Baobab Laboratory Information Management System (LIMS)

There has been overwhelming interest internationally for our LIMS especially after we presented our work at the Global Biobank Week in Stockholm. This exposure coupled with our contribution to the B3Africa consortium led to a range of training workshops for potential users of our LIMS, including:

- Biobanks in low and middle income countries – we hosted a training session in Lyon, France. Through this event we secured users for our software from research teams in Egypt, Uganda, Indonesia, Gambia, Senegal, Cote d'Ivoire and The Gambia MRC unit.
- Center for Proteomics and Genomics Research (CPGR) announced in October 2017 that they will be using Baobab LIMS for their laboratories.
- User testing in our collaborator's laboratory at Heamatology Pathology Tygerberg Medical School.
- At least three manuscripts in this space (i.e. biobank informatics).
- Started our LIMS twitter account (@BaobabLIMS).

2. M.tuberculosis research

Publications relating to our drug discovery efforts:

- Proteins that could be potential inhibitors of key enzymes in mycobacteria.
- Predicting the effect of genetic variation on enzyme function in humans. The consequence of this is that we need to think carefully of patient-specific treatment.

3. Southern African Human Genome Project

It was rewarding to see a national project that started in 2011 reap its rewards when we published our findings on the analysis of South African population groups. This work provides a genomics resource (data) that will be the impetus for many specialised genetics studies when researchers investigate the consequence of genetic variation on human health.

4. Graduations

We celebrated two student graduations in 2017, namely:

- Ibrahim Ahmed , PhD
- Antoinette Colic, MSc registered at Stellenbosch University but co-supervised by me and Prof Samantha Sampson. This project represented a joint effort by two SARChi holders.

5. BRICS funding

For the past two years our lab was part of a TBResist consortium spanning teams in Russia, SA, China and USA. During 2017, groups in SA, Russia and China secured a grant to analyse *M.tuberculosis* genomic data across these countries.

6. African Society of Bioinformatics and Computional Biology (ASBCB)

We were part of the organising committee of the regional bioinformatics conference held in Uganda in October 2017. At least 5 members of the lab attended and presented at this conference.

7. Improper Scientist

In December 2017, we launched the quarterly magazine of the ASBCB called the *Improper Scientist*. My vision for this magazine is to highlight the wealth of omics research underway in Africa and to strengthen the bioinformatics community.

RESEARCH PROJECTS

The projects below underpin our translational work:

Tuberculosis

A comprehensive research programme is underway that includes investigating genetic determinants in both host (Human) and pathogen (*Mycobacterium tuberculosis*) to understand drug resistance, and protein structure determination to inform patient-centric drug design. These findings inform the development of a scientific workflow management system to support reproducible high throughput computational experiments. These workflows form the basis of our COMputational BActerial analytical toolkit for Tuberculosis research (COMBAT-TB) (combattb.sanbi.ac.za) initiative to deploy analytic tools across the African continent. A new BRICS grant was secured with teams in China, South Africa and Russia to explore genetic differences in strains of *M.tuberculosis*.

Search for new anti-malarials

The medicinal benefits of plant extracts generates a plethora of data. In collaboration with the UWC School of Pharmacy, we are exploring the use of computational tools to identify signatures in common between registered anti-malarial drugs and plant extracts that show anti-plasmodial activity.

Infrastructure for biobanking

The collection and storage of biological material and its associated data are essential to any genetics project. To this end we have contributed to the the development of a computational environment to manage and analyse genomic data that is linked to a biobank. Specifically, we have developed an open-source laboratory information management system – Baobab LIMS (baobablims.org) that is being rolled out across Africa.

Genome Annotation

We continue to extend our genome annotation footprint through our participation in international projects to define the gene repertoire in newly sequenced genomes. We have initiated a new project with a Trypanosome consortium with partners in Kenya and the UK.

RESEARCH COLLABORATIONS

1. Prediction of human-tuberculosis interaction networks

Collaborating Parties:

Prof Peter Witbooi - Department of Mathematics, UWC.

Nature and Purpose:

Use machine-learning techniques to improve the prediction of host-pathogen interactions between human mycobacteria.

Output in the last 12 months:

A PhD student graduated and a manuscript submitted.

2. Develop a biobank informatics management system

Collaborating Parties:

B3Africa Consortium funded through Horizon2020 (b3africa.org).

Nature and Purpose:

Development of a laboratory management system that will integrate with other biobanking tools through an open-source platform called eB3Kit (demo.bibbox.org).

Output in the last 12 months:

We extended our user-base in multiple African countries.

3. Chemosensory genes in Tsetse

Collaborating Parties:

Dr Dan Masiga and Dr Merid Getahun - ICIPE, Kenya.

Nature and Purpose:

To annotate chemosensory genes in tsetse species and to identify chemicals that bind to these chemosensory receptors.

Output in the last 12 months:

1 PhD student registered.

4. Identification of novel drug targets for drug resistant tuberculosis

Collaborating Parties:

Prof Samantha Sampson, Prof Rob Warren and Dr Cedric Wereley – University of Stellenbosch.

Nature and Purpose:

Current TB drugs are more than 30 years old and have unacceptable efficacy and safety profiles, emphasising the need for new drugs. We mapped drug resistance genes, derived from comparative genome analysis of three *M. tuberculosis* strains (susceptible, multidrug and extensively drug resistant), and from published literature to metabolic pathways and identified nine potential drug target candidate genes. These genes

were ranked for further computational analyses in the quest to identify inhibitors that could be lead compounds.

Output in the last 12 months:

Identified 2 compounds through in-silico screens that bind to drug targets. These compounds are undergoing *in-vitro* testing in the laboratory. Continued screening of *M.tuberculosis* genomes for genes that are implicated in drug resistance.

5. Chemo-informatics profiling of plant extracts that show anti-plasmodial activity

Collaborating Parties:

Profs Sarel Malan and James Syce - School of Pharmacy, UWC

Nature and Purpose:

At least 1000 compounds were identified in plant extracts and show anti-plasmodial activity. We are using a range of computational tools to compare these compounds with currently registered anti-malarial drugs to identify common signatures and ultimately prioritise the list of compounds identified in plant extracts.

Output in the last 12 months:

2 manuscripts submitted. The PhD graduate holds a faculty position and has started his own research team.

6. Computational Bacterial analytical toolkit for Tuberculosis (COMBAT-TB)

Collaborating Parties:

Prof Samantha Sampson and Prof Rob Warren – University of Stellenbosch
Profs Manormoney Pillay and Balla Pillay - UKZN
Prof Kevin Naidoo - UCT

Nature and Purpose:

Develop a computational platform to store Tuberculosis omic data and to provide a visualisation tool.

Output in the last 12 months:

Thoba Lose presented our graph database in Berlin to demonstrate the power of using a no SQL database for the scale and diverse nature of the tuberculosis omics data.
Peter Van Heusden presented the Galaxy platform that we are using for our analytics, at the Galaxy meeting in France.

7. Exploiting protein signatures in Colon Cancer archival biospecimens

Collaborating Parties:

Professor Schneider and Dr Johnathan Rigby – University of Stellenbosch

Nature and Purpose:

There are two phases to this project. The first is to assess the stability of the formalin embedded tumor blocks that have been archived for 1-10 years. Our assessment uses proteomic signatures to verify the biological stability of the material. Once this has been completed then we will stratify a cohort of patients to screen for biomarkers.

Output for the last 12 months

Catherine Rossouw (PhD student) started protein extractions from these tumors to optimise the experimental protocol.

8. Dental Genetics Programme

Collaborating Parties:

Prof Manogari Chetty and Dr Tina Roberts - Dentistry Faculty, UWC

Nature and Purpose:

We aim to utilise bioinformatics tools to analyze the genetic basis of some of the syndromes observed in the dental clinics.

Output for the last 12 months:

We registered our first project and recruited a MSc student.

RESEARCH LABORATORY OF JUNAID GAMIELDIEN

HIGHLIGHTS OF THE LAB FOR 2017:

- Exome studies identified likely oligenic mehanisms in cases of ALS and limb-girdle muscular dystrophy.
- Provisional exome-based diagnoses of MODY-diabetes were made using bespoke tools.
- Two lab members were authors on the South African Human Genome Programme landmark paper.
- Prof Gamieldien visited the British Columbia Cancer Agency in November 2017 and presented his work on semantic databases.
- Prof Gamieldien established a PTSD semantic analysis collaboration with ex-postdoc Mahjoubeh Jalali at the Pacific Northwest Research Institute and Dr Kai Wang at the Institute for Systems Biology.

Prof Junaid Gamieldien

RESEARCH PROJECTS

1. Cognitive computing for disease genomics

Our BioOntological Graph Relationship Database (BORG) database assimilates and integrates multiple biomedical ontologies and disparate sources of genomic and biomedical knowledge and metadata to model and learn the phenotypic features and molecular mechanisms associated with diseases, which it utilizes to automatically prioritize novel candidates identified in genomics studies. In essence, the system uses search paradigms that simulate the cognitive processes applied by clinical geneticists and biomedical researchers. In practice, it performs a directed walk on a stored biomedical knowledge graph, which enables the identification of non-obvious yet biologically plausible links that may explain a gene's potential role in the clinical phenotype based relevant prior annotations, or transitively via the functions of proteins that the gene's expressed product physically interacts with. This has proven especially useful in our disease variant and transcriptomics collaborative studies, where unobvious yet biologically and biomedically plausible links have been prioritized and progressed into for further studies.

2. Cloud-based analytical pipelines for NGS

The cost of ownership of specialised hardware required for very large NGS projects is very high, and scaling up infrastructure is not sustainable in the current era of rapidly decreasing data cost. We have thus implemented cloud based exome, genome and RNAseq data processing platform that are effectively infinitely scalable and can reduce the dreaded data processing bottleneck on qualifying studies from several months to weeks or even days.

3. Development of SOPs for variant prioritization

While workflows and associated software to process the raw data and to produce high-confidence variant calls have significantly improved, filtering the tens of thousands of candidates to produce a subset relevant to the study at hand is still complex. We are thus continually developing rules and SOPs for identifying potentially functional variants from whole genome and exome variant calls and prioritizing those potentially associated with a phenotype of interest. Used in conjunction with our cognitive analytics approach, we have been able to identify strong candidates for further analysis in several in-house and collaborative disease studies.

4. Development of an exome based genetic diagnosis framework for monogenic and atypical diabetes

Among the monogenic forms of diabetes mellitus (DM), Mature Onset Diabetes of the Young (MODY), which has an age of onset of less than 25 years, has been the most intensively investigated in recent years. MODY is a group of clinically heterogeneous, often non-insulin- dependent forms of DM that are defined at the molecular genetic level by deleterious mutations in different genes. As it is estimated that up to 80% of MODY cases go undiagnosed and therefore sub-optimally treated, we have developed a whole exome sequencing (WES) based monogenic diabetes diagnostic workflow as part of our SAMRC Strategic Health Innovation Partnerships grant. Early indications are that the diagnostic yield of the approach is relatively high for South African patients with atypical diabetes.

5. WES-driven clinical research sequencing projects

We have several ongoing disease/risk variant discovery projects that benefit from customized versions of our WES pipeline and semantic discovery process:

- Muscular dystrophy
- Tendinopathy
- Motor neuron disease/ALS

RESEARCH COLLABORATIONS

1. Exome sequencing of sporadic flail arm ALS

Collaborating Parties:

Dr Mahjoubeh Jalali – Pacific North West Research Institute, Seattle, USA
Prof Jeanine Heckmann – University of Cape Town

Nature and Purpose:

To identify causative variants in two unrelated but similar cases of atypical flail arm amyotrophic lateral sclerosis.

Output in the last 12 months:

Participants have been exome sequenced and using our semantic model of motor neuron disease, strong candidate causative variants have been identified and Sanger confirmed in an affected-child- unaffected-parents trio study. A manuscript reporting strong evidence of distinct oligogenic inheritance patterns of the disease in each case is currently in review.

Future Direction:

We will search for opportunities to apply the lessons learned and technology developed to other ALS cases, most likely using data from sporadic cases that have not yielded any candidates and that have made data available in public repositories.

2. Exome sequencing in a case of limb-girdle muscular dystrophy (LGMD)

Collaborating Parties:

Dr Mahjoubeh Jalali – Pacific North West Research Institute, Seattle, USA
Prof Jeanine Heckmann – University of Cape Town

Nature and Purpose:

To use whole genome sequencing to identify candidate causative variants in a seemingly atypical case of LGMD.

Output in the last 12 months:

We have applied our filtering approach to a) rule out calpainopathy and b) identify several strong candidate mutations in multiple genes, which suggests that the inheritance mechanism is complex and oligogenic.

Future Direction:
Candidate variants will be confirmed and further filtered by Sanger sequencing in the proband, parents and unaffected sibling.

3. Exome sequencing of myasthenia gravis-related ophthalmoparesis

Collaborating Parties:
Dr Mahjoubeh Jalali – Pacific North West Research Institute, Seattle, USA
Prof Jeanine Heckmann (PI), Dr Melissa Nel - University of Cape Town

Nature and Purpose:
To identify the genetic cause of myasthenia gravis-related ocular muscle complications seen only in African patients.

Output in the last 12 months:
A manuscript reporting findings has been published.

Future Direction:
None yet planned.

4. Exome sequencing of atypical diabetes

Collaborating Parties:
Dr Mahjoubeh Jalali – Pacific North West Research Institute, Seattle, USA
Prof Alison September (co-PI) and Prof Naomi Levitt - University of Cape Town

Nature and Purpose:
To test a clinical exome-sequencing pipeline developed as part of an SAMRC SHIP funded project by evaluating its ability to identify causative mutations in patients suspected of having mature onset diabetes of the young (MODY) or ketosis prone diabetes.

Output in the last 12 months:
Cases of possible MODY have been exome sequenced and likely pathogenic mutations have been identified in several patients.

Future Direction:
Confirmation of mutations through Sanger sequencing and prediction of MODY class based therapeutic interventions. We will also search for unsolved cases with data available in public repositories and reanalyse using our filtering and prioritisation tools and strategies.

5. Identification of genetic biomarkers for risk for musculoskeletal soft-tissue injuries (MSI)

Collaborating Parties:
Dr Colleen Saunders, Prof Alison September and Prof Malcolm Collins - University of Cape Town

Nature and Purpose:
To identify candidate Achilles tendinopathy susceptibility genes and variants by comparing whole exome sequences of individuals with severe tendinopathy phenotypes and healthy physically active controls.

Output in the last 12 months:
A cohort of suitably matched cases and controls at the extreme ends of the phenotypic spectrum has been exome-sequenced and candidate predisposing variants identified. Selected candidates have been genotyped in a larger cohort and have shown positive association with the phenotype.

Future Direction:
Several manuscripts in preparation. Genotyping additional candidates.

6. Identification of signatures of circulating miRNAs predictive of smoke inhalation injury in burn victims

Collaborating Parties:

Dr Kareemah Gamieldien (PI) - Cape Peninsula University of Technology
Dr Wayne Kleintjies - Head of the Tygerberg Hospital Burns Unit

Nature and Purpose:

To identify microRNAs in blood that may be used in the early diagnosis of smoke inhalation related lung injury in burn victims.

Output in the last 12 months:

Sample collection has been completed.

Future Direction:

Small RNAs will be isolated from blood of patients meeting the clinical criteria and will be used for miRNA profiling.

7. Identification of signatures for prediction of treatment-linked survival of cancer patients

Collaborating Parties:

Dr Azeez Fatai – Senior Lecturer, Lagos State University, Nigeria

Nature and Purpose:

To build reusable machine learning tools for identifying predictive and prognostic signatures in omics data generated from cancer patients.

Output in the last 12 months:

We identified and assessed the prognostic value of a 35-gene expression signature selected by pathway and machine learning based methods in adjuvant therapy-linked glioblastoma multiforme (GBM) patients from the Cancer Genome Atlas (TCGA). Survival analysis on an entirely independent study, showed that the 35-gene signature reproducibly discriminated between the survival groups in all cases and could accurately predict survival irrespective of the subtype. In a multivariate analysis, the signature predicted progression-free and overall survival independently of other factors considered. A manuscript was submitted and is in final review at *BMC Cancer*.

Future Direction:

We will continue to refine the approach and apply it to available data from other cancers in the public repositories.

RESEARCH LABORATORY OF GORDON HARKINS

Dr Gordan Harkins

My research primarily focuses on the evolution and molecular epidemiology of single-stranded DNA and RNA viral pathogens of animals and plants. My research seeks to determine the evolutionary underpinnings of the emergence and spread of the numerous viral diseases that seriously threaten the health and food security of Africa and the rest of the developing world.

HIGHLIGHTS OF THE LAB:

- Two publications obtained.
- One PhD submitted for examination in 2017.
- Member of of an international collaboration that was awarded substantial funding from the French National Funding agency, ANR.
- Funding received from the NRF/RISA to sequence "ancient" viral genomes present in fynbos plant specimens stored within French and South African herbaria to more accurately estimate when and where, key ancestral plant viruses existed.
- Awarded funding from the UWC Senate Research Committee for the investigation of the spatio-temporal dynamics and evolution of novel geminiviruses within uncultivated plants distributed across the Western and Southern Cape regions of South Africa.

A summary of some of the research projects that members of my research group have been involved in 2017 is presented below.

PLANT-INFECTING VIRAL PATHOGENS

Virulence Evolution

We have received funding from the National Research Foundation (NRF) to investigate how the traits of virulence and transmission rates are evolving in viral pathogens important to human health. Using a combination of laboratory experimentation and state-of-the-art computational analyses, we are investigating the spatiotemporal dynamics and evolution of a broad range of viral pathogens focusing primarily on viruses that cause important crop diseases (Maize Streak Disease, Cassava Mosaic Disease and Tomato Yellow Leaf Curl Disease). Collectively, these studies are aimed at improving our understanding of the different life history strategies of pathogens, including reproduction, migration between and within hosts and virulence. By explicitly connecting our pathogenicity and virulence measurements to the natural history of the pathogen, we are confident this will provide a broader perspective on the evolution of pathogens for which current findings are mainly limited to theoretical or experimental settings.

Viral Emergence

In collaboration with researchers from CIRAD France, the State University of Arizona and the University of Cape Town, we have adopted a "geo-metagenomics" approach to investigate whether ecological disturbances such as intensive agriculture, cause measureable changes in the spatial and temporal diversity, demographics and evolutionary dynamics of viral communities inhabiting natural ecosystems that are linked to the emergence of socially- relevant crop-infecting geminiviral diseases. Our study population is the fynbos ecosystem situated in the Western Cape region of South Africa – a biodiversity hot-spot which is extremely threatened by increasing human population densities, urbanization and agriculture. To complement the data on contemporary viral diversity derived from field studies, we are sequencing "ancient" viral genomes present in fynbos plant specimens stored within French and South African herbaria to more accurately estimate when and where, key ancestral plant viruses existed. By producing infectious clones of both these ancient genomes and re-synthesized computationally inferred ancestral genomes, it will be possible to pinpoint when and where over the past few centuries, major changes in virus virulence have occurred. It is hoped that this project will complement the results obtained from our previous work on virulence evolution in Maize streak virus and increase our understanding of the evolutionary and epidemiological processes by which members of this important group of crop pathogens (i) emerged during the colonization era from their indigenous African hosts, (ii) began causing diseases in crop species introduced by Europeans and (iii) disseminated across the continent from sites of initial emergence to become what are today among the most important biotic threats to African food security.

HUMAN–INFECTING VIRAL PATHOGENS IN THE ORDER PICORNAVIRALES

(+)ssRNA viruses

We are investigating the degree to which RNA viral evolution is constrained by secondary structure within the genomes of positive sense single-stranded RNA viruses in the order Picornavirales (one of the most genetically diverse of the positive-sense single-stranded RNA viral orders and the most common cause of infections in humans in developed countries. Among the most notable serious health threats include clinical syndromes such as HIV/AIDS, aseptic meningitis, paralytic poliomyelitis, SARS and hepatitis.

Human immunodeficiency virus

In collaboration with researchers from the University of KwaZulu Natal and the University of Cape Town we have received funding from the Sub-Sahara Africa Network for TB-HIV Research Excellence (SANTHE) to computationally reconstruct, the early spatio-temporal history of the HIV-1M epidemic in order to i) better understand the evolutionary dynamics of HIV-1M before the onset of the global pandemic and ii) elucidate the factors that mediated the emergence of this important human pathogen. Such an understanding could be crucial both for retracing the earliest evolutionary steps during the emergence of HIV-1M in humans, and accurately reconstructing the ancestral sequences of the major contemporary circulating HIV-1M lineages.

RESEARCH COLLABORATIONS

1. HIV-1 Recombinants and Epidemiologically Important Subtypes in Africa and Abroad

Collaborating Parties:

Marcel Tongo, Jeffrey R. Dorfman and Tulio de Oliveira - University of KwaZulu-Natal
Darren P. Martin - University of Cape Town

Nature and purpose:

To reconstruct the evolutionary dynamics of HIV-1M before the onset of the global pandemic we analysed all published subtype A near full length sequences together with subtype-A derived genome fragments from 22 different Circulating Recombination Forms (CRFs) using Bayesian spatial diffusion models.

Output in the last 12 months:

An article was published in the journal *Virus Evolution* in 2017.

Future Direction:

To investigate the other HIV-1 group M subtypes using a similar analytical approach and perform a comparative analyses with subtype A.

2. Positive sense single-stranded ribonucleic acid viral genomes

Collaborating Parties:

Darren Martin - University of Cape Town
Brejnev Muhire - Stanford University
Emile Tanov - University of the Western Cape

Nature and purpose:

We are investigating the degree to which RNA viral evolution is constrained by secondary structure within the genomes of positive sense single-stranded RNA viruses in the order Picornavirales.
A student has analysed full genome datasets from 51 pathogenic virus species including the causal agents of among others, poliomyelitis, HIV/AIDS, hepatitis A and E and dengue fever using a wide selection of advanced computational methods to detect and study highly conserved secondary structures in relation to recombination, natural selection and co-evolution of base-paired sites. The results of this research are of interest to a wide audience of medical virologists.

Output in the last 12 months:

One PhD thesis entitled "Identification and ranking of pervasive secondary structures in positive sense single-stranded ribonucleic acid viral genomes" was submitted for examination in November 2017.

Future Direction:

Two manuscripts reporting the findings of this PhD thesis will be submitted in 2018.

3. Virulence Evolution

Collaborating Parties:

Philippe Roumagnac - CIRAD, Montpelier, France
Darren P. Martin, Adérito Luis Monjane, Dionne Natalie Shepherd - University of Cape Town, South Africa
Philippe. Lemey and Simon Dellicour - Katholieke Universiteit, Leuven, Belgium
Pierre Lefeuvre and Jean-Michel Lett - CIRAD, Reunion Island
Arvind Varsani - University of Canterbury, Christchurch, New Zealand

Nature and purpose:

We have received funding from the National Research Foundation (NRF) to investigate how the traits of virulence and transmission rates are evolving in viral pathogens important to human health. Using a combination of laboratory experimentation and state of-the-art computational analyses, we are investigating

the spatiotemporal dynamics and evolution of a broad range of viral pathogens focusing primarily on viruses that cause important crop diseases (MSD, CMD and TYLCD).

Output in the last 12 months:

None

Future Direction:

It is hoped that at least one manuscript reporting the findings of this study will be submitted for publication in 2018.

4. Molecular Mechanisms of Viral Emergence

Collaborating Parties:

Pauline Bernardo, Sarah François, Maëlle Deshoux, Denis Filloux, Emmanuel Fernandez, Serge Galzi, Romain Ferdinand, Martine Granier, Michel Peterschmitt and Philippe Roumagnac - CIRAD-INRA-SupAgro, Montpellier, France

Brejnev Muhire, Darren P. Martin - University of Cape Town, South Africa

Simona Kraberger, Arvind Varsani - University of Canterbury, Christchurch, New Zealand

Anna-Liisa Laine, Mikko J. Frilander - University of Helsinki, Finland

Armelle Marais, Thierry Candresse - INRA, UMR Villenave d'Ornon Cedex, France

Pablo Monge, Fernando Escriu - Unidad de Sanidad Vegetal Zaragoza, Spain.

Nature and purpose:

To investigate whether ecological disturbances such as intensive agriculture, cause measureable changes in the spatial and temporal diversity, demographics and evolutionary dynamics of viral communities inhabiting natural ecosystems that are linked to the emergence of socially- relevant crop-infecting geminiviral diseases such as Tomato yellow leaf curl disease (TYLCD), Maize streak disease (MSD) and Cassava mosaic disease (CMD).

Output in the last 12 months:

A manuscript was published in the ISME journal *Nature* in 2017 and a second review article will be submitted in 2018.

Future Direction:

The rate of discovery of "emergent" viruses has increased over the past two decades. However, it remains difficult to determine whether these viruses are truly emergent, or, if they have simply remained undetected until now. We have shown that one such plant-infecting single-stranded DNA geminivirus species named Euphorbia caput medusae latent virus displays a high prevalence in samples of the wild spurge Caput medusae collected between Cape Town and Paternoster on the west coast of South Africa is related to other newly discovered species that cause severe infections in French beans in India and alfalfa in France.

5. Viral dynamics

Collaborating Parties:

Alexandre De Bruyn, Mireille Harimalala, Innocent Zinga, Murielle Hoareau, Bernard Reynaud, Jean-Michel Lett and Pierre Lefeuvre - CIRAD, Reunion Island

Arvind Varsani and Matthew Walters - University of Canterbury, Christchurch, New Zealand

Darren P. Martin - University of Cape Town, South Africa

Virginie Ravigné - CIRAD, Montpellier, France

Simona Kraberger - Colorado State University, Fort Collins, USA

Andrew D.W. Geering, Sharonvan Brunschot and John E. Thomas - University of Queensland, Brisbane, Australia

Kaveh Bananej - Iranian Research Institute of Plant Protection (IRIPP), Iran

Nature and purpose:

To apply recombination analyses and Bayesian phylogeographic inference methods to all available MSV viral sequences and reconstruct a plausible history of ongoing diversification and movements throughout the southern hemisphere with a specific focus on it's introduction(s) into Madagascar.

Output in the last 12 months:

This work resulted in a single publication in 2017.

Future Direction:

In 2017 we secured funding from the National Research Foundation to collect and sequence "ancient" viral genomes present in fynbos plant specimens stored within French and South African herbaria to more accurately estimate when and where, key ancestral plant viruses existed.

6. Endogenous Viral Elements (EVE's)

Collaborating Parties:

Philippe Roumagnac, Denis Filloux and Emmanuel Fernandez - CIRAD, Montpelier, France

Darren Martin - University of Cape Town

Véronique Jamilloux, Florian Maumus - INRA, URGI, Versailles, France

Pierre-Yves Teycheney - CIRAD, UMR AGAP, Guadeloupe, France

Pierre Lefeuvre and Jean-Michel Lett - CIRAD, UMR PVBMT, Réunion Island, France

Thierry Candresse, Armelle Marais, Sébastien Theil and Chantal Faure - INRA, UMR BFP, Villenave d'Ornon France

Andrew Geering - the University of Queensland, Australia

Nature and purpose:

To study endogenous viral elements (EVE's) and their role in virus evolution and their functions in plants to address fundamental biological issues that are critical to the understanding of plant virus evolution and the contribution of plant viruses to both the host genome evolution and biology. This involves approaches in virology, bio-informatics, genomics, proteomics, evolution and synthetic biology, and aims at disentangling key mechanisms of adaptive plasticity of plant and viral genomes such as gene domestication, genetic and epigenetic regulation of plant gene expression and deciphering mechanisms that are central to molecular plant-virus interactions such as recombination, transcriptional or post transcriptional gene silencing.

Output in the last 12 months:

None

Future Direction:

We received funding to support this research from the French National Research Foundation (ANR) in December 2017 and have begun generating data. An initial meeting of all of the major stakeholders was held at CIRAD Réunion Island in September 2017 and a second meeting/workshop, is scheduled to be held at SANBI in December 2018 to assess the progress made after one year.

RESEARCH LABORATORY OF ULJANA HESSE

My primary research interests focus on the genomics of South African medicinal plants and their symbiotic microorganisms. South Africa is home to one of the six floral kingdoms of the world, harboring nearly 9000 plant species (20% of the plant species on the continent), 70% of which are endemic. Over 3000 plant species are known to possess medicinal properties, and approximately 1000 are actively traded.

Yet, locally, commercial plant production, plant breeding, and biotechnological exploitation of this immense biological resource are virtually missing. The genomic background of these plant species has not been investigated, and biosynthetic pathways for the diverse medicinal compounds remain unknown. Reduced sequencing costs have sparked a renewed worldwide interest in plant biotechnology as means to develop novel drugs and herbal supplements. Governments across the globe actively promote plant genomics research (e.g. USA, Canada, China, South Korea). My goal is to establish all procedures essential for plant genome analysis, data mining and biotechnological utilization locally, so as to provide South Africa with a leading position in the field of medicinal plant biotechnology, and to advance development of commercial production systems and protection of this endemic natural resource. The program was initiated in 2016 through successful NRF grant applications for sequencing the genome and diverse transcriptomes of rooibos, one of the few indigenous South African medicinal plant with a well-established production and marketing system. In 2016, rooibos plants from diverse locations in the Cederberg Mountains were sampled and characterized biochemically. We also established RNA extraction methods suitable for plant material rich in polyphenols.

Members of the Hesse lab.

HIGHLIGHTS OF THE LAB:

Students in the research team:

- One PostDoc: Wesley William
- One PhD student: Emily Stander
- Two MSc students: Yamkela Mgwatyu, Ben Muteba
- Three Honours students from the Department of Biotechnology, UWC namely, Min-Ghah Kariem, Danielle Davids and Tyler Hartel. All three students will be graduating in April 2018 with "*cum laude*".

All students attended the 5-week Introduction to Bioinformatics course and a workshop on Variant Structure Prediction, both hosted at SANBI.

Collaborations:

- Rooibos Ltd (the largest rooibos processing company in South Africa)
- Fair Wupperthal Agricultural Primary Co-Operative Limited
- Northern Cape Department of Agriculture
- Small holder farmers, as well as medium and large rooibos producers in the larger Cederberg Mountain region (Clanwilliam, Nieuwoudtville, Wupperthal)
- Cape Peninsula University of Technology (CPUT)
- WestCape Biotech

Yamkela Mgwatyu and Wesley Williams taking morphological measurements from rooibos ecotypes for taxonomic classification analyses.

Research Highlights:

We have established a flow cytometry method to investigate the genome sizes of plants with high levels of polyphenolds and other secondary metabolites. The genome sizes of diverse growth forms of rooibos as well as of the important South African medicinal plant Sutherlandia frutescens (cancer bush) have been determined.

From the nearly 50 rooibos plants that were biochemically characterized in 2016 and 2017, we have selected five plants for genome and transcriptome analysis. In 2017, we extracted DNA and RNA from these samples, and sequenced the genome from one rooibos ecotype and six transcriptomes from morphologically and biochemically diverse rooibos ecotypes using the Illumina sequencing technology. We are now in the process of biocomputational data analysis.

Rooibos, like most other legumes, undergoes a symbiotic interaction with nitrogen fixing bacteria. In contrast to the plant RNA, bacterial RNA is not polyadenylated, and therefore is rarely captured during sequencing of plant material. To investigate plant and bacterial genes involved in the symbiosis, we have established laboratorial procedures that allow simultaneous extraction, reverse transcription and sequencing of bacterial and host plant RNA from nodulated rooibos roots. One of the transcriptomes sequenced in 2017 was constructed from such plant material. Data analysis will be conducted in 2018.

We have identified a number of diverse fungal species that colonize rooibos endophytically, i.e. live within the plant. Such endophytes are of interest as they may affect the production of medicinal compounds, or even produce some themselves. In 2017, we have focused on the taxonomic characterization of one such

endophyte, which were found to colonize cultivated as well as wild growing rooibos plants in different regions of the Cederberg mountains, i.e. seems to be a common symbiont of rooibos.

Publication:

Virome assembly and annotation: a surprise in the Namib Desert.
(2017) Hesse U, van Heusden P, Kirby BM, Olonade I, van Zyl LJ, Trindade M. Frontiers in Microbiology 8.
Impact Factor: 4.2
Among top 20% most viewed and downloaded articles in the first quarter of 2017.

Journal Review:

Laffy et al. "Coral Reef Viromics: Diversity, Host-Specificity & Functional Capacity"
Environmental Microbiology

Community Engagement:

On 25 May 2017, E. Stander and Y. Mgwatyu presented at the Research Day of Rooibos Ltd on the rooibos genomics program and the thin layer chromatography (TLC) method for aspalathin visualization. More than 50 farmers and a journalist from the national agricultural magazine *Landbou Weekblad* attended the meeting.

To introduce our project to the farmers in the Wupperthal area, we presented on our work and conducted a practical demonstration of the TLC method for the farmers of the Fair Wupperthal Agricultural Primary Co-Operative Limited in Wupperthal. This meeting was conducted on the 30 th of August and was attended by 40 farmers.

Presentation to small holder farmers from the Fair Wupperthal Agricultural Primary Co-Operative Limited in Wupperthal.

Presentation to small holder farmers from the Fair Wupperthal Agricultural Primary Co-Operative Limited in Wupperthal (method demonstration).

RESEARCH LABORATORY OF SIMON TRAVERS

HIGHLIGHTS OF THE LAB

Prof Simon Travers

2017 was an incredibly busy year working to expand the capabilities of Exatype (our online solution that enables low cost drug resistance testing) which was launched in 2016.

In late 2016 we had been awarded a grant by the South African Medical Research council to expand Exatype to support TB drug susceptibility testing. Work 2017 involved expanding our core algorithms to be able to work with sequence data from entire TB genomes and to be able to call mutations in both coding and non-coding DNA. In parallel with this we developed a complete TB resistance scoring algorithm that can be used to call resistance to the broad spectrum of TB drugs currently in use from DNA sequences obtained from a TB infection. We have partnered with researchers in Stellenbosch University as well as the Critical Path to TB Drug Regimens and partners in the US Center for Diseases Control (CDC) to undertake validation of our TB solution.

In March 2017 I travelled to Washington DC to take part in the Critical Path to TB Drug Regimens Workshop where key players in the TB treatment and drug resistance space come together to showcase the progress both in TB drug development and diagnostics. I sat on an expert panel that discussed the role of regulators and policymakers in enabling and evolving the diagnostic landscape.

2017 was also the year where we saw real traction in the use of our Exatype HIV drug resistance pipeline through our spinoff company Hyrax Biosciences.

One of the largest diagnostic laboratories in the US is now using Exatype for its routine HIV drug resistance testing. Further, we have been working with the Kenyan Medical Research Institute (KEMRI), the Clinton Health Access Initiative in Kenya and an industry partner to facilitate the rollout of routine HIV drug resistance testing in Kenya. 2018 will be an exciting year where we are aiming to facilitate the rollout of routine HIV drug resistance testing in a number of other resource limited settings.

2017 was a productive year in our HIV vaccine work with the publication of two papers and the work completed for another 3 publications that will be submitted in early 2018. Further, two students, Roux-Cil Ferreira and Phillip Labuschagne, have completed their work and are now writing up their PhD theses for submission in early 2018.

RESEARCH PROJECTS

1. The development of computational approaches for highly sensitive analysis of next-generation sequencing (NGS) data.

One of the biggest challenges with NGS is the complexity involved in the analysis of the data. This is particularly difficult for infectious diseases, such as HIV, where it is important to be able to identify pertinent mutations in the infection in the presence of a lot of noise introduced by sample preparation and the sequencing process.

Following on from the launch of Exatype for HIV drug resistance testing in 2016, 2017 focused on the expansion of Exatype to support TB drug susceptibility testing. Exatype for HIV involves mapping sequence data to a single coding region of the HIV genome, while expansion to TB meant that we needed to expand our algorithms to be able to call mutations in both coding and non-coding DNA.

This was successfully achieved in 2017 meaning that the Exatype platform is now capable of mapping sequence reads to multiple reference sequences in parallel thereby enabling the accurate, fast analysis of both whole genome and targeted sequencing data.

2017 saw the completion of the underlying database and internal communication protocols to enable the transfer of Exatype outputs into a real-time drug resistance surveillance solution. The next body of work involves the development of analytics and reporting solutions that will mine the underlying database in order to provide end-users with an easy to use surveillance solution.

2. The use of high-throughput sequencing approaches to explore viral and host diversity.

Our work on developing a data analysis pipeline for undertaking sieve analysis to explore the mechanism of viral evasion of vaccine-induced immune pressure focused on the characterization of recombination in primer ID data. Such data enables the comparison of HIV-1 sequences from vaccine-experienced and vaccine-naïve individuals and allows attribution of vaccine-versus-placebo viral genetic signatures to vaccination. PCR-induced recombination during the sample preparation step, however, can introduce errors into the estimates. We have developed an approach to identify the presence of recombination and minimise the effect that such data would have on analysis.

3. The application of molecular dynamics approaches to characterise the viral epitopes of broadly cross-clade neutralising (BCN) antibodies with a view to identifying novel targets for HIV vaccine design.

The HIV-1 'glycan shield' - the sugars that form a dense layer on the viral surface protein, Envelope - is not only important for its protective qualities but is also a key focus in vaccine research. In recent years, various studies have identified specific glycans that are essential components of broad and potent neutralising antibody epitopes; laboratory experiments have therefore focused on editing the glycan shield in such a way to create, or expose, a glycan-dependent epitope that would increase its sensitivity to neutralisation antibodies. The current understanding is that the removal of a glycan, or a cluster of glycans, creates a gap in the glycan shield, thereby allowing antibodies to bind and neutralise the virus. Our research over the last year has focussed on describing the mechanism by which this happens, and we have presented evidence, beyond the glycan hole/gap view, on how the glycan shield can rearrange to compensate for the loss of a glycan and escape antibody neutralisation. We have carried out a detailed investigation of two different HIV-1 Envelope glycoproteins, using computational molecular dynamics simulations, and show how the loss of a single glycan results in distinct conformational changes of the remaining glycans that form the glycan shield. To our knowledge, this is the first report of how a ripple effect occurs across the surface of the protein when the glycan shield is disrupted, and how this ripple effect is dependent on the original composition of the glycan shield. Our research provides an in-depth description of both the direct and indirect effects of the loss of a specific glycan from two different HIV-1 Envelope glycoproteins, where a series of movement and interaction events has a unique influence on the integrity of an epitope located far from the original mutation. Thus, our work presents novel considerations for laboratory scientists focussing on eliminating or exposing glycan-associated HIV-1 Envelope epitopes.

RESEARCH COLLABORATIONS

1. The role of N-linked glycosylation in the escape of HIV from, and susceptibility to, neutralizing antibodies.

Collaborating Parties:

Dr Oliver Grant, Prof Robert Woods - University of Georgia
Dr Elisa Fadda - Maynooth University, Ireland
Prof Penny Moore - NICD, South Africa
Dr Jeffrey Dorfman, Dr Natasha Wood - University of Cape Town

Nature and purpose:

As part of the post-translational processing of a HIV virion carbohydrates are added to the surface of the virion by the hosts glycosylation mechanism. The binding of such N-linked glycans conveys protection to a virion's surface proteins by acting as a shield to avoid detection by the host's immune system. These carbohydrates, however, may comprise a novel target for HIV therapeutics and we are using molecular dynamic simulations to further understand the conservation and dynamics of these carbohydrates and their effect on the accessibility of the underlying protein.

Output in the last 12 months:

Colin Anthony, Talita York, Valerie Bekker, David Matten, Philippe Selhorst, Roux-Cil Ferreria, Nigel J Garrett, Salim S Abdool Karim, Lynn Morris, Natasha T Wood, Penny L Moore, Carolyn Williamson (2017) Cooperation between strain-specific and broadly neutralizing responses limited viral escape and prolonged the exposure of the broadly neutralizing epitope. Journal of Virology 91, 18 (e00828-17).

Thandeka Moyo, Roux-Cil Ferreira, Reyaaz Davids, Zarinah Sonday, Penny L Moore, Simon A Travers, Natasha T Wood, Jeffrey R Dorfman Chinks in the armor of the HIV-1 Envelope glycan shield: Implications for immune escape from anti-glycan broadly neutralizing antibodies (2017) Virology501 (12-24).

Poster presentation at the 24th HIV Dynamics and Evolution meeting – Using molecular dynamics to illustrate the changes in the glycan shields of two HIV-1 envelope trimers after the loss of a glycan.

Future Direction:

This work will continue to focus on studying the effects of glycan addition/removal on the underlying structure of the glycan shield.

2. Development and application of next-generation sequencing for HIV drug resistance genotyping.

Collaborating Parties:

Prof Maria Papathanasopolous - University of the Witwatersrand
Dr Robert Shafer - Stanford University

Nature and purpose:

Next-generation sequencing holds great promise for low-cost, highly scalable HIV drug resistance genotyping but there are a number of hurdles that need to be addressed before these approaches can be used routinely in a clinical setting. We are working on a number of projects with our collaborators to explore the optimal approach to produce and analyse NGS data for HIV drug resistance genotyping.

Output in the last 12 months:

Oral presentations at a number of national and international meetings.
Signature of a contract for use of the Exatype HIV drug resistance testing pipeline through the spinoff company Hyrax Biosciences.

3. Using next-generation sequencing to explore the diversity of the HIV viral quasispecies.

Collaborating Parties:

Prof Lynn Morris - NICD, South Africa

Prof Carolyn Williamson - UCT, South Africa

Dr Paul Edlefsen - Fred Hutchinson Cancer Research Center, Seattle The HIV vaccine trials network (HVTN)

Nature and Purpose:

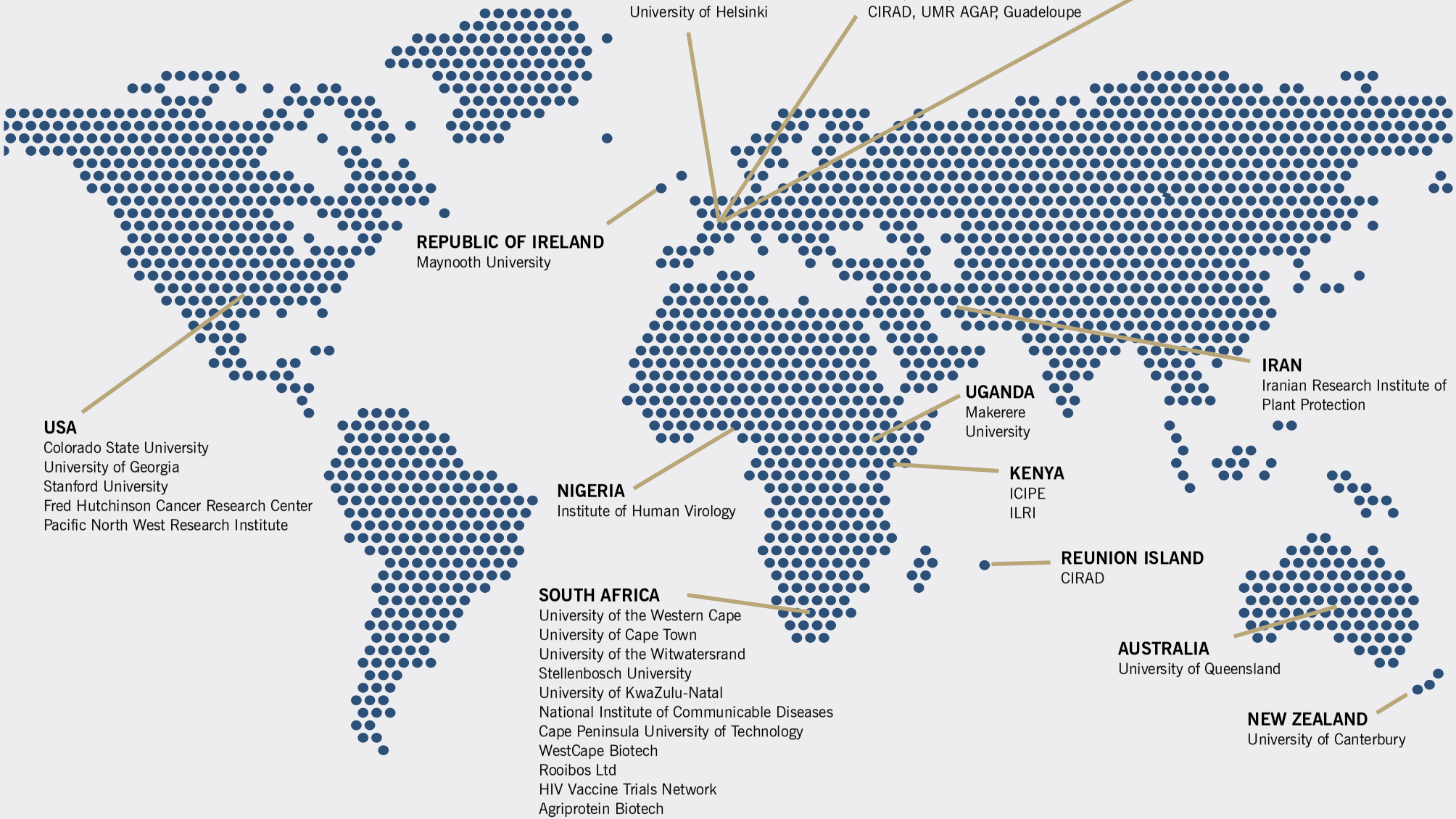
We are involved in a number of research projects that are harnessing the power of NGS to enable us to fully understand the diversity of the viral quasispecies in HIV infected individuals. The vast majority of this work is focused on the identification of the potential of individuals to produce broadly cross-neutralising antibodies as well as exploring the viral diversity of breakthrough infections following vaccination with HIV vaccine candidates.

Future Direction:

This pipeline will continue to be developed to support the analysis of data originating from a number of HIV vaccine trails being undertaken in South Africa.

Collaborations

SANBI researchers have established a vast network of partnerships and collaborations with peers all over the world.



2017 Financials

SANBI's income and expenditure trends for 2017 are shown in this section.

FIGURE 1. DISTRIBUTION OF INCOME RECEIVED FROM ALL SOURCES.

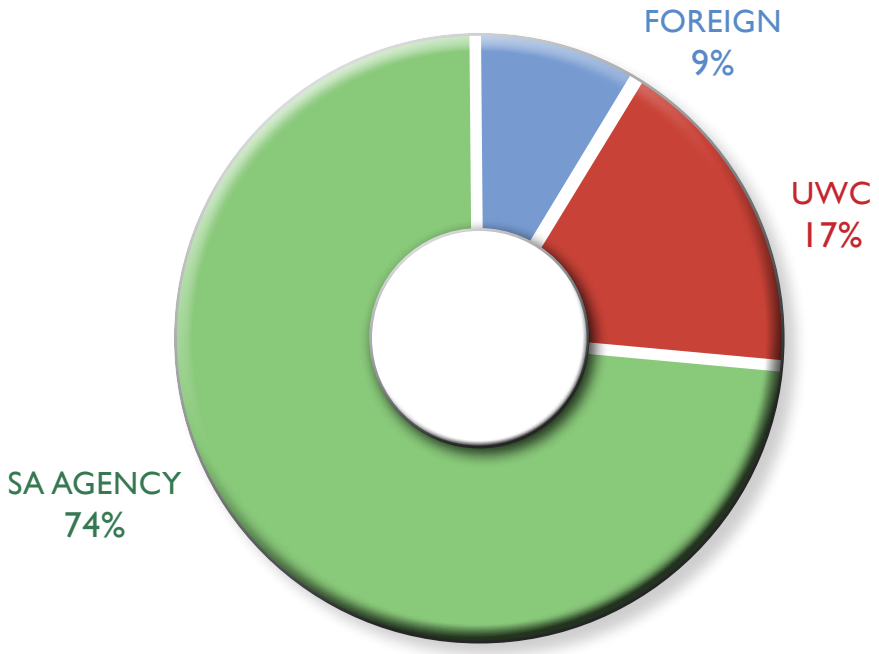
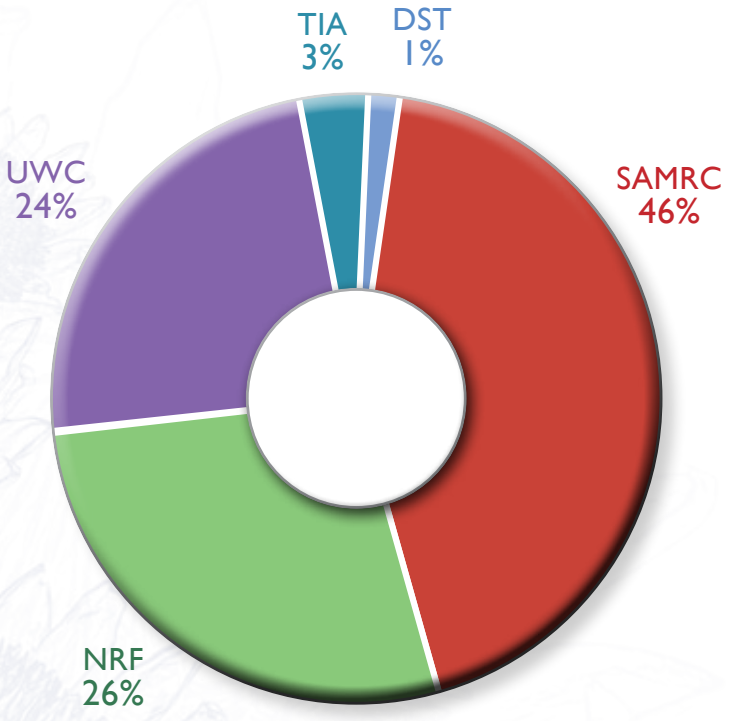


FIGURE 2. INCOME RECEIVED FROM SA SOURCES.



KEY:	
TIA	= Technology Innovation Agency
DST	= Department of Science and Technology
SAMRC	= South African Medical Research Council
NRF	= National Research Foundation
UWC	= University of the Western Cape

FIGURE 3. INCOME VS. EXPENDITURE 2011-2017.

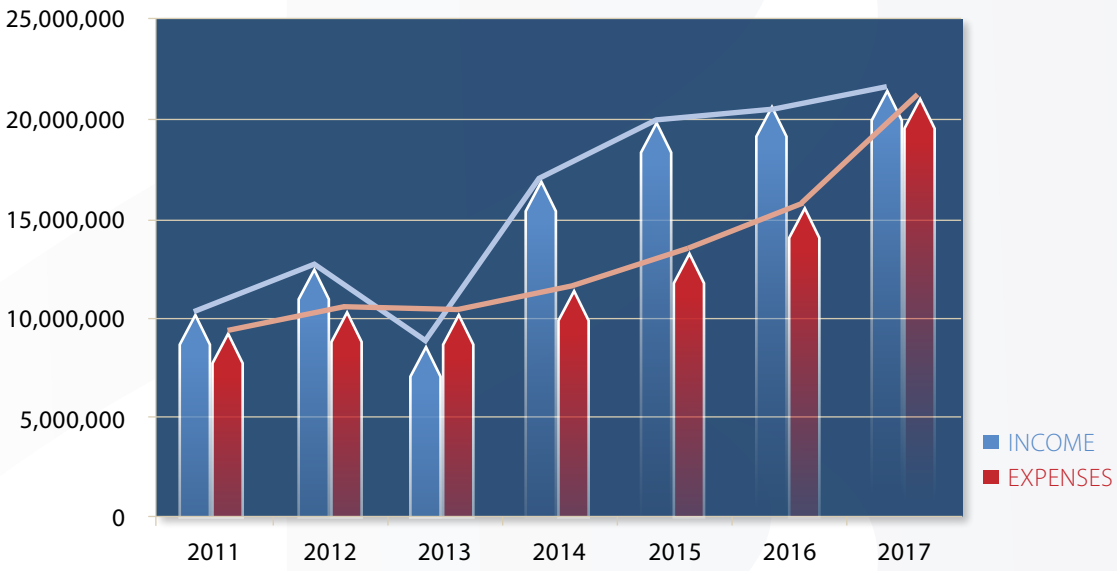
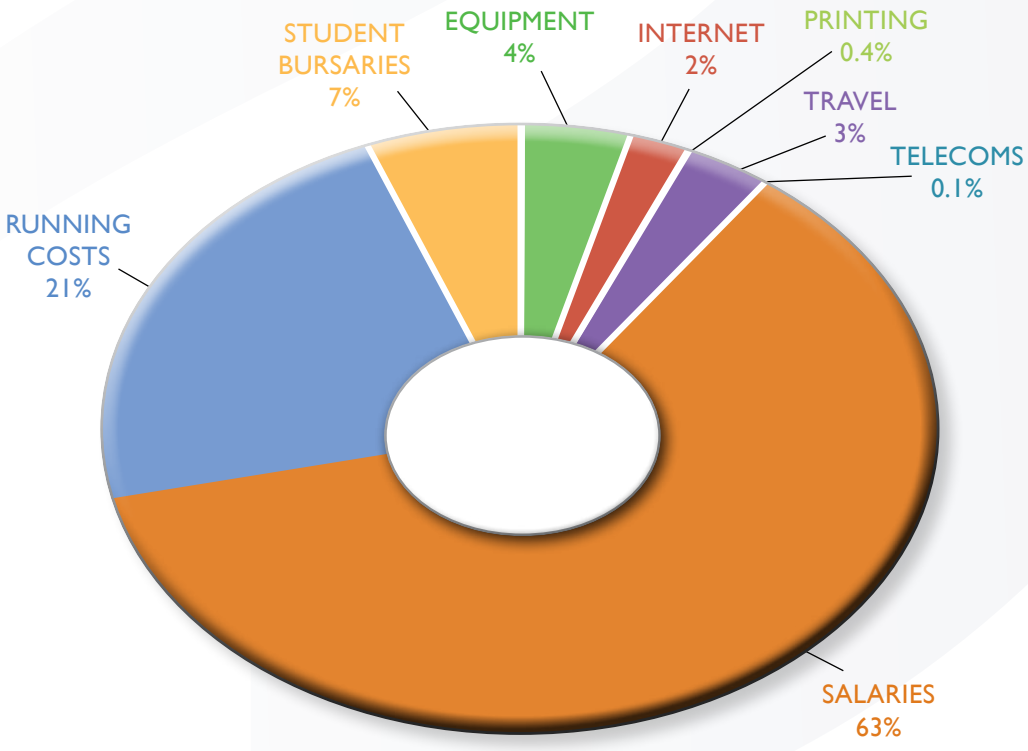


FIGURE 4. EXPENDITURE IN 2017.



Alumni

STAFF:

Name	Currently
Winston Hide	Chair of Computational Biology, University of Sheffield Adjunct Professor of Bioinformatics and Computational Biology, Harvard School of Public Health, Harvard Stem Cell Institute, Director, HSCI Center for Stem Cell Bioinformatics
Vladimir Bajic	Director & Professor, Computational Bioscience Research Center, King Abdullah University of Science and Technology
Heikki Lehvaslaiho	Senior Research Scientist, Computational Bioscience Research Centre, King Abdullah University of Science and Technology
Tulio de Oliviera	Senior Bioinformatics Researcher, Africa Centre for Health and Population Studies, University of KwaZulu-Natal
Nicky Mulder	Head Computational Biology Group, Institute of Infectious Disease and Molecular Medicine, University of Cape Town
Cathal Seoighe	Stokes Professor of Bioinformatics, School of Mathematics, Statistics and Applied Mathematics, National University of Ireland, Galway
Dale Gibbs	IT Consultant
Samantha Alexander	Administrative Assistant, UCT Faculty of Commerce
Mario Jonas	Data Manager, UCT Computational Biology Group
Nicki Tiffin	Researcher, UCT School of Public Health and Family Medicine
Natasha Schoeman	UWC Administration
Long Yi	DevOps Manager, Konga.com

POSTDOCTORAL FELLOWS:

Name	Date Completed	Currently
Soraya Bardien-Kruger	2002	Associate Professor, University of Stellenbosch
Vladimir Babenko	2002	Senior Staff Scientist, IC&G
Janet Kelso	2004	Max Planck Institute for Evolutionary Anthropology
Raphael Isokpehi	2004	Director of the Center for Bioinformatics & Computational Biology, Jackson State University
Konrad Scheffler	2005	Theodore Gildred Research Facility, University of California, San Diego
Gwen Koning	2006	Global Seed Core Manager – Syngenta Crop Protein AG, Basel, Switzerland
Chris Maher	2007	Assistant Professor, Washington University School of Medicine
James Patterson	2009	Unknown
Adam Dawe	2009	SANBI Staff, 2012
Sunil Sagar	2009	Research Scientist, KAUST
Mandeep Kaur	2009	University of the Witwatersrand School of Molecular and Cell Biology (MCB)
Stuart Meier	2009	Research Scientist, KAUST
Adele Kruger	2010	Wayne State University, Detroit, Michigan

Oliver Hofmann	2010	Affiliated Faculty, Harvard Stemcell Institute, Associate Director at Harvard School of Public Health
Sundarajan Seshadri	2010	Nanyang Technology University, Singapore
Ashley Pretorius	2010	Chief Scientific Officer, Aminotek
Sumir Panji	2012	H3ABioNet Project Manager, UCT
Samson Muyanga	2012	Researcher, Monash South Africa
John Pool	2012	University KwaZulu-Natal
Uljana Hesse	2013	SANBI Staff
Barbara Picone	2013	Unknown
Monique Maqungo	2013	Unknown
Edwin Murungi	2013	Department of Biochemistry and Molecular Biology, Egerton University, Njoro, Kenya
Hannah Ajoge	2013	Postdoctoral Fellow, University of Western Ontario, Canada
Natasha Wood	2014	Lecturer, UCT Computational Biology Department
Sarah Mwangi	2016	Karolinska Institute, Sweden
Zahra Jalali	2016	Postdoctoral Fellow, University of British Columbia, Centre for Molecular Medicine and Therapeutics
Mahjoubeh Jalali	2016	Postdoctoral Fellow, Pacific Northwest Diabetes Research Institute, Washington

PHD:

Name	Date Completed	Currently
Alan Christoffels	2001	Director, SANBI; DST/NRF Research Chair
Ekow Oppon	2002	SAMRC
Junaid Gamieldien	2002	Senior Lecturer, SANBI, UWC
Zhuo Zhang	2007	Research Scientist, University of Singapore
Alan Chong	2009	Research Fellow, Beth Israel Deaconess Medical Centre, Harvard Medical School
Magbubah Essack	2009	Research Scientist, KAUST
Sebastian Schmeier	2009	New Zealand
Ulf Schaefer	2009	Research Scientist, KAUST
Aleksander Radovanovic	2010	Research Scientist, KAUST
Mark Wamalwa	2011	International Livestock Research Institute, Kenya
Musa Gabere	2011	Principal investigator, King Abdullah International Medical Research Center (KAIMRC)
Samuel Kwofie	2011	Biomedical Engineering Lecturer, University of Ghana
Mushal Ali	2013	National Institute of Communicable Diseases, Johannesburg
Kavisha Ramdayal	2014	SAP ERP Analyst, City of Cape Town
Michael Berry	2015	Roche
Azeez Fatai	2015	Senior Lecturer, Lagos State University
Alecia Naidu	2015	Postdoc, UCT
George Obiero	2015	Postdoc, Germany
Adugna Woldesemayat	2015	Unisa, Pretoria
Darlington Mapiye	2016	IBM, Johannesburg
Rosaline Macharia	2016	Lecturer, University of Nairobi

MSC:

Name	Date Completed	Currently
Tzu-Ming Chern	2003	PhD, KAUST
Elana Ernstoff	2003	Unknown
Etienne Swart	2003	Graduate Student, Princeton University
Victoria Nembaware	2003	Training Co-ordinator, H3ABioNet
Zayed Albertyn	2003	Bioinformatics Director, Malaysia
Anelda Boardman	2004	Sequencing Facility Manager, Stellenbosch University
Faisel Mosoval	2005	Senior Professional Officer, Information Systems and Technology, City of Cape Town
Nothemba Gwija-Kula	2005	Health Consultant, World Bank
Farahnaz Ketwaroo	2005	PhD, UCT
Bukiwe Lupindo	2005	SA Government Administration
Oliver Bezuidt	2007	PhD, University of Pretoria
Eugene Duvenhage	2009	Software Developer, Corporate
Frederick Kamanu	2009	PhD, KAUST
Feziwe Mpondo	2009	SAMRC Research Scientist
Saleem Adam	2011	Unknown
Firdous Khan	2012	Human Sciences Research Council, Cape Town
Wisdom Akurugu	2013	Bioinformatician, H3ABioNet, Ghana
Fred Nindo	2013	Computational Biology Department, UCT
Ram Shrestha	2013	London, UK
Leendert Cloete	2014	Bioinformatics Scientist, KappaBiosystems, Cape Town
Batsirai Mabvakure	2015	NICD, Johannesburg
Stephanie Pitts	2016	PhD, University of Stellenbosch
Eugene Madzokere	2016	Zimbabwe

HONOURS:

Name	Date Completed
Clifford Omorogie	2001
Grant Carelse	2002
Thurayah Davids	2005
Halimit Ebrahim	2009
Katlego Motlhatlego	2012
Siyanda Tsaba	2012
Stacey Moses	2012
Lynley Abdoll	2015
Warren Jacobus	2015

Funders

Research Institute at the University of the Western Cape since 1997

National Research Foundation funding since 1998

South African Medical Research Council Bioinformatics Unit since 2002

Department of Science and Technology
National Research Foundation Research Chair in Bioinformatics and Public Health Genomics since 2007

B3 Africa funding since 2015

Technology Innovation Agency funding since 2016

Bill and Melinda Gates Foundation funding since 2014

Postal Address:
South African National Bioinformatics Institute
University of the Western Cape | Private Bag X17 | Bellville | 7535

Physical Address:
South African National Bioinformatics Institute
5th Floor | Life Sciences Building | University of the Western Cape
Robert Sobukwe Road | Bellville | 7530 | South Africa

Telephone: +27 (0)21 959-3645
Facsimile: +27 (0)21 959-2512
Email: info@sanbi.ac.za

www.sanbi.ac.za