SANBI

ANNUAL REPORT 2019

SANBI
South African National
Bioinformatics Institute

UNIVERSITY of the
WESTERN CAPE

South African National Bioinformatics Institute

CONTENTS

2019

Research Laboratories:

SANBI
South African National
Bioinformatics Institute

RESPICE PROSPICE
UNIVERSITY of the
WESTERN CAPE

ABOUT SANBI

WHO WE ARE

Bioinformatics is a specialist discipline straddling the fields of biology, mathematics and computer sciences and is integral to modern biological research. The South African National Bioinformatics Institute (SANBI) is situated at the University of the Western Cape (UWC) in Cape Town. Our primary focus is the development and implementation of computational methodologies which allow biomedical 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 South Africa; and
- The development and implementation of genome annotation methods.

OUR VISION

- To become a center of excellence, achieving the highest level in biomedical 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 biomedical 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 innovations, improved efficiency and enhanced impact…"**

THE SA MRC 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…"

2018 DRAFT WHITE PAPER ON SCIENCE, TECHNOLOGY AND INNOVATION

The research and development at SANBI aligns with the Department of Science and Innovation White Paper on Science, Technology and Innovation 2019. For example, contributing to the development of **"local innovation systems"**, and **"developing human capacities"**.

NATIONAL RESEARCH FOUNDATION STRATEGY 2020

The vision and mission of SANBI aligns with the NRF's Strategy 2020. Specifically, **"Promoting globally and competitive research and innovation."**

DIRECTOR'S MESSAGE

The recent Academy of Science of South Africa Consensus Study on "Recognising individual contributions to collaborative research" is a reminder of the value of inter-disciplinary research. At SANBI, our funding stream and certainly our ability to innovate has benefited from multi-national projects. In this 2019 report, we clearly demonstrate the value of inter-disciplinary research that has resulted in postgraduate training, publications and innovative software.

In a domain that is constrained by limited data analytics skills we are excited at the recruitment of new staff and at the same time the role that SANBI has played in contributing to the workforce through staff migration to other sectors in South Africa.

STAFF RECRUITMENT

Our academic standing has been strengthened by the addition of two new staff members. Dominique Anderson was recruited as a senior researcher with a focus on biobank informatics. Hocine Bendou has joined our precision oncology team as a senior researcher.

Professor Chandra Verma was appointed as an Extraordinary Professor at SANBI. He is based at the A-Star Bioinformatics Institute in Singapore.

STAFF MIGRATION TO INDUSTRY

While we are sad to see Simon Travers move to Hyrax Biosciences in April 2020, we are proud that SANBI was at the forefront to catalyse the bioinformatics industry in South Africa.

Junaid Gamieldien has taken up the position as head of bioinformatics at Roche based in Cape Town. We will explore ways to continue to strengthen our academic-private partnerships.

INTERNATIONAL RECOGNITION

For the second year in a row, Ravnit Grewal was appointed to the WHO Strategic Advisory Group of Experts (SAGE) on In vitro Diagnostics.

The growth at SANBI over the past 12 months and the launch of new global programmes such as a Public Health Alliance for Genomic Epidemiology (PHA4GE; **www.pha4ge.org**) has provided new strategic partnerships as we strive to impact the African Continent and beyond.

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

YEAR IN REVIEW

SANBI's national and international footprint continues to grow and 2019 has been no different. We have excelled at translating our research for the good of society and have developed a cohort of next generation scientists.

RESEARCH HIGHLIGHTS

In 2019, a total of 13 journal publications and 6 book chapters were produced, and we continued development of 2 computational tools.

PATHOGEN SURVEILLANCE

Access to pathogen sequencing enables a rapid response to public health emergencies including insights into transmission dynamics and disease progression. Pertinent to this eco-system is the need for reproducible analytics, standardized protocols for data storage and access to pathogen biological specimens. At SANBI, we have been working on different elements of pathogen surveillance including data storage (COMBAT-TB-NeoDB), workbench (COMBAT-TB; **www.combattb.org**), visualisation (COMBAT-TB-Explorer), biospecimen tracking (Baobab LIMS; **www.baobablims.org**) and a genomics archive prototype (**https://github.com/jamietyger/AfricanGenomeArchive**). Our continued research and development in this space will be to integrate these projects with other international initiatives such as meta-data standards development in the Canadian IRIDA project.

The Christoffels lab created a strategic partnership with African CDC to strengthen and support public health institutes in Africa. Phase I of this plan was to survey 11 institutes spread across Zambia, Mozambique, Swaziland, Kenya, Mali, Gabon and Nigeria.

DRUG DISCOVERY

Our external collaborations in the drug discovery space continue to yield good results. We have identified novel compounds against *M.tuberculosis* and provided experimental validation.

VISITS TO PUBLIC HEALTH SITES

At the Manhiça Health Research Center, (CISM), Mozambique

Inspecting a laboratory at the CISM in Mozambique

At the National Malaria Elimination Centre in Zambia

At the Zambia National Public Health Institute

DATA SECURITY

Biomedical researchers often work with either pathogen or host (human) data but that unintended boundary is starting to blur as integration of data is sought in an attempt to better interpret the research outcomes. This data integration mirrors the philosophy behind the concept of a one health approach. The announcement of the introduction of the Protection of Personal Information Act (POPIA) in South Africa and soon to be implemented (April 2020) requires that we develop a code of conduct for biomedical research to be compliant within universities. To this end we are assessing the status of universities including our own with regard to information security as it pertains to technological, organizational and environmental factors. This assessment will inform a data management plan for our research and in cross-border collaboration, especially as we seek to implement an African Pathogen Genomics Archive with a view to replicate it on other continents.

PUBLIC HEALTH ALLIANCE FOR GENOMIC EPIDEMIOLOGY (PHA4GE)

Developments in data storage, pathogen analysis and movement of code via the Cloud provided the basis for a partnership with the Africa CDC in Ethiopia and saw the establishment of Phase1 of a pathogen genomics initiative based at the Africa CDC to strengthen public health institutes on the African content. This work has resonated with international funding agencies including the Bill & Melinda Gates Foundation who convened a stakeholder meeting among bioinformatics and public health practitioners in March 2019 to align bioinformatics and public health. By October 2019, we launched a global initiative called the Public Health Alliance for Genomic Epidemiology (PHA4GE) at the GrandChallenges meeting in Addis, Ethiopia with partners including Africa CDC, Oxford University, Washington University, and the University of Melbourne, USA CDC, University of Birmingham UK, University of British Columbia and BC Centre for Disease Control, the Broad Institute in Boston, and H3Africa. The secretariat for this programme is based at SANBI-UWC.

IMPACT OF STUDENT THESES

In 2019, MSc student, Eugene de Beste, completed a thesis with distinction on moving code to the data using openstack (**https://bit.ly/2qjPCtW** and **https://github.com/Banshee1221/Nikeza**). This work has formed part of new projects in Africa to strengthen public health institutes responsible for managing pathogen genomic data.

SOFTWARE DEVELOPMENT

Combat-TB

In 2019, 30 international biomedical researchers were invited to Berlin to share their use of graph databases in the life sciences (**https://neo4j.com/blog/neo4j-life-sciences-healthcare-workshop-berlin/**). Software developer, Thoba Lose shared SANBI R&D in this space. It was encouraging to see that our research is at the cutting edge of applying new data archive technology in the biomedical space. Our in-house implementation of the graph database (Neo4j) in COMBAT-TB (Lose et al., 2019) has had 114 views and 77 users of our system distributed across China (14 users), South Africa (13), Japan (10), USA (6), Belgium (5), Italy (5), Germany (3), India (3), Argentina (2) and Canada (2).

Baobab LIMS

While initially developed using a non-communicable disease-focused biobank as a testcase, (Baobab LIMS (**www.baobablims.org**; Bendou et al., 2017), other labs started to show interest in the functionality of managing biospecimens, freezers and inventory.
Impact: Within two years since publication, the following Baobab LIMS footprint is evident:

1. 342 views of the code online and 64 unique downloads (github)
2. Installation in African laboratories in Uganda (Makerere University), Ivory Coast (Abidjan Biobank), Tunisia (Pasteur Tunis) and Nigeria (Lagos Biobank)
3. SANBI hosted 11 members of the Precise Network (**www.precisenetwork.org**) for training on the use and application of Baobab LIMS. The Precise Network is a UK MRC funded programme with field sites in Mozambique, Kenya and The Gambia. In partnership with the Precise Network, we customized a unique implementation across their field sites. This development identified gaps in our software that slowed down importing of thousands of biospecimens. This limitation is now part of our new round of customization.
4. After a site visit to the Abidjan biobank in 2017, Alan Christoffels was funded through the West African Health Organisation to customize Baobab LIMS for the Abidjan Biobank (2019-2020) where there is a focus on infectious disease specimens. In partnership with the Abidjan biobank, the LIMS has been translated into French.

5. The Lagos Ministry of Health, Nigeria has funded the implementation of Baobab LIMS in the Lagos Biobank over a two-year period starting November 2019.

Hyrax Biosciences

The HIV drug resistance testing software developed in the research group is now being rolled out worldwide as a partnership between Hyrax Biosciences (a company spun-out from Simon Traver's research group) and Thermo Fisher Scientific. Any users who purchase Thermo Fisher's low cost HIV drug resistance (HIVDR) testing kit are provided with free access to the software for the analysis and interpretation of the data generated by the kit. Access to a low-cost, scalable, end-to-end HIVDR solution is an exciting development in the global fight against HIV and we are very proud that part of this solution was developed at SANBI. 2019 saw many national labs from Africa, Asia, Central and North America sign up to use the service. Further, 2019 saw the expansion of the software to enable TB drug susceptibility testing – this will also be rolled out globally through Hyrax Biosciences.

Galaxy Workshop

The research and development work on COMBAT-TB has cemented our annual contribution to the International Galaxy Consortium Developers' Conference that alternates between USA and Europe. We launched the first Galaxy meeting outside of USA and Europe called the GalaxyAfrica conference (**galaxyafrica.sanbi.ac.za**) in 2018 in Cape Town. This event is a data analytics forum for African researchers and technical staff to exchange ideas and technology expertise on genomics data management and analyses. We subsequently hosted this event in Kumasi, Ghana in November 2019.

AWARDS

Alan Christoffels and team received the UWC "Innovation with Social Impact" award for the development of their **www.skill4life.org** resource which has been distributed to 76000 learners in the Western Cape.

A publication co-authored by Ruben Cloete, which was submitted to the Academy of Pharmaceutical Sciences SA (APSSA) Conference held in October 2019 , won the award for the best publication in Pharmaceutical Chemistry.

UWC 15 Years' Long Service Award

SANBI Finance Administrator, Ferial Mullins
received an award for 15 years' service to UWC.

STAFF

As a research institute, the Director of SANBI reports through the faculty of Natural Sciences to the University of the Western Cape. SANBI comprises a diverse group of research scientists, technical and administrative staff who all contribute to a dynamic productive working environment.

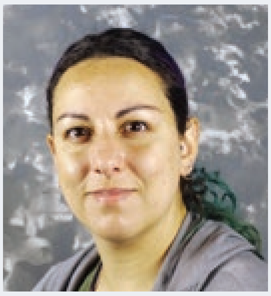
Academic staff are tasked with conducting research in their laboratories, securing funding, student graduations and capacity development. The computing infrastructure is maintained and developed by a team of technical staff while administrators ensure the smooth running of daily operations.

STAFF DEVELOPMENT

Staff are encouraged to keep their skills updated by attending workshops or courses throughout the year.

Name	Organised by	Date	Course Name + Purpose of Course
Ferial Mullins	UCT	April	Post Award Grants Administration and Legalities Workshop
Ferial Mullins	UCT and Price Waterhouse Cooper	May	VAT on Research Grants Workshop

ACADEMIC



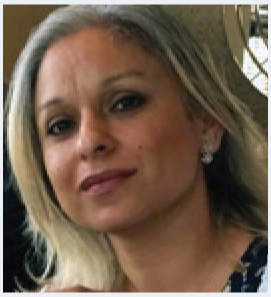
Dominique Anderson, Dr
Senior Researcher
UWC

Hocine Bendou, Dr
Senior Researcher
SA Medical Research Council

Alan Christoffels, Prof
Director & DST/NRF Research Chair in Bioinformatics & Health Genomics, SA MRC Bioinformatics Unit
DST/NRF Research Chairs Programme

Ruben Cloete, Dr
Lecturer
UWC

Junaid Gamieldien, Prof
Head of Bioinformatics
*** Roche Diagnostics*



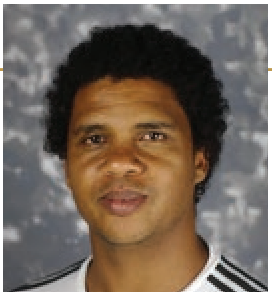
Ravnit Grewal, Dr
Senior Lecturer
SA Medical Research Council

Gordon Harkins, Dr
Senior Lecturer
UWC

Simon Travers, Prof
CEO
*** Hyrax Biosciences*

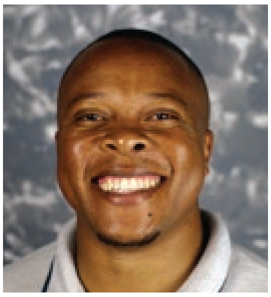
TECHNICAL

Eugene de Beste
Systems Administrator
SA Medical Research Council



Quinton Coert
Software Developer
Baobab LIMS

Thoba Lose
Development Operations
*** Entersekt*



Zipho Mashologu
Software Developer
SA Medical Research Council

Campbell Rae
Web Developer (part-time)
DST/NRF Research Chairs Programme

Peter van Heusden
Senior Systems Developer
UWC

** New employer

ADMINISTRATION

Duane Kellerman
Receptionist
SA Medical Research Council

Fungiwe Mpithi
PA/Administrator
SA Medical Research Council

Ferial Mullins
Finance Administrator
UWC

Junita Williams
Marketing Administrator (part-time)
DST/NRF Research Chairs Programme

CAPACITY DEVELOPMENT

SANBI offers training programmes which are in keeping with its vision of becoming a center of excellence in biomedical research and education in the global, African and South African context.

UNDERGRADUATE TRAINING PROGRAMME

Students who are interested in Bioinformatics as a career path are encouraged to take a combination of relevant subjects in Life or Health Sciences, Statistics, Computer Science and Mathematics during their undergraduate degree

Bioinformatics Module (BTN 315)

Each year the UWC undergraduate Bioinformatics Module is taught to approximately 85 third-year students. During April and May 2019 the SANBI faculty presented on the following topics:

- Ruben Cloete - Comparative Genomics
- Alan Christoffels - Blast and sequence alignments
- Gordon Harkins - Phylogenetics
- Junaid Gamieldien - Biodatabases

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 project.

Internship Programme

As part of the DST/NRF Research Chair Programme, SANBI provides a 1-year internship programme to students who have graduated with a BSc degree.

POSTGRADUATE TRAINING PROGRAMME

Postgraduate training at SANBI is well-established and alumni are now working all over South Africa and at bioinformatics research sites around the world.

Masters Programme

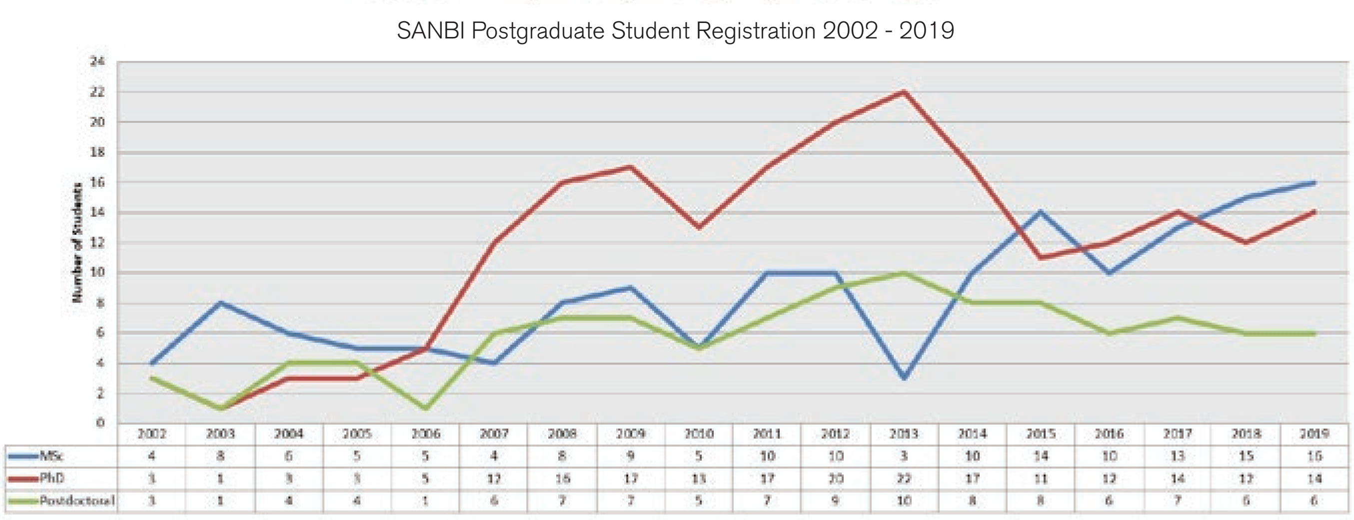
SANBI offers a research MSc in Bioinformatics. Candidates with an Honours (BSc) degree or equivalent in a related scientific subject area may apply. The MSc degree is usually completed within two years.

Doctoral Programme

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 and Engineering. The PhD degree must be completed within five years although most students aim to complete the degree in three years.

Postdoctoral Programme

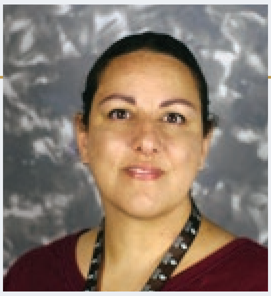
Postdoctoral fellows are admitted to the research programme after consultation with a potential SANBI supervisor. Fellowships may last for a maximum of three years.


SANBI Postgraduate Student Registration 2002 - 2019

	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
MSc	4	8	6	5	5	4	8	9	5	10	10	3	10	14	10	13	15	16
PhD	3	1	3	3	5	12	16	17	13	17	20	22	17	11	12	14	12	14
Postdoctoral	3	1	4	4	1	6	7	7	5	7	9	10	8	8	6	7	6	6

SANBI Students 2019

POSTDOCTORAL FELLOWS



Dominique Anderson

Supervisor: Alan Christoffels

Mohd Shahbaaz

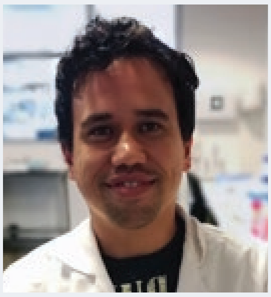
Supervisor: Alan Christoffels

Johann Josef Eicher

Supervisor: Simon Travers

Rajan Sharma

Supervisor: Alan Christoffels



Wesley Williams

Supervisor: Uljana Hesse

Ibrahim Ahmed

Supervisor: Alan Christoffels

DOCTORAL STUDENTS

Anati Nkaule

Supervisor: Alan Christoffels

Catherine Rossouw

Supervisor: Alan Christoffels

Mmakamohelo Direko

Supervisor: Alan Christoffels

Tracey Calvert-Joshua

Supervisor: Alan Christoffels

Roux-Cil Ferreira

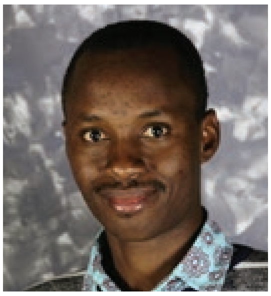
Supervisor: Simon Travers

Philip Labuschagne

Supervisor: Simon Travers

Emily
Stander
Supervisor: Uljana Hesse

Sarah
DeRaedt
Supervisor: Alan Christoffels



Souleymane
Diallo
Supervisor: Alan Christoffels

Bridget
Langa
Supervisor: Junaid Gamieldien

Hocine
Bendou
Supervisor: Alan Christoffels

Abiola Abidemi
Babjide
Supervisor: Alan Christoffels

Nasr
Eshibona
Supervisor: Junaid Gamieldien

Abdulazeez
Giwa
Supervisor: Junaid Gamieldien

MASTERS STUDENTS

Ridaa
Fredericks
Supervisor: Alan Christoffels

Farzaana
Diedericks
Supervisor: Alan Christoffels

Eugene
de Beste
Supervisor: Alan Christoffels

Samson
Oselusi
Supervisor: Alan Christoffels

Peter
Abiodun
Supervisor: Alan Christoffels

Lynley
Abdoll
Supervisor: Alan Christoffels

Fanechka
Esterhuysen
Supervisor: Junaid Gamieldien

Yamkela
Mgwatyu
Supervisor: Uljana Hesse

Jamie Southgate
Supervisor: Alan Christoffels

Ben Ilunga Muteba
Supervisor: Uljana Hesse

Peter van Heusden
Supervisor: Alan Christoffels

Susan Alecia Fernol
Supervisor: Alan Christoffels

Olabode Ajayi
Supervisor: Alan Christoffels

Darren Isaacs
Supervisor: Ruben Cloete

Rumbidzai Chitongo
Supervisor: Ruben Cloete

Maryam Hassan
Supervisor: Ruben Cloete

WORKSHOPS AND COURSES

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

Bioinformatics Training Course

This annual Western Cape course ran from 18 February – 15 March 2019. The first week aimed to get students from different backgrounds to a similar level to start the modules so that biologists could learn basic programming and programmers could learn the fundamentals of molecular biology. The rest of the course was aimed at providing an overview of a range of bioinformatics topics so that students are exposed to the broad field.

There were 25 participants in attendance from UWC, UCT, US and CPUT. Lectures were presented by specialists from the various universities - Peter van Heusden, Dr Wesley Williams, Dr Nicki Tiffin, Dr Ibrahim Ahmed, Prof Nicola Mulder, Prof Junaid Gameldien, Dr Imane Allali, Dr Ruben Cloete and Prof Darren Martin.

Attendees of the 2019 Bioinformatics Training Course held at SANBI

Galaxy Africa Workshop

After a successful GalaxyAfrica2018 conference in South Africa, SANBI held a GalaxyAfrica workshop in collaboration with the ASBCB/ISCB Africa conference in Kumasi, Ghana in November 2019. The workshop included presentations from SANBI and other researchers and training on how to use Galaxy for microbial bioinformatics as well as how to set up and administer a Galaxy server. The workshop was attended by 12 participants from 4 countries in Africa, one of whom (Michael Thompson from KNUST in Ghana) subsequently was awarded an Open Bioinformatics Foundation Travel Fellowship to further their Galaxy server administration knowledge at a workshop in Barcelona, Spain.

2019 SANBI GRADUATIONS

Name	Degree	Supervisor	Thesis Title
Hocine Bendou	PhD	Alan Christoffels	Baobab LIMS: an open-source laboratory information management system for human biobanks.
Roux-Cil Ferreira	PhD	Simon Travers	Exploring the role of the “glycan-shield” of human immunodeficiency virus in susceptibility to, and escape from, broadly neutralising antibodies.
Philip Labuschagne	PhD	Simon Travers	Development of a data processing toolkit for the analysis of next-generation sequencing data generated using the primer ID approach.
Emily Stander	PhD Bioinformatics	Uljana Hesse	Unraveling the transcriptome of Aspalathus linearis (Rooibos) towards identification of novel genes involved in polyphenol biosynthesis.
Mqondisi Tshabalala	PhD (University of Pretoria)	Alan Christoffels (co-supervisor)	Human Leukocyte Antigen (HLA) genetic diversity in South African populations.
Adetayo Emmanuel	PhD (Stellenbosch University)	Ruben Cloete	Multidisciplinary viral analyses in People Living with HIV-1C and receiving second-line combination antiretroviral therapy (cART) in South Africa.
Olabode Ajayi	MSc	Alan Christoffels	An evaluation of Galaxy and Ruffus workflows system for DNA-seq analysis.
Eugene de Beste	**MSc** ***Cum Laude***	Alan Christoffels	Enabling the processing of bioinformatics workflows where data is located through the use of cloud and container technologies.
Fanechka Esterhuyzen	**MSc** ***Cum Laude***	Junaid Gamieldien	Development of a simple artificial intelligence method to accurately subtype breast cancers based on gene expression barcodes.
Yamkela Mgwatyu	MSc Bioinformatics	Uljana Hesse	DNA barcoding of commercial and wild rooibos ecotypes.
Ben Ilunga Muteba	MSc Bioinformatics	Alan Christoffels	Data science techniques for predicting plant genes involved in secondary metabolite production.
Carla Griesel	MMed (Stellenbosch University)	Ravnit Grewal	The diagnostic utility of flow cytometry on finer needle aspirate (FNA) samples at Tygerberg Hospital (TBH).
Hassan Elamin	MSc Dentistry	Ravnit Grewal	Incidence of plasmablastic lymphoma in HIV positive patients at a tertiary hospital in South African(2005-2017).
Boiketlo Sebate	MSc (Stellenbosch University)	Ruben Cloete	The utilisation of whole exome sequencing to dissect the genetic aetiology of familial Parkinson’s disease in a South African Afrikaner Caucasian Family.
Farzaana Diedricks	BSc (Honours)	Alan Christoffels	Evaluating In SIlico spoligotyping approaches for *Mycobacterium tuberculosis* strains in South Africa.
Ridaa Fredricks	BSc (Honours)	Alan Christoffels	In-silico identification of microsatellite markers in Black soldier fly.

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.

In 2019 SANBI moved to bring its legacy infrastructure (a Dell Equallogic SAN and in-house developed research cloud) to end of life as we have migrated to a more modern Ceph distributed storage system, Slurm-based HPC cluster and OpenStack based research cloud.

On a hardware level both HPC and research cloud are largely based on Dell Blade servers providing a total of 232 CPU cores (spread across the HPC and research cloud infrastructure). Ceph storage is based on SuperMicro servers and provides 246 TB of raw storage. Servers are interconnected using 10 Gb Ethernet on fibre.

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 centres.

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

SANBI HPC Cluster

The cluster hardware configuration was not expanded in 2019 and remains on par with the configuration from 2018. We have 232 CPU cores and 1952 GB of RAM on our Dell cluster.

Virtual Machine Infrastructure

As of the end of 2018, servers providing 64 CPUs and 340 GB of RAM have been provisioned as part of our research cloud. We will complete the decommissioning of legacy virtualisation infrastructure in 2020.

DEVELOPMENT ACTIVITIES

RSSE Africa Forum

SANBI launched the Research Software and Systems Engineers (RSSEs) in Africa Forum in December of 2019. This is a discussion and networking forum for professionals involved in building software and computing systems to support research on the African continent. It is found online at **https://rsse-africa.sanbi.ac.za/** and hosts discussions and tutorials for RSSEs.

Cluster at NICD

Following on from SANBI-alumnus Batsirai Madzokere's Masters thesis, Eugene de Beste and Peter van Heusden are building a small cluster with low-cost high-performance computing infrastructure at the NICD.

Health Information Systems Program

Originally a UWC startup, this small data centre in Newlands provides opensource District Health Infromation System software (DHIS2) to large parts of Africa where resources are constrained. Eugene de Beste is currently contracted to upgrade and modernise their systems.

EMERGING RESEARCH PROJECTS

Africa CDC/PHA4GE

From August – December 2019, Alan Christoffels, Peter van Heusden and Eugene de Beste travelled to 11 labs in Africa (Zambia, Mozambique, Swaziland, Kenya, Mali, Nigeria, Gabon) to survey public health institutes. From 2020 the focus will be on setting up the Secretariat office at SANBI. Working groups will be established to align data standards and tools for public health.

Alan Christoffels and the Africa CDC team on a site visit to the Eswatini Malaria Programme.

MENTORSHIP AND TEACHING

Student Cluster Competition

At the beginning of 2019 our Systems Developer, Eugene de Beste, began training two groups of students from the Computer Science department at the UWC for the Centre for High Performance (CHPC) Student Cluster Competition (SCC). The CHPC SCC pits teams from various universities against one another in building an efficient scientific-based cluster computer system and compiling and using scientific software in a race to be first to finish as well as generate the most accurate or best results. This training involves building skills in Linux operation, computer hardware, systems administration and scientific software. The two teams were selected to compete at the national final round in December 2019 in Johannesburg, where one placed 5th and the other placed 2nd overall. One of the members of the team that placed 2nd was selected to accompany the winning team in the International Supercomputing Conference Student Cluster Challenge in 2020. Another member from the 2nd place team was selected as a reserve for said competition.

The skills that this program affords the students that are involved are invaluable, and many of the students proceed to work at important companies or enter the sciences. It has also proven to be somewhat of a channel into introducing students to bioinformatics and has built interest in post-graduate work.

Data Carpentry Workshop

Peter van Heusden, taught a 2-day Data Carpentry workshop at SANBI on 10 - 11 September 2019. Aimed at those who have little to no prior computational experience, the workshop taught fundamental data skills needed to conduct research. The workshop was organised by the UWC Library Services and e-Research Office and was attended by 30 UWC staff and postgraduate students.

SANBI IN THE MEDIA

FROM THECONVERSATION.COM

How We Used Computers To Figure Out Drugs That Can Beat Drug-Resistant TB

Ruben Cloete - January 14, 2019

Tuberculosis (TB) is one of the leading causes of death from infectious diseases. Globally, it accounts for around 1.3 million deaths and 10.4 million people develop the disease every year.

The standard first-line treatment regimen used for newly diagnosed TB patients involves taking four different drugs over a period of six months. Some patients experience varying degrees of adverse reactions as a result. These factors have led to patients defaulting on their treatment. This, in turn, has led to a rise in drug resistance.

TB drug resistance – when the bacteria become resistant to at least one anti-TB drug – is a growing problem across the world. Current treatments that include more than one drug known as combination therapy have become inadequate because the bacteria have devised ways to survive even when antibiotics are used. An estimated 3.5% of new cases and 18% of previously treated cases of TB are drug resistant. There is therefore still a need to identify alternative drugs...

FROM INSECTTECHNOLOGY.COM

Agriprotein Completes Sequencing Of Black Soldier Fly Genome

Cape Town - November 28, 2019

AgriProtein Technologies (part of the Insect Technology Group) in collaboration with the South African National Bioinformatics Institute at the University of the Western Cape, have sequenced the genome of the black soldier fly (Hermetia illucens) across both sexes and strains originating from Kenya and South Africa. In addition to the whole genome sequence, miRNA sequencing has been carried out on four life stages of the fly, namely: egg, larvae, pupae and unmated male and female fly in order to create a map of expression across the life of the fly.

Within the Insect Technology Group, we have over 50 skilled and experienced scientists working as part of our global Research & Development team to expand the understanding of the insects that we work with. Our team has embarked on unravelling the blueprint of the black soldier fly and continues to push the boundaries of scientific exploration, knowledge and understanding of this remarkable insect to drive our factory operating metrics.

Read more about what we do here –
www.agriprotein.com – and discover how we harness the power of insects to help repair the future.

FROM UWC.AC.ZA

Innovating With Impact: UWC At Innovation Bridge/Science Forum South Africa 2019

Nicklaus Kruger - December 5, 2019

The University of the Western Cape (UWC) believes in innovations that make a difference - improving the way people live, learn and communicate. So it's no surprise that UWC is taking part in the 2019 Innovation Bridge (IB) Technology Matchmaking and Showcasing event, aimed at encouraging and accelerating the utilisation and commercialisation of intellectual property and technologies developed by public research institutions.

"Innovation isn't just about the successful implementation of new or improved products, processes and services derived from new ideas and inventions," says Dr Ana Casanueva, Director of UWC's Technology Transfer Office (TTO). "It's also about the creation of social and/or monetary value - making life better in some ways..."

FROM UWC.AC.ZA

Biobanking And Me: The Book That Brings Genomics To The People

Nicklaus Kruger - December 20, 2019

Science requires trust and meaningful engagement. That's especially true for human sample biobanking, where donors need to make informed decisions. SANBI's Biobanking and Me speaking book helps them do just that - in isi-Xhosa, English and Afrikaans.

Biosciences and health sciences are advancing quickly, with new technologies and techniques invented every day. But how do we help ordinary people - and especially children - understand topics such as DNA, biobanks and diseases?

Biobanking and Me, a new speaking book recently published by Professor Christoffels and his team from SANBI, communicate these concepts to children and adults alike through informative text, colourful illustrations and bilingual audio (English-Xhosa and English-Afrikaans versions of the book are available).

"The value of creating and increasing public understanding of science through meaningful engagement platforms cannot be overestimated," says SANBI's Professor Alan Christoffels. "Failure to do so can cause irreparable damage to meaningful community engagement as we have seen in other parts of the world"...

RESEARCH OUTPUTS

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

JOURNAL PUBLICATIONS

#	Publication **(SANBI contributors)**	Published Date
1.	**Estimation of pH effect on the structure and stability of kinase domain of human integrin-linked kinase.** Syed SB, **Shahbaaz M**, Khan SH, Srivastava S, Islam A, Ahmad F, Hassan MI. Journal of Biomolecular Structure and Dynamics Volume number: 37, 2019 Issue 1 DOI: 10.1080/07391102.2017.1420492	31 January
2.	**Visualization of Aspalathin in Rooibos (Aspalathus linearis) Plant and Herbal Tea Extracts Using Thin-Layer Chromatography.** **Amor Stander, E. Williams, W.** Rautenbach, F. Le Roes-Hill, M. Mgwatyu, Y. Marnewick, J. **Hesse, U.** Molecules 2019, 24, 938.	07 March
3.	**Designing novel possible kinase inhibitor derivatives as therapeutics against *Mycobacterium tuberculosis*: An in silico study.** **Mohd Shahbaaz, Anati Nkaule & Alan Christofels.** Scientific Reports Volume 9, Article number:4405 https://doi.org/10.1038/s41598-019-40621-7	13 March
4.	**FTO haplotyping underlines high obesity risk for European populations.** Babenko, Vladimir, Babenko, Roman, **Gamieldien, Junaid** & Markel, Arcady. BMC Medical Genomics. 12. 10.1186/s12920-019-0491-x.	13 March
5.	**Ten simple rules for organizing a webinar series.** Fadlelmola Faisal M., Panji Sumir, Ahmed Azza E., Ghouila Amel, Akurugu Wisdom A., Entfellner Jean-Baka Domelevo, Souiai Oussema, Mulder Nicola, Fadlelmola Faisal M., Panji Sumir, Ahmed Azza E., Ghouila Amel, Akurugu Wisdom A., **Entfellner Jean-Baka Domelevo**, Souiai Oussema, Mulder Nicky, Abdelhak Sonia, Adebiyi Ezekiel, Allali Imane, Alzo-Hairy Ahmed Mansour, Amzazi Saaid, Badaoui Bouabid, Bbosa Nicholas, Benkhala Alia, Bensellak Taoufik, Bishop Ozlem Tastan, Botha Gerrit, Bouhaouala-Zahar Balkiss, Chaouch Malek, Chaouni Bouchra, Eljadid Sara, Fatumo Segun, **Gamieldien Junaid**, Gaseitsiwe Simane, Ghansah Anita, Hassan Atinuke, Hazelhurst Scott, Joromea Kerapetse, Kishk Ali, Koita Almamy I., Masiga Daniel, Messaoud Olfa, Mmbando Bruno, Moussa Ahmed, Mungloo-Dilmohamud Zahra, Musicha Patrick, Odia Trust, Omotuyi Idowu, Patterton Hugh-George, Samtal Chaimae, Ssemwanga Deo-Gratius, Wells Gordon. PLOS Computational Biology 15(5): 10.1371/journal.pcbi.1006671	01 April
6.	**Evidence for both intermittent and persistent compartmentalization of HIV-1 in the female genital tract.** **Mabvakure BM**, Lambson BE, **Ramdayal K**, Masson L, Kitchin D, Allam M, Abdool Karim S, Williamson C, Passmore J-A, Martin DP, Scheepers C, Moore PL, **Harkins GW**, Morris L. Journal of Virology 93:e00311-19. https://doi.org/10.1128/JVI.00311-19.	01 May
7.	**Immune Correlates of Disease Progression in Linked HIV-1 Infection.** Tuen M, Bimela JS, Banin AN, Ding S, **Harkins GW**, Weiss S, Itri V, Durham AR, Porcella SF, Soni S, Mayr L, Meli J, Torimiro JN, Tongo M, Wang X, Kong X-P, Nádas A, Kaufmann DE, Brumme ZL, Nanfack AJ, Quinn TC, Zolla-Pazner S, Redd AD, Finzi A, Gorny MK, Nyambi PN and Duerr R. Frontiers in Immunology 10:1062. doi: 10.3389/fimmu.2019.01062	14 May
8.	**Structure based identification of novel inhibitors against ATP synthase of *Mycobacterium tuberculosis*: A combined in silico and in vitro study.** **Mohd Shahbaaz, Ruben Cloete**, Melanie Grobbelaar, Samantha Sampson and **Alan Christoffels**. International Journal of Biological Macromolecules 135 (2019) 582–590	20 May

9.	**Combining Viral Genetics and Statistical Modeling to Improve HIV-1 Time-of-Infection Estimation towards Enhanced Vaccine Efficacy Assessment.** Raabya Rossenkhan, Morgane Rolland, **Jan P.L. Labuschagne, Roux-Cil Ferreira**, Craig A. Magaret, Lindsay N. Carpp, Frederick A. Matsen, Yunda Huang, Erika E. Rudnicki, Yuanyuan Zhang, Nonkululeko Ndabambi, Murray Logan, Ted Holzman, Melissa-Rose Abrahams, Colin Anthony, Sodsai Tovanabutra, Christopher Warth, Gordon Botha, David Matten, Sorachai Nitayaphan, Hannah Kibuuka, Fred K. Sawe, Denis Chopera, Leigh Anne Eller, **Simon Travers**, Merlin L. Robb, Carolyn Williamson, Peter B. Gilbert and Paul T. Edlefsen. Viruses 2019, 11, 607 doi:10.3390/v11070607	03 July
10.	**Sperm Proteome Analysis and Identification of Fertility-Associated Biomarkers in Unexplained Male Infertility.** Panner Selvam MK, Agarwal A, Pushparaj PN, Baskaran S, **Bendou H.** Genes 2019, 10(7), 522 https://doi.org/10.3390/genes10070522 ISSN: 2073-4425	11 July
11.	**Advancing HIV Vaccine Research With Low-Cost High-Performance Computing Infrastructure: An Alternative Approach for Resource-Limited Settings.** **Batsirai M Mabvakure**, Raymond Rott, Leslie Dobrowsky, **Peter Van Heusden**, Lynn Morris, Cathrine Scheepers and Penny L Moore. Bioinformatics and Biology Insights Volume 13:1-8 https://doi.org/10.1177/1177932219882347	07 November
12.	**The structural basis of acid resistance in *Mycobacterium tuberculosis*: insights from multiple pH regime molecular dynamics simulations.** **Mohd Shahbaaz**,Vladimir Potemkin, Maria Grishina, Krishna Bisetty & Imtaiyaz Hassan. Journal of Biomolecular Structure and Dynamics (0739-1102) https://doi.org/10.1080/07391102.2019.1682676	08 November
13.	**Understanding biobanking: An assessment of the public engagement speaking book intervention Biobanking and Me.** **Anja Bedeker, Dominique Anderson, Thoba Lose, Yamkela Mgwatyu**, Retha Luus, Renete Blignaut, **Alan Christoffels.** South African Journal of Bioethics Law. DOI:10.7196/SAJBL.2019.v12i2.00691	December

CHAPTERS IN BOOKS

#	Chapter Title in Book
1.	**Alan Christoffels and Peter van Heusden** (2019) Genome Annotation: Perspective from Bacterial Genomes. In: Guenther, R and Steel, D. (eds) Encyclopedia of Bioinformatics and Computational Biology. Vol. 3, pp 152-156. Oxford: Elsevier
2.	**Ravnit Grewal** WHO Strategic Advisory Group of Experts (SAGE) on In vitro Diagnostics (2019). Essential Diagnostic List https://www.who.int/medical_devices/publications/Standalone_document_v8.pdf?ua=1
3.	Samuel Egieyeh, Sarel Malan and **Alan Christoffels** (2019) Cheminformatics techniques in antimalarial drug discovery and development from natural products 2: Molecular scaffold and machine learning approaches. In: Giamberini, Mart et al (Eds) Physical Sciences Reviews. https://doi.org/10.1515/psr-2019-0029
4.	Samuel Egieyeh, Sarel Malan and **Alan Christoffels** (2019) Cheminformatics techniques in antimalarial drug discovery and development from natural products 1: basic concepts. In: Giamberini, Mart et al (Eds) Physical Sciences Reviews. Vol 4 (Issue 7). https://doi.org/10.1515/psr-2018-0130

5. Lett Jean-Michel, Scussel Sarah, Campos Paola, Filloux Denis, Roumagnac Philippe, **Harkins Gordon William**, Martin Darren Patrick, Roumagnac Philippe, Martin Darren Patrick, Lefeuvre Pierre, Rieux Adrien, Becker Nathalie. 2019.
Night at the museum: Contribution of small RNA from historical herbarium specimens in the reconstruction of evolutionary histories of geminiviruses.
In : Livre des résumés des 17 ème Rencontres de virologie végétale. INRA, CIRAD, CNRS, SFP, SFV. Aussois : INRA, Résumé, p. 52. Rencontres de Virologie Végétale (RVV 2019).
17, Aussois, France, 27 January 2019/31 January 2019.

6. Filloux Denis, Lefeuvre Pierre, Lett Jean-Michel, Varsani Arvind, **Harkins Gordon William**, Malmstrom Carolyn M., Martin Darren Patrick, Roumagnac Philippe. 2019.
Deciphering plant virus ecology and evolution using metagenomics-based approaches.
State College : s.n., Résumé, 1 p. International Congress on Virus Evolution.
10, State College, États-Unis, 7 March 2019/9 March 2019.

CONFERENCE PARTICIPATION

Presenter	Conference Name and Venue	Date	Title	Poster/ Oral
Oluwafemi Peter Abiodun	eResearch Africa 2019, UCT, Cape Town, SA	5-18 April	Exploring The Influence Of Organizational, Environmental, And Technological Factors On Information Security Policies And Compliance At South African Higher Education Institutions.	Oral
Darren Isaacs	9th SA AIDS Conference, CTICC, Durban, SA	11-14 June 2019	Structural comparison of diverse HIV-1 subtypes using Molecular modelling and Molecular Docking of integrase inhibitors.	Poster
Rumbidzai Chitongo	9th SA AIDS Conference, CTICC, Durban, SA	11-14 June	Prioritization of sequence variants in HIV-1C Integrase associated with Raltegravir drug resistance using structural methods.	Poster
Ruben Cloete	9th SA AIDS Conference, CTICC, Durban, SA	11-14 June	Presented novel findings from my research laboratory and built collaborations.	Attended
Oluwafemi Peter Abiodun	5th African Conference on Emerging Infectious Diseases & Biosecurity (GET) Abuja, Nigeria (**www.getafrica.org**)	7-9 August	Exploring The Influence Of Organizational, Environmental, And Technological Factors On Information Security Policies And Compliance At South African Higher Education Institutions, With A Focus On Implications For Biomedical Research.	Oral
Sampson Selusi	5th Conference on SA Society for Basic and Clinical Pharmacology (SASBCP)	05-07 October	Attended the conference to learn and build collaborations.	Attended
Oluwafemi Peter Abiodun	ISCB Africa ASBCB 2019 Kwame Nkrumah University of Science and Technology (KNUST), Kumasi, Ghana	11-15 November	Development Of An Effective Information Security Policy For Biomedical Researchers.	Oral

Tracey Calvert-Joshua	ISCB Africa ASBCB 2019 Kwame Nkrumah University of Science and Technology (KNUST), Kumasi, Ghana	11-15 November	Regulatory Networks in *Mycobacterium tuberculosis.*	Oral
Susan Fernol	2nd Arab Association of Genetic Research Conference, Cairo, Egypt	27 November	A whole genome sequencing approach to identifying disease-causing mutations in South African Osteogenesis Imperfecta III mixed ancestry siblings.	Oral

Participants of the ISCB Africa ASBCB 2019 held at Kwame Nkrumah University of Science and Technology (KNUST), Kumasi, Ghana.

INVITED TALKS

Presenter	Venue	Date	Title
Alan Christoffels	Armauer Hansen Research Institute, Ethiopia	February	Establishing genomics capacity for disease surveillance.
	Bill & Melinda Gates Foundation, Seattle, USA	March	Bioinformatics in LMIC. Stakeholder meeting to explore bioinformatics for public health.
	Vavilov Institute of General Genetics, Moscow, Russia	May	Computational analysis of non-coding SNPs in SA *M.tb* strains.
	Governance and Utilization of African Data for Development, Nairobi, Kenya	June	Data governance and the role of the African University.
	The Governance of Data Sharing in Health Research in South Africa, Johannesburg, SA	July	Data governance for biospecimens – lessons learnt from Baobab LIMS. Sponsored by the Wellcome Trust.

Ruben Cloete	Department of Biochemistry, University of Cambridge, UK	September	SEMINAR: Structural computational methods to tackle HIV and tuberculosis drug resistance.
Alan Christoffels	GrandChallenges Africa. Addis Ababa, Ethiopia	October	Improving openness and interoperability in public health bioinformatics.
Dominique Anderson	Noguchi Institute, Ghana	November	Biological sample management and community engagement for biorepositories.

Ruben Cloete with Emeritus Professor Sir Tom Blundell from Cambridge University.

RESEARCH MEETINGS AND WORKSHOPS ATTENDED

Attended by	Meeting Name	Date	Purpose of meeting
Dominique Anderson	Governance of data sharing in Africa workshop	February	Invited workshop on data sharing, POPIA and data governance
Alan Christoffels	BRICS-ST1 Project, Moscow	May	Report on MDR analysis and planning for future projects
Peter van Heusden	ABPHM Meeting, Cambridge	June	Discussion on pathogen surveillance and reporting on COMBAT-TB
Dominique Anderson	Medical biobank cluster meeting	June	DSI-medical collections platform meeting
	NHLS Pathology Research and Development Congress (PathRed)	July	Biobanking and clinical pathology workshop
Fanechka Esterhuyzen	Deeplearning Indaba, Kenya	August	Training and practicals on artificial intelligence and deeplearning

HOSTING OF ACADEMICS

Visitor	Institution	Date	Purpose
Philip Onyebujoh C Onyebujoh, Manjinder Sandhu, Ajay Thirumala, Akhona Tshangela, Yewande Habibat Alimi, Tessema, Sofonias, Pascale Ondoa, Amha Kebede, Jennifer Gardy (by Skype), Mushal Allam Ali, Basil Desmond Brooke, Maria Venter	African Union (AU) Africa CDC, Ethiopia Wellcome Sanger Institute/ Cambridge, UK Bryan Greenhouse's Lab, Univeristy of California San Francisco, USA African Society for Laboratory Medicine, Ethiopia Bill and Melinda Gates Foundation, Seattle, USA NICD, Wits Research Institute for Malaria Research, University of Pretoria	June	Africa CDC proposal development workshop on establishing regional partnerships for implementing Next Generation Sequencing (NGS) in public health institutions in member states of the Africa Union.
Ismael Kone	Institute Pasteur Côte d'Ivoire	September	Collaborative Baobab LIMS customisation and training.
Jennifer Gardy, Philip Onyebujoh C Onyebujoh	Bill and Melinda Gates Foundation, Africa CDC, Ethiopia	September	To continue development of the proposal to build and implement the African Genomics Archive. The AGA will be based across 5 regional centers of the African CDC, Africa CDC headquarters in Addis and SANBI-UWC. These nodes will interact and facilitate data flow/analytics from and between National Public Health Institutes.

Peter van Heusden, Dr Philip Onyebujoh (Advisor to the Africa CDC), Dr Jennifer Gardy (Deputy Director, Surveillance, Data, and Epidemiology, Malaria, Global Health, Bill and Melinda Gates Foundation) and Alan Christoffels.

EXPERT PANEL OR COMMITTEE MEMBERSHIP

Name	Membership and Purpose	Role
Dominique Anderson	ISCB-Africa 2019 Conference, Ghana	Scientific Committee Chair
	DSI-medical collections platform	Co-Chair
Alan Christoffels	African Society for Bioinformatics and Computational Biology (**http://www.asbcb.org**).	Governing Council member
	ISCB-Africa 2019 Conference, Ghana	Co-Chair
	TBGenAfrica H3Africa funded TB consortium	Scientific Advisory Board Member
	PHINDAccess EU-funded project in Tunisia to build capacity in Host-pathogen interaction omic analysis	Scientific Advisory Board Member
	East African Bioinformatics Network Fogarty Funded Bioinformatics Network in East Africa	Advisory Board Member
	Global Emerging Pathogens Consortium to create awareness of biosecurity threats across Africa, and to empower African scientists to respond to disease outbreaks	Co-Chair the annual conference and develop training modules
	African Academy of Sciences to develop guidelines for biospecimen and data management in Africa	Biospecimen and Data Governance Expert Panel
	NRF Rating Panel	Assessor
	Center for Proteomics and Genomics Research	Advisory Board Member
	Department of Science and Technology Genomic and Precision Medicine Think Tank to develop a strategic framework for personalized medicine in South Africa	Member
	National Health Laboratory Services: biospecimen and data access ethics	Review guidelines for data sharing
Ravnit Grewal	WHO Strategic Advisory Group of Experts (SAGE) on In vitro Diagnostics Essential Diagnostic List.	Reviewed the Haematology related diagnostic tests to be implemented in middle to low income countries
Peter van Heusden	PHA4GE Infrastructure Working Group	Chair

JOURNAL EDITING AND REVIEWS

PI Name	Journal
Alan Christoffels	Editor: The Improper Scientist: Quarterly magazine for the African Society for Bioinformatics and Computational Biology
	Editorial Board Member: Data Journal
	Reviewer: BMC Genomics, BMC Bioinformatics, Bioinformatics, PLoSONE
Hocine Bendou	Reviewer: Oxford Academic Bioinformatics
Gordon Harkins	Reviewer: AIDS Research and Human Retroviruses Transboundary and Emerging Diseases

THESIS EXAMINED FOR STUDENTS AT OTHER INSTITUTIONS

Examiner	Institution	Degree
Alan Christoffels	Fort Hare	MSc
	University Pretoria	PhD
Ruben Cloete	Rhodes University	MSc

EXTERNAL MODERATION

Name	Institution	Course
Alan Christoffels	Stellenbosch University	Bioinformatics Modules for Biochemistry Department: 3rd year and Honours modules
Gordon Harkins	Stellenbosch University	CBIO Honours Course

RESEARCH PROJECT THEMES

SANBI's vibrant research environment is reflected in the diversity and number of projects that are underway.

CURRENT PROJECTS AND THEIR SUBPROJECTS

PI	Theme	Projects
Dominique Anderson	Biobank informatics	▪ Biobank informatics - multiple collaborations for biocollection sample quality management and e-infrastructure with a focus on Baobab LIMS.
	Biomedical data governance	▪ Collaboration with biobank and law experts to evaluate data governance at tertiary institutions and biomedical research facilities.
	Dental Genetics	▪ Dental genetics in collaboration with the Faculty of Dentistry investigating oral microbiomes.
	OneHealth	▪ Projects to start in 2020
Hocine Bendou	Cancer Bioinformatics	▪ Novel genomic biomarkers for Pediatric and Adult Acute Myeloid Leukemia. ▪ Computational analysis of multi-omic data for the elucidation of molecular mechanisms of neuroblastoma.
	Pipeline development	▪ Developement of a Nextflow pipeline for the detection of enhancer-associated noncoding variants using ChIP-seq data.
	Machine learning	▪ Prediction of synthetic lethal gene interaction in human cancer using supervised machine learning.
	Gene regulatory networks	▪ Gene regulatory network study of diffuse large B-cell lymphoma.
	Viral Study	▪ Identify impact of viral inserts in human esophageal cancer. A collaborative work with the University of Cape Town.
Alan Christoffels	Pathogen genomics	▪ In collaboration with the Tygerberg MRC Unit, we are developing methods to analyse high throughput sequencing data for microbial genomes. ▪ Identification of novel drug targets in pathways known to contain drug resistant genes. ▪ Predicting the interaction networks between human and mycobacteria ▪ Identification of operons & non-coding RNAs to understand gene regulation in pathogen genomes.
	Pathogen surveillance	▪ Developing systems to support reproducible analyses ▪ Development of global standards for data sharing and integration ▪ Public Health Alliance for Genomic Epidemiology (PHA4GE)
	Drug discovery	▪ Identifying drug molecules against multi-drug resistant *M.tuberculosis* ▪ In partnership with School of Pharmacy at UWC, delineate the pharmacokinetics of drugs in response to host genetic factors ▪ In partnership with School of Pharmacy at UWC, use of Machine learning techniques to prioritize molecules in natural compounds.
Ruben Cloete	Molecular modelling and drug design	▪ NGS sequencing of tuberculosis isolates and HIV plasma samples allowed the identification of novel mutations. To identify mutations associated with drug resistance we develop structural computational pipelines to further our understanding of HIV-1 and tuberculosis drug resistance. ▪ Previous pathway mapping and gene prioritization methods allowed us to identify known and novel *Mycobacterium tuberculosis* drug targets. Here we exploit the drug targets using virtual screening methods to identify novel drugs to treat drug resistant tuberculosis. The drugs identified are tested for activity against *Mycobacterium tuberculosis* using a bioluminescent reporter assay system.

Ruben Cloete	Molecular modelling and drug design	▪ Whole genome sequencing identified variants that co-segregate within South African families with Parkinson's disease. Here we use various software tools to prioritize these variants to identify causal variants that might be associated with Parkinson's disease. ▪ Using machine learning methods to identify putative compounds with activity against *Mycobacterium tuberculosis*. ▪ To identify drug resistance pathways in HIV-1 using Phylogenetics. ▪ Develop structural methods to aid in the identification of biomarker discovery during tuberculosis infection and DNA aptamer identification for diagnostic purposes.
Ravnit Grewal	Precision Medicine	▪ The role of precision medicine in the diagnosis of HIV related lymphomas.
Gordon Harkins	HIV Immunology	▪ Studies of heterologous and cross-neutralizing antibody responses, antibody binding, IgA/IgG levels, antibody-dependent cellular cytotoxicity (ADCC), viral epitopes and evolution, and host factors including HLA-I alleles to determine whether correlates of protection are associated with clinical outcome in natural infection
	HIV Dynamics	▪ Investigation of the properties of viruses replicating in the female genital tract, and whether these properties differ from those replicating in the blood
	Virulence Evolution	▪ Investigation of how virulence and pathogenicity are evolving in viral crop pathogens
	Viral Emergence	▪ Investigation of 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
	Endogenous Viral Elements	▪ Investigation of the role of EVE's in virus evolution and their functions in plants to address fundamental biological issues that are critical to the understanding of plant virus evolution
	Phytovirus Diversity	▪ An analysis of the phytoviral species richness of plant communities within several unmanaged and managed ecosystems
	Rift Valley fever virus epidemiology	▪ Phylogeographic analyses of Rift Valley fever virus complete genomes to investigate the epidemiological history of these viruses circulating in South Africa and other African countries.
Peter van Heusden	Research Computing Infrastructure	▪ Studying the deployment of research computing infrastructure and development of associated skills on the African continent.
	Pathogen surveillance Platforms	▪ Researching the use of pathogen genomics in a public health context for drug resistance prediction and genomic surveillance.
	Scientific workflows for pathogen genomics	▪ Developing workflow systems for pathogen genomics with a focus on Galaxy.

RESEARCH LABORATORIES

Members Of The Anderson Lab

HIGHLIGHTS OF THE LAB

- Co-supervisor of 1 MSc student
- Supervisor of 1 PhD student
- Funding grant awarded: Technology Innovation Agency Seed fund award.
- Invited talk: Dominique Anderson. Biological sample management and community engagement for biorepositories. Noguchi Institute, Ghana. 14 November 2019.
- UWC Technology Transfer Office exhibited Baobab LIMS at the Innovation Bridge held in December 2019, as one of the many game-changing technologies UWC has helped produce.

RESEARCH PROJECTS

The research group is investigating several areas with an interest in dental metagenomics, data security and privacy, biomedical data governance, informatics of biobanking and One Health. For two and a half years, I have been part of a team that has focused on informatics solutions for biobanking with continued development, enhancement and training in the open-source Baobab laboratory information management system. Information management and quality management remains a core area of activity in my research group with hopes to expand the collaborative network of developers and researchers. In relation to my interest in biobanking, the group is also involved in community engagement and effective science communication and training.

With regards to data privacy and governance, my collaborators and I combine expertise in biobanking, bioinformatics and big data, and law to investigate how both regulations and best practices in ICT influence the biomedical data and data sharing landscape in Africa.

Research Laboratory of Dominique Anderson

The University of the Western Cape Dentistry faculty and SANBI have also started a collaborative effort to increase research into dental genetics. In the future, I hope to establish networks within the agricultural, veterinary, medical, and environmental sphere in order to intensify research in the area of One Health.

RESEARCH COLLABORATIONS

1 - Data Governance For Biobanks

Collaborating Parties:
Dr Carmen Swanepoel – NHLS/Stellenbosch University Tygerberg
Dr Ciara Staunton – School of Law, Middlesex University, UK

Nature and purpose:
To investigate the impact of local and international regulation on the sharing of biomedical research data in Africa.

Output in the last 12 months:
One publication:
Ciara Staunton, Rachel Adams, Dominique Anderson, Talishiea Croxton, Dorcas Kamuya, Marianne Munene and Carmen Swanepoel. *Protection of Personal Information Act 2013 and data protection for health research in South Africa.* International Data Privacy Law. Jan 2020. https://doi.org/10.1093/idpl/ipz024

Future Direction:
Policy development for ethical cross-border material and data sharing in health research.
Research data management for big data in health.

2 - Dental Genetics

Collaborating Parties:
Dr Manogari Chetty – Department of Dentistry, UWC

Nature and purpose:
Incorporating molecular biology, genetics and bioinformatics into dentistry research.

Output in the last 12 months:
None

Future Direction:
Dental metagenomics.
Oral cancer genomics.
Impact of dental health in disease.

3 - Biobank Informatics and Community Engagement

Collaborating Parties:
Prof Alan Christoffels - SANBI, UWC
Dr Carmen Swanepoel - NHLS/Stellenbosch University Tygerberg

Nature and purpose:
Enhancement and customization of Baobab LIMS for LMICs. Standardization of e-infrastructure. Science communication and community engagement.

Output in the last 12 months:
TIA seed fund grant award.
West Africa Health Organisation/Institute Pasteur Côte d'Ivoire LIMS customization.

One publication:
Anja Bedeker, Dominique Anderson, Thoba Lose, Yamkela Mgwatyu, Retha Luus, Renete Blignaut, Alan Christoffels. *Understanding biobanking: An assessment of the public engagement speaking book intervention Biobanking and Me.* South African Journal of Bioethics Law. Dec 2019

Future Direction:
New collaborations within Africa as well as training initiatives on the continent. Maintenance and enhancement of Baobab LIMS. Community engagement initiatives.

Members Of The Bendou Lab

HIGHLIGHTS OF THE LAB

Student supervision:

Abdulazeez Giwa	PhD, SANBI
Nasr Eshibona	PhD, SANBI
Wardah Jassiem	MSc, SANBI
Nomlindelo Mfuphi	MSc, SANBI
Victoria Patten	PhD, UCT

Grants application
NIH grant on Diffuse Large B-Cell Lymphoma.
Self-initiated research SAMRC grant on pancreatic adenocarcenoma.

RESEARCH PROJECTS

As a new research group, our interests lie in the area of genomics focusing on identification of genetic abnormalities driving the genesis and progression of human cancer, using bioinformatics.
The group is working on developing bioinformatics pipelines, and applying RNA-Seq and Whole Genome Sequencing (WGS) methods in analysing different types of cancer including; Neuroblastoma, Acute Myloid Leukemia and Diffuse Large B-Cell Lymphoma.

The group works with external collaborators to identify the role of integrated viral DNA in oesophageal cancer by using WGS combined with ERVCaller tool.

Research Laboratory of Hocine Bendou

RESEARCH COLLABORATIONS

1 - Differential Expression Protein On Unexplained Male Infertility

Collaborating Parties:
Prof Ashok Agarwal – American Center for Reproductive Medicine, Cleveland Clinic, Cleveland
Dr Hocine Bendou – SANBI, UWC

Nature and purpose:
Provide bioinformatics and statistics expertise.

Output in the last 12 months:
Sperm Proteome analysis and identification of fertility-associated biomarkers in unexplained male infertility.

2 - Molecular Biology and Genetics of Oesophageal Cancer

Collaborating Parties:
Prof Iqbal Parker – UCT
Dr Hocine Bendou – SANBI, UWC

Nature and purpose:
Provide bioinformatics expertise and PhD student co-supervision.

Output in the last 12 months:
Ongoing work on viral DNA integrated in human oesophageal cancer.

Future Direction:
Whole genome sequencing analysis on paired tumor-normal samples from 60 participants.
Prepare and publish a manuscript.

Research Laboratory of Alan Christoffels

Members Of The Christoffels Lab

HIGHLIGHTS OF THE LAB

The research and development work on COMBAT-TB has cemented our annual contribution to the International Galaxy Consortium Developers' Conference. The International Galaxy consortium has annual developers' conferences alternating between USA and Europe. We launched the first Galaxy meeting outside of USA and Europe called the GalaxyAfrica conference (**galaxyafrica.sanbi.ac.za**).

In 2019 my MSc student, Eugene de Beste, completed a thesis with distinction on moving code to the data using openstack **(https://bit.ly/2qjPCtW** and **https://github.com/Banshee1221/Nikeza.)**. This work has formed part of new projects in Africa to strengthen public health institutes responsible for managing pathogen genomic data.

In December 2019, my MSc student Jamie Southgate submitted his thesis "Development of an African Genomics Archive". This work represents a prototype of what a regional pathogen data archive could look like. We published a novel implementation of a Neo4j as a backend to our COMBAT-TB project.
The above-mentioned developments in data storage, pathogen analysis and movement of code via the Cloud provided the basis for a partnership with the Africa CDC in Ethiopia and saw the establishment of Phase1 of a pathogen genomics initiative based at the Africa CDC to strengthen public health institutes on the African content. By October 2019, we launched a global initiative called the Public Health Alliance for Genomic Epidemiology (PHA4GE) at the GrandChallenges meeting in Addis, Ethiopia with partners including Africa CDC, Oxford University, Washington University, and the University of Melbourne, USA CDC, University of Birmingham UK, University of British Columbia and BC Centre for Disease Control, the Broad Institute in Boston, and H3Africa. The secretariat for this programme is based at SANBI-UWC.

A total of 6 students graduated in 2019 namely: 2 Hons students (Ridaa Fredricks and Farzaana Diedricks), 2 MSc students (Eugene de Beste and Ajayi Olabode Oluwaseum) and 2 PhD students (Hocine Bendou and Mqondisi Tshabalala).

RESEARCH PROJECTS

Tuberculosis

A comprehensive research programme is underway that includes investigating genetics 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) (**www.combattb.org**) initiative to deploy analytic tools across the African continent.

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 have published new predictive models using machine learning techniques 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 with a long term vision. To this end we have contributed to 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 (**www.baobablims.org**) that is being rolled out across Africa.

HLA Typing

In collaboration with Professor Michael Pepper at the University of Pretoria we are analysing the HLA data for African populations. There remains an important link between HLA diversity and susceptibility to infectious disease. This work will expand with reference to host susceptibility to tuberculosis.

RESEARCH COLLABORATIONS

1 - Prediction Of Human-Tuberculosis Interaction Networks

Collaborating Parties:
Prof Peter Witbooi - Mathematics Department, UWC.

Nature and Purpose:
Use machine-learning techniques to improve the prediction of host-pathogen interactions between human and pathogens.

Output in the last 12 months:
Resources has been moved to predicting genetic determinates of multi-drug resistance in M.tb.

2 - Develop A Biobank Informatics Management System

Collaborating Parties:
Dr Dominique Anderson – SANBI, UWC; Lagos Biobank, Nigeria; USA Center for Diseases and Prevention.

Nature and Purpose:
Development of a laboratory management system for infectious disease labs that will integrate with other biobanking tools.

Output in the last 12 months:
We have extended the functionality of Baobab LIMS for utility in the Abidjan Biobank.

3 - Analyzing Genetic Networks In *M.tuberculosis*

Collaborating Parties:
Prof Rob Warren - University of Stellenbosch
Dr Hocine Bendou - SANBI, UWC

Nature and Purpose:
Identification of Operons: We rewrote an algorithm previously published by a team at Johns Hopkins University.
Identification of non-coding RNAs.

Output in the last 12 months:
Our operon detection algorithm has been tested with different metrics. And currently benchmarked against two international methods.

Using a sequence conversation approach and a RNA folding approach, we are identifying putative small RNAs.

4 - Identification Of Novel Drug Targets For Drug Resistant Tuberculosis

Collaborating Parties:
Prof Samantha Sampson - University of Stellenbosch
Prof Rob Warren - University of Stellenbosch
Dr Cedric Wereley - University of Stellenbosch
Prof Sarel Malan - School of Pharmacy, UWC

Nature and Purpose:
Current TB drugs are more than 30 years old and have unacceptable efficacy and safety profiles, emphasizing the need for new drugs. In this regard we published a paper on kinase inhibitors in *M.tuberculosis*, and efflux pump inhibitors.

Output in the last 12 months:
2 compounds were identified as potential inhibitors and experimental assays confirmed their role as inhibitors of mycobacterial growth. We also identified compounds that bind more strongly than current first line drugs (isoniazid and Rifampicin). Both papers were published in 2019.

5 - Chemo-Informatics Profiling Of Plant Extracts That Show Anti-Plasmodial Activity

Collaborating Parties:
Prof Sarel Malan and Dr Samuel Egieyeh - 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:
Recruited 3 postgrad students for this project. Our research to date was summarized in two book chapters.

6 - Computational Bacterial Analytical Toolkit For Tuberculosis (Combat-TB)

Collaborating Parties:
Prof Samantha Sampson - US
Prof Rob Warren - US.

Nature and Purpose:
Develop a computational platform to store tuberculosis omic data and to provide a visualisation tool.

Output in the last 12 months:
Graph database was published in Bioinformatics Journal. Extended the COMBAT-TB toolkit.

7 - Exploiting Protein Signatures In Colon Cancer Archival Biospecimens

Collaborating Parties:
Prof Schneider, 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 in the last 12 months:
Submitted manuscript on optimization of experimental protocols for extracting protein from FFPE samples. Thesis submitted.

8 - Identification Of SNPs Implicated In Rare Diseases

Collaborating Parties:
Prof Manogari Chetty – Dentistry Faculty, UWC.

Nature and Purpose:
We aim to utilise bioinformatics tools to analyse the genetic basis of some of the syndromes observed in the dental clinics.

Output in the last 12 months:
Work on Osteogenesis Imperfecta presented at 2nd Arab Association of Genetic Research Conference, Egypt.

9 - Biosecurity Preparedness In Africa

Collaborating Parties:
Dr Dominique Anderson, SANBI, UWC; Global Consortium on emerging infectious diseases (**www.getafrica.org**).

Nature and Purpose:
We aim to establishing infrastructure and research capacity to respond to highly infectious emerging pathogens.

Output in the last 12 months:
We have reviewed the information security policies at SA universities with a view to propose guidelines for data management that impacts biomedical researchers. Co-Chaired the GET conference in Abuja, Nigeria.

10 - Public Health Alliance For Genomic Epidemiology (Pha4ge)

Collaborating Parties:
Africa CDC, Oxford University, Washington University, University of Melbourne, USA CDC, University of Birmingham UK, University of British Columbia and BC Centre for Disease Control, Broad Institute in Boston, and H3Africa.

Nature and Purpose:
The PHA4GE consortium was established to bring Bioinformatics closer to Public Health - to build data standards.

Output in the last 12 months:
The consortium was launched in October 2019 and is funded by the Bill & Melinda Gates Foundation.

11 - African Genomics Archive

Collaborating Parties:
Africa CDC

Nature and Purpose:
To establish a data archive that can be used to store and retrieve pathogen data generated in Africa with a view to strengthen public health institutes to improve disease surveillance.

Output in the last 12 months:
We completed a prototype to move code to data. The manuscript in prep and code available on github. Prototype for a genomic archive nearly completed.

Research Laboratory of Ruben Cloete

Members of the Cloete Lab

HIGHLIGHTS OF THE LAB

Students:
This past year saw the lab growing. Masters student, Maryam Hassan joined my lab.
Two students who I co-supervised at Stellenbosch University, graduated: Boiketlo Sebate, MSc and Adetayo Emmanuel Obasa, PhD.
Additionally, I co-supervised PhD student Given Mikasi at Stellenbosch University.

Grants/Funding Awarded:
The lab was awarded funding by the Poliomyelitis Research Foundation for a project "Understanding HIV-1 subtype C integrase drug resistance to provide novel treatment strategies for South African HIV-1 infected patients".

RESEARCH PROJECTS

My primary research interests focus on understanding drug resistance in *Mycobacterium tuberculosis* and Human Immune deficiency virus (HIV-1) integrase protein and the identification of causal variants in Parkinson's disease development. Tuberculosis and HIV-1 drug resistance is a huge problem in South Africa and calls for the identification of newer drugs to curb the spread of these diseases. The use of computational methods to speed up the process of drug discovery can reduce the cost and time spent pursuing drugs that later fail in clinical trials.

We have developed computational pipelines within my laboratory to interrogate drug targets to screen for drugs that inhibit enzyme targets that can be purchased and tested in vitro for activity against *Mycobacterium tuberculosis*.

We also use computational methods to validate mutations and their effects on protein structure and function. This is also very useful to determine if a drug remains within an enzyme active pocket and if the mutation results in reduced binding or no binding.
The result of this research may provide improve treatment regimens for South African populations to combat infectious disease and non-communicable diseases like Parkinson's.

RESEARCH COLLABORATIONS

1 - Novel Drug Discovery Against *Mycobacterium Tuberculosis*

Collaborating Parties:
Prof Alan Christoffels – SANBI, UWC
Dr Melanie Grobbelaar – Stellenbosch University
Prof Samantha Samson – Stellenbosch University

Nature and purpose:
To identify drugs with a new mode of action against *Mycobacterium tuberculosis* protein targets.

Output in the last 12 months:
One published article.
Structure based identification of novel inhibitors against ATP synthase of *Mycobacterium tuberculosis*: A combined in silico and in vitro study. Mohd Shahbaaz, Ruben Cloete, Melanie Grobbelaar, Samantha Sampson, Alan Christoffels. International Journal of Biological Macromolecules 135 (2019) 582–590.

Future Direction:
Rv2421c manuscript in preparation.

2 - Structural Impact Of Resistance Associated Mutations In The South African HIV-1C Integrase Protein

Collaborating Parties:
Dr Graeme Jacobs – Stellenbosch University

Nature and purpose:
Firstly, to understand genetic diversity in HIV-1 subtype C integrase gene in South African HIV-1 infected patients and recombinant subtype AG in Cameroonian patients. Secondly, to determine if second-line integrase inhibitors will be a viable option for South African and Cameroonian patients infected with HIV-1.

Output in the last 12 months:
Sello Given Mikasi, Josiah Otwoma Gichana, Cheri Van der Walt, Dominik Brado, Adetayo Emmanuel Obasa, Duncan Njenda, Martha Messembe, Emilia Lyonga, Okomo Assoumou, Ruben Cloete, George Mondinde Ikomey, and Graeme Brendon Jacobs. AIDS Research and Human Retroviruses, 2020. ahead of print: http://doi.org/10.1089/aid.2019.0264

Future Direction:
Two more manuscripts in preparation.

3 - Prioritising Mutations Identified In South African Parkinson's Disease Patients Using Structural Methods

Collaborating Parties:
Prof Soraya Bardien – Stellenbosch University

Nature and purpose:
To identify novel genes associated with Parkinson's disease development using Whole Exome sequencing and using structural computational methods to understand the impact of mutations on protein structure and function.

Output in the last 12 months:
None

Future Direction:
Manuscript in preparation.

4 - Identification Of *Mycobacterium Tuberculosis* Efflux Pump Inhibitors

Collaborating Parties:
Prof Sarel Malan – School of Pharmacy, UWC
Prof Jacques Joubert – School of Pharmacy, UWC
Miss Erika Kapp – School of Pharmacy, UWC

Nature and purpose:
To identify alternative drugs to verapamil and piperine to inhibit *Mycobacterium tuberculosis* efflux pump Rv1258c and restore sensitivity of *Mycobacterium tuberculosis* to Rifampicin.

Output in the last 12 months:
None

Future Direction:
Manuscripts in preparation.

Research Laboratory of Junaid Gamieldien

Junaid Gamieldien

HIGHLIGHTS OF THE LAB

2 publications in 2019:

FTO haplotyping underlines high obesity risk for European populations. Babenko, Vladimir, Babenko, Roman, Gamieldien, Junaid & Markel, Arcady. BMC Medical Genomics. 12. 10.1186/s12920-019-0491-x.

Ten simple rules for organizing a webinar series Fadlelmola Faisal M., Panji Sumir, Ahmed Azza E., Ghouila Amel, Akurugu Wisdom A., Entfellner Jean-Baka Domelevo, Souiai Oussema, Mulder Nicola, **Gamieldien Junaid** et al. PLOS Computational Biology 15(5): 10.1371/journal.pcbi.1006671

1 MSc *Cum Laude* Graduation:
Fanechka Esterhuyzen, Thesis Title: Development of a simple artificial intelligence method to accurately subtype breast cancers based on gene expression barcodes.

RESEARCH PROJECTS

Graph Knowledgebases For Disease Genomics

Our oldest project, the BioOntological Graph Relationship (BORG) database, assimilates and integrates multiple disparate sources of genomic and biomedical knowledge and metadata into a unified knowledge graph. We have successfully used it to model and learn the phenotypic features and molecular mechanisms associated with diseases as a means to rapidly prioritise novel candidates identified in genomics studies. The system simulates search paradigms, cognitive processes and biological 'rules' applied by clinical geneticists and biomedical researchers when filtering candidates.
The knowledgebase has proven especially useful in our Next Generation Sequencing (NGS) work focused on tendinopathy.

Development Of An NGS Diagnostic Framework For 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 and have identified candidate mutations in South African patients, several of which are not in typical MODY genes.

Exome and genome sequencing for discovery of candidate causative variants in human diseases

We have several disease NGS projects that benefit from customised versions of our variant discovery pipelines and BORG database:

- Muscular dystrophy
- Motor neuron disease/ALS
- Unknown neuromuscular disorders
- Achilles tendinopathy and cruciate ligament ruptures

Transcriptomics profiling to identify circulating biosignatures

- For detection of smoke inhalation injury in burns victims
- For early prediction of likely rapid progressors in cystic fibrosis patients

Machine Learning

We apply established machine learning algorithms and latest advances in artificial intelligence approaches to high-value publicly available big biomedical data establish models classification and prediction. Both supervised and unsupervised learning approaches are used to identify features and patterns that can be used to predict clinical endpoints and also potentially novel pathways/networks that contribute to a phenotype of interest. Existing work that has reached a degree of maturity include the use of microarray and RNA sequencing data from repositories such as The Cancer Genome Atlas and the Gene Expression Omnibus to develop models capable of predicting e.g. rate of disease progression and survival in glioblastoma and disease subtype in breast cancer. We are expanding our research to include the use of Computer Vision and Deep Neural Networks to build predictive/explanatory models from both omics and biomedical imaging data.

Research Laboratory of Ravnit Grewal

Ravnit Grewal

As a Haemato-Pathologist my primary research interest is in HIV related lymphomas. As a consequence of this enquiry, I spearheaded a research group that studied lymphomas as it became evident that we were experiencing a surge in HIV associated complex lymphoma pathology in our laboratory. As people with HIV on Antiretroviral treatment live longer, research nationally and internationally has demonstrated a rise in certain lymphomas that occur as a consequence of infection with HIV, amongst other pathogenetic mechanisms. Understanding the pathophysiology of these lymphomas in our population specifically is of paramount importance as oncology management practices move towards targeted therapy.

Thus and in keeping abreast with current international research I aim to investigate the role of Precision/ Personalised Medicine in Pathology, specifically focusing on the detection of biomarkers for diagnosis, prognostication and management of HIV related lymphomas in South Africa and other African countries.

Our first aim is to setup an African Precision Oncology group with a mission to identify the different molecular (genomic, proteomic) profiles within cancer types, and assess the effectiveness of proposed drugs and therapies. Thus providing African pathologists with clinical decision tools to classify patient's tumor by molecular profile, and provide the most suitable existing therapy. We have expanded the group with a bioinformatics scientist who is supervising two students on algorithms for precision oncology.

HIGHLIGHTS OF THE LAB

Student supervision

- 2 PhD students (self-funded) currently working on algorithm development with application to cancer – supervised by Dr Bendou
- 1 PhD student studying Oesophageal cancer at the University of Cape Town – co-supervised by Dr. Bendou
- 1 Mmed student registered at Stellenbosch University – co-supervised by Dr Grewal
- Identified 2 new MSc students for 2020 – Dr Grewal
- 2 postgrad Graduations: 1 Mmed student, University of Stellenbosch with thesis title “The diagnostic utility of flow cytometry on finer needle aspirate (FNA) samples at Tygerberg Hospital (TBH)” and 1 MSc in Dentistry, UWC with thesis title “Incidence of plasmablastic lymphoma in HIV positive patients at a tertiary hospital in South African(2005-2017)”.

Collaborations

- Established collaboration with international researchers at the University of Washington, Seattle, USA.
- Ongoing collaborative work with the Molecular Biology and Genetics of Oesophageal Cancer at UCT.
- Participation in a successful international collaborative work with the Cleveland Clinic Lerner College of Medicine in USA. In this research study we were able to find new biomarkers (proteins) in unexplained male infertility. The project was managed by Professor Ashok Agarwal, one of the known scientists in the field of translational research in human infertility and assisted reproduction.

Networking Opportunities

- Dr Grewal was recently invited to serve on an advisory group of Experts on In vitro diagnostics for the World Health Organisation. This will be a unique opportunity to extend the networking with regard to Precision Oncology.
- Dr Grewal has initiated collaborations with:
  - University of Washington, Seattle, USA –co-applicants on NIH grant.
  - Basel University, Basel, Switzerland - Reciprocated visits and planning a precision oncology course to be held in Cape Town.

Outputs

- Panner Selvam, M.K., Agarwal, A., Pushparaj, P.N., Baskaran, S., and Bendou, H. Sperm Proteome Analysis and Identification of Fertility-Associated Biomarkers in Unexplained Male Infertility. Genes 10, 522.
- Cristina Bagacean, Ciprian Tomuleasa, Adrian Tempescul, Ravnit Grewal, Wesley H Brooks, Christian Berthou & Yves Renaudineau Apoptotic resistance in chronic lymphocytic leukemia and therapeutic perspectives (2019): Critical Reviews in Clinical Laboratory Sciences, 0.1080/10408363.2019.1600468Apoptotic resistance in chronic lymphocytic leukemia and therapeutic perspectives.
- Dr Grewal is a member of the WHO Strategic Advisory Group of Experts (SAGE) on In vitro Diagnostics (2019). Essential Diagnostic List (Reviewed the Haematology related diagnostic tests to be implemented in middle to low income

countries. The book has now been published). **https://www.who.int/medical_devices/publications/Standalone_document_v8.pdf?ua=1**

- Invited speaker: Second meeting of the Strategic Advisory Group of Experts on In Vitro Diagnostics (SAGE IVD)
- 18 - 22 March 2019, WHO HQ, Geneva: Invited as an expert reviewer for the following haematology related diagnostic tools:
- Presentations Given:
    1. Plasma and urine protein electrophoresis and immunofixation
    2. Essential panel of antibodies for flow cytometry for leukemia
    3. Basic panel with Immunohistochemical testing (IHC) for lymphoma
    4. Numeration of lymphocyte subtypes, namely CD4 (as in HIV), CD8, CD20, CD16/56 cells
- Grant Application: July 2019, NIH-MRC grant, RFA-AI-19-022, for a project titled : The relationship between genetic aberrations, transcriptomic profiles as well as the microenvironment of Diffuse Large B-cell lymphoma (DLBCL) in South African HIV patients.

Research Laboratory of Gordon Harkins

My research primarily focuses on the evolution and molecular epidemiology of single-stranded DNA and RNA viral pathogens of humans other animals and plants. I seek to computationally 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.

Gordon Harkins

HIGHLIGHTS OF THE LAB

Four manuscripts were submitted for publication in 2019 three of which were published within the calendar year and the fourth in 2020. Also in 2019 our paper entitled "Geometagenomics illuminates the impact of agriculture on the distribution and prevalence of plant viruses at the ecosystem scale" that was published in 2018 in the ISME Journal earned the "Editors choice" award as one of the most cited and most shared. The article has been accessed 1931 times and cited 28 times and is in the 97th percentile (ranked 9,078) of the 318, 355 tracked articles of a similar age in all journals and in the 93rd percentile (ranked fourth) of the 62 tracked articles of a similar age in the ISME Journal. A summary of some of the **research projects** that my lab has been involved in during 2019 is presented below.

HUMAN-INFECTING VIRAL PATHOGENS

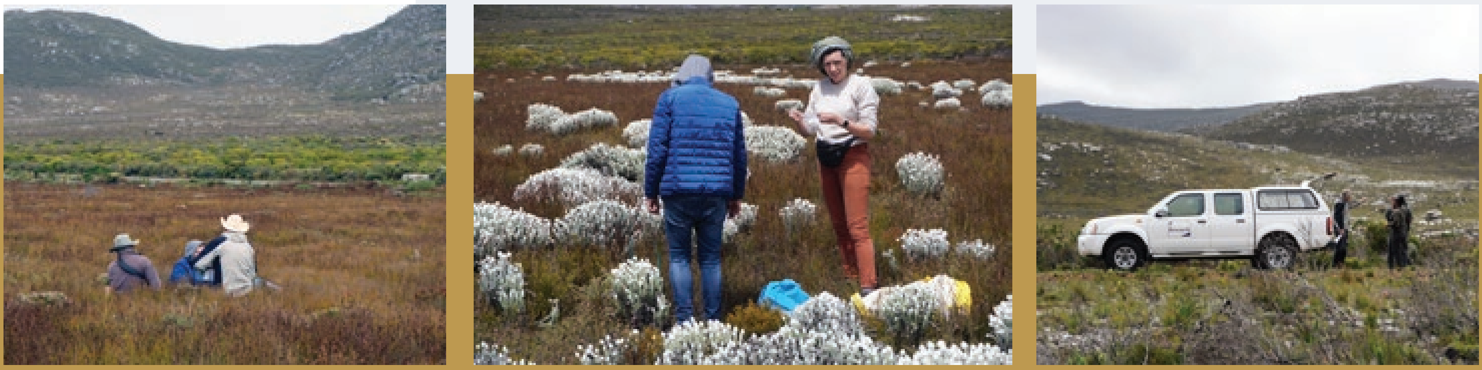
HUMAN IMMUNODEFICIENCY VIRUS

Immunological Responses And Their Impact On Clinical Outcomes In Epidemiologically Linked HIV-1 Transmission

In collaboration with researchers from New York University School of Medicine, Johns Hopkins University and the University of KwaZulu-Natal, South Africa among others; we studied four HIV-1+ Cameroonian individuals, three of them epidemiologically linked in a polygamous heterosexual relationship and one incidence-matched case, over 15 years for heterologous and cross-neutralizing antibody responses, antibody binding, IgA/IgG levels, antibody-dependent cellular cytotoxicity (ADCC), viral epitopes and evolution, and host factors including HLA-I alleles to determine whether correlates of protection are associated with clinical outcome in natural infection.

Evidence For A Mucosal Sieve Effect Among Different Anatomical Tissues In HIV Infection

Because most HIV-1 infections in sub-Saharan Africa are acquired heterosexually through the genital mucosa, understanding the properties of viruses replicating in the female genital tract, and whether these properties differ from those replicating in the blood, is important. In collaboration with researchers from the National



Gordon Harkins and team searching for plant viruses in the field at Cape Point

Institute for Communicable Diseases (NICD) and the University of Cape Town (UCT), we showed in our published findings that in some women, distinct viral populations exist in the female genital tract and the blood that may impact the efficacy of microbicides and vaccines designed to provide mucosal immunity.

PLANT-INFECTING VIRAL PATHOGENS

Virulence Evolution

This is a long-term collaborative effort 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. 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. By producing infectious clones of ancient genomes from herbarium specimens 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. 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 that shape the diversity in an important group of crop pathogens.

Viral Emergence

In collaboration with researchers from CIRAD France, the State University of Arizona and UCT, 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.

Phytovirus Diversity

Emergent diseases of plants, a high proportion of which are caused by phytoviruses, are a significant burden on the food security and economic stability of society. However, a comprehensive view of the geographical distribution of phytovirus diversity does not exist, including both the numbers or richness of virus species and the evenness of their distribution in any individual environment on Earth. In collaboration with researchers from CIRAD France, the State University of Arizona and UCT we received funding from the French National Research Agency (ANR) in 2019 for a project that had three primary objectives (i) test whether plant species richness influences phytovirus species richness in natural and cultivated areas (ii) experimentally study the effect of plant communities on phytoviral species richness (iii) search for evolutionary footprints associated with emergence within phytovirus genomes.

ANIMAL-INFECTING VIRAL PATHOGENS

Rift Valley Fever Virus Epidemiology

Rift Valley fever is a mosquito-borne zoonotic disease that affects humans and some species of ruminants and is responsible for major losses in livestock production in sub-Sahara African countries. The causative agent, Rift Valley Fever Virus belongs to the genus Phlebovirus in the family Bunyaviridae. Outbreaks of the disease in South Africa have occurred regularly every 10-15 years since 1943. Between 2008 and 2010 outbreaks were reported in all nine provinces with 232 confirmed infections in humans, and 26 confirmed deaths. We are investigating the epidemiological history of these viruses circulating in South Africa and in other African countries. The results of this study have the potential to inform governmental policy aimed at restricting the movements of infected livestock across national borders.

RESEARCH COLLABORATIONS

1 - Immune Correlates Of Disease Progression In Linked HIV-1 Infection

Collaborating Parties:
Ralf Duerr Michael Tuen, Jude S. Bimela, Andrew N. Banin; Sonal Soni; Luzia Mayr; Aubin J. Nanfack; Miroslaw K. Gorny and Phillipe N. Nyambi - New York University School of Medicine, New York, New York, USA. Jude S. Bimela; Judith N. Torimiro - University of Yaoundé, Yaoundé, Cameroon.
Shilei Ding - Centre de Recherche du CHUM, and Université de Montréal, Montréal, QC, Canada.
Vincenza Itri - Icahn School of Medicine at Mount Sinai, New York, USA.
Allison R. Durham - National Institutes of Health-National Institute of Allergy and Infectious Diseases, Bethesda, Maryland, USA.
Stephen F. Porcella - Rocky Mountain Laboratories, Division of Intramural Research, NIAID, NIH, Hamilton, Montana, USA. Josephine Meli - Yaoundé General Hospital, Yaoundé, Cameroon.
Marcel Tongo - Institute of Medical Research and Study of Medicinal Plants (IMPM), Yaoundé, Cameroon and Nelson R Mandela School of Medicine, University of KwaZulu-Natal, South Africa.
Xiaohong Wang - Veterans Affairs New York Harbor Healthcare Systems, New York, USA. Xiang-Peng Kong - Arthur Nádas New York University School of Medicine, New York, USA. Daniel E. Kaufmann - The Scripps Research Institute, La Jolla, California, USA.
Zabrina L. Brumme - Simon Fraser University, and British Columbia Centre for Excellence in HIV/AIDS, St. Paul's Hospital, Vancouver, Canada.
Thomas C. Quinn; Andrew D. Redd -Department of Medicine, Johns Hopkins University, Baltimore, USA.
Andrés Finzi - McGill University, Montréal, Canada.

Nature and purpose:
To determine whether correlates of protection are associated with clinical outcome in natural HIV-1 infection in epidemiologically linked individuals.

Output in the last 12 months:
One manuscript was published in Frontiers in Immunology in 2019

Future Direction:
This is an ongoing collaboration on a cohort of HIV+ individuals in Cameroon and further investigations are currently being conducted on the breadth and potency of neutralizing antibody responses in elite controllers

2 - Evidence For Both Intermittent And Persistent Compartmentalization Of HIV-1 In The Female Genital Tract.

Collaborating Parties:
Batsirai M. Mabvakure, Bronwen E. Lambson,
Dale Kitchin, Cathrine Scheepers, Penny L. Moore, Lynn Morris - Centre for HIV and STIs, National Institute for Communicable Diseases of the National Health Laboratory Service, Johannesburg, South Africa; Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa
Lindi Masson, Carolyn Williamson, Jo-Ann Passmore, Darren P. Martin - Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Cape Town, South Africa. Kavisha Ramdayal, Gordon W. Harkins – MRC Capacity Devopment Unit, South African National Bioinformatics Institute.
Mushal Allam - Sequencing Core Facility, National Institute for Communicable Diseases of the National Health Laboratory Service, Johannesburg, South Africa. Salim Abdool Karim, Centre for the AIDS Programmed of Research in South Africa (CAPRISA), KwaZulu-Natal, South Africa

Nature and purpose:
To investigate whether distinct viral populations exist in the female genital tract and the blood that may impact the efficacy of microbicides and vaccines designed to provide mucosal immunity.

Output in the last 12 months:
One manuscript published in the Journal of Virology in 2019

Future Direction:
There are no intentions to investigate tissue compartmentalization in HIV+ individuals in the near future however, several other investigations of this virus are planned for 2020 with some of the collaborators on this project.

3 - Virulence Evolution

Collaborating Parties:
Philippe Roumagnac - CIRAD, Montpelier, France.
Darren P. Martin, Adérito Luis Monjane, Dionne Natalie Shepherd - UCT, South Africa.
Philippe. Lemey and Simon Dellicour - Katholieke Universiteit, Leuven, Belgium.
Pierre Lefeuvre and Jean-Michel Lett - CIRAD, Reunion Island.
Arvind Varsani – Arizona State University, USA.

Nature and purpose:
We 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 investigated the spatiotemporal dynamics and evolution of a broad range of viral pathogens focusing primarily on viruses that cause important crop diseases such as Tomato yellow leaf curl disease (TYLCD), Maize streak disease (MSD) and Cassava mosaic disease (CMD).

Output in the last 12 months:
One manuscript accepted by the scientific journal E-Life in December 2019

Future Direction:

This is a long-term ongoing collaboration that is currently attempting to develop rigorous statistical methods to analyze the effect of environmental factors on the dispersal history of viral epidemics and the evolution of virulence.

4 - Molecular Mechanisms Of Viral Emergence

Collaborating Parties:

Pauline Bernardo, Sarah François, Maëlle Deshoux, Denis Filloux, Emmanuel Fernandez, Serge Galzi, Romain Ferdinand, Martine Granier, Cécile Richet, Michel Peterschmitt and Philippe Roumagnac - CIRAD-INRA-SupAgro, Montpellier, France.
Brejnev Muhire, Darren P. Martin - UCT, South Africa.
Simona Kraberger, Arvind Varsani – Arizona State University, USA.
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 Archives of virology in 2019.

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 Euphorbia 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. Three other geminiviruses viruses that we discovered infecting the uncultivated plant species Limeum africanum, Exomis microphylla and Polygala garcinii DC that group in the with begomoviruses, becurtoviruses and capulaviruses respectively are currently being investigated by a MSc student at SANBI.

5 - Endogenous Viral Elements (EVEs)

Collaborating Parties:

Philippe Roumagnac, Denis Filloux and Emmanuel Fernandez - CIRAD, Montpelier, France.
Darren Martin - UCT.
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 (EVEs) 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:

A manuscript is in preparation and will be submitted for publication in 2020.

Future Direction:

We received funding to support this research from the French National Research Foundation (ANR) in December 2017 and subsequently begun generating data. Two annual meetings of all of the major stakeholders were held at CIRAD Réunion Island in September 2017 and Montpellier in 2018 to assess the progress to date and final meeting is scheduled for 2020.

6 - Phytovirus Diversity

Collaborating Parties:

Philippe Roumagnac, Denis Filloux, Charlotte Julian and Emmanuel Fernandez - CIRAD, Montpelier, France. Pierre Lefeuvre, Frederic Chiroleux and Jean-Michel Lett - CIRAD, UMR PVBMT, Réunion Island, France. Thierry Candresse, Armelle Marais, Marie Lefebvre and Chantal Faure - INRA, UMR BFP, Villenave d'Ornon France. Darren Martin - UCT South Africa. Arvind Varsani - Arizona State University, USA. Pascal Gentit and Benoir Remenant - ANSES France. Francois

Mesleard and Hugo Fontes - La Tour du Valat, France. Christophe Levergne - CBN CPIE Mascarin. Anne Duputie and Francois Massol UMR Evo-Eco-Paleo, France.

Nature and purpose:
While it is well established that emerging viruses generally originate in uncultivated hosts with which they have well-established interactions, we currently know almost nothing about the diversity and the distribution of phytoviruses circulating in either natural or managed ecosystems. This project proposes (i) to develop an integrated approach to reveal the phytoviral species richness of plant communities within several unmanaged and managed ecosystems, (ii) to experimentally study the effect of controlled plant communities on phytovirus species richness, and (iii) to search phytovirus nucleotide sequence data in silico evidence of the evolutionary footprints adaptation to a new host species.

Output in the last 12 months:
The first year of funding for this project was 2019 and to date we have completed the plant sampling efforts in South Africa in and Arizona for 2019. Further sampling trips are planned for both countries in 2020 while efforts to detect phytoviruses in our plant samples are currently being performed in the laboratory of Philippe Roumagnac in Montpelier France.

Future Direction:
This is an ongoing project that is funded until 2023.

7 - Rift Valley Fever Virus Epidemiology

Collaborating Parties:
Batsirai Mabvukere - Johns Hopkins Medical School, Baltimore NY, USA. Antoinette Grobbelaar - National Institute of Communicable Disease (NICD) Tshwane, South Africa

Nature and Purpose:
To analyze 224 Rift Valley fever virus complete genomes to investigate the epidemiological history of these viruses circulating in South Africa and other African countries. The results of this study will help inform governmental policy aimed at restricting the movements of infected livestock across national borders.

Output in the last 12 months:
None

Future Direction:
None

Uljana Hesse

Research Laboratory of Uljana Hesse

South Africa is home to an exceptional biological resource - the unique flora of the Cape Floristic Region includes over 3000 plant species actively used in traditional medicine. These plants produce a wide range of medicinally active compounds, many of which are rare or even species specific. One of the more famous endemic South African medicinal plant species is rooibos, best known as a herbal tea. It only grows in the Cederberg Mountain Region and represents a key cash crop for the local farming communities. Rooibos is increasingly recognized as a potential phytopharmaceutical: the species produces a wide range of phenolic compounds, which are associated with diverse medicinal properties of the plant (including anti-diabetic, cardioprotective, antispasmodic, anti-aging effects).

It therefore represents a suitable pilot plant species to initiate medicinal plant genomics research that focuses on the endemic flora of South Africa.

The Rooibos Genomics Programme aims to 1) locally establish all methodologies essential for medicinal plant genome analysis, 2) generate a high-quality assembly of the rooibos genome with extensive functional annotation of the rooibos genes, 3) identify rooibos genes involved in medicinal compound production and plant stress tolerance.

COLLABORATION WITH SANBI

The rooibos genome and eight rooibos transcriptomes have been sequenced using Illumina sequencing technologies. Establishment of biocomputational data analyses procedures is conducted at SANBI. In 2019, my research team has completed:

1. Computational analysis of the rooibos transcriptome data (manuscript in preparation);
2. Rooibos genome size estimation using k-mer analyses (published);
3. A first assessment of biocomputational tools for plant genome assembly using the computing clusters at SANBI and at CHPC (Centre for High Performance Computing) Cape Town, SA.

Students Graduated In 2019:

Yamkela Mgwatyu - MSc Bioinformatics, Thesis Title: Investigations into genome size and genetic diversity of distinct rooibos growth forms.
Emily Stander - PhD Bioinformatics, Thesis Title: Unraveling the transcriptome of Aspalathus linearis (Rooibos) towards identification of novel genes involved in polyphenol biosynthesis.

Publication:
Emily Amor Stander, Wesley Williams, Fanie Rautenbach, Marilize Le Roes-Hill, Yamkela Mgwatyu, Jeanine Marnewick, Uljana Hesse. "Visualization of Aspalathin in Rooibos (Aspalathus linearis) Plant and Herbal Tea Extracts Using Thin-Layer Chromatography" Molecules 24, no. 5 (2019): 938.

Research Laboratory of Simon Travers


Simon Travers

HIGHLIGHTS OF THE LAB

In 2019 we saw the **graduation of two PhD students**, Roux-Cil Ferreira and Phillip Labuschagne. Roux-Cil worked on understanding how the HIV virus is protected by sugars that binds to its surface and how we might be able to target these conserved elements with new treatments. Phillip's PhD focused on developing methods to characterize viral diversity and the timing of infection – these methods were developed in order to study data from HIV vaccine trials to understand the features of the virus that resulted in individuals becoming infected with HIV despite having received a vaccine.

The HIV drug resistance testing software developed in the research group is now being **rolled out worldwide** as a partnership between Hyrax Biosciences (a company spun-out from Simon's research group) and Thermo Fisher Scientific. Any users who purchase Thermo Fisher's **low cost HIV drug resistance (HIVDR) testing kit** are provided with free access to the software for the analysis and interpretation of the data generated by the kit. Access to a low-cost, scalable, end-to-end HIVDR solution is an exciting development in the **global fight against HIV and we are very proud that part of this solution was developed at SANBI.** In 2019, many national labs from Africa, Asia, Central and North America sign up to use the service. Further, 2019 saw the **expansion of the software to enable TB drug susceptibility testing** – this will also be rolled out globally through Hyrax Biosciences.

Finally, an agreement between UWC and Hyrax Biosciences was finalised for the **assignment of the intellectual property developed in the lab of Simon Travers to Hyrax Biosciences**. This agreement means that Hyrax Biosciences will now undertake the commercialisation and continued expansion of the Exatype drug resistance testing solution, ensuring that research undertaken in the laboratory will have maximal impact in global healthcare.

RESEARCH PROJECTS

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, recent work has focused on the expansion of Exatype to support TB drug susceptibility testing as well as the expansion to support sequencing data generated by the Sanger sequencing approach.

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.
Adding the capability for Sanger sequencing involved included the WHO-recommended analysis pipeline within the Exatype platform meaning that end-users can view and edit chromatograms when required.

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

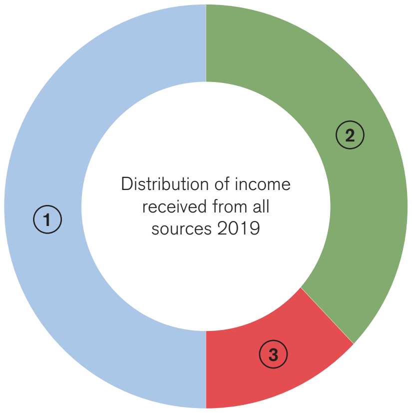
Our work on developing a data analysis pipeline for undertaking sieve analysis to explore the mechanism of viral evasion of vaccine-induced immune pressure focused on the characterisation of recombination in primer ID data. Such data enables the comparison of HIV-1 sequences from vaccine-experienced 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.

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.

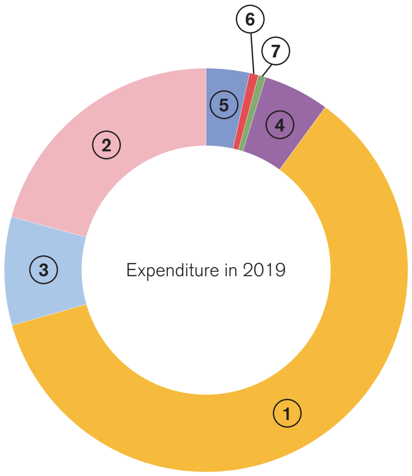
2019 FINANCIALS

SANBI's income and expenditure trends for 2019 are shown in this section.



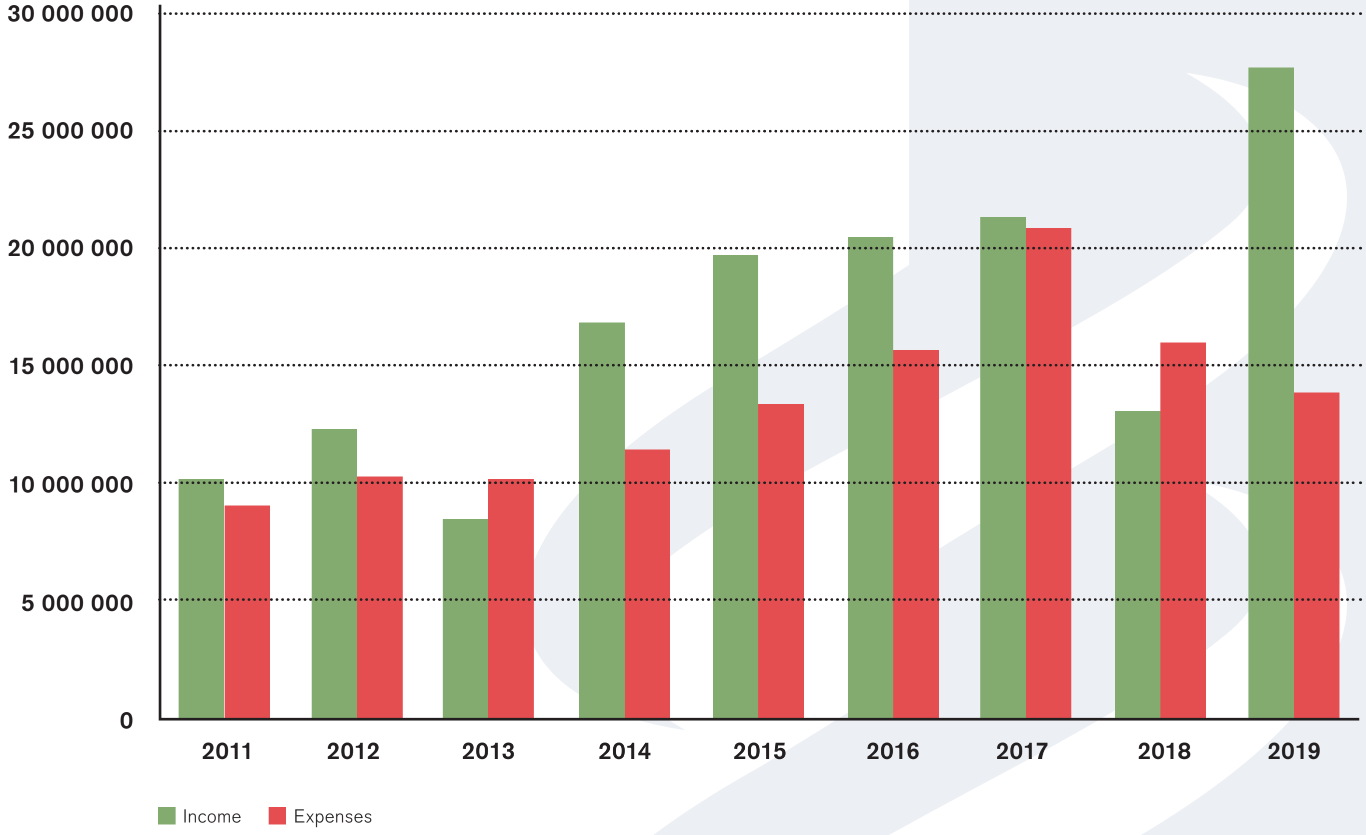
1. Foreign 50%
2. SA Agency 38%
3. UWC 12%

1. NRF 44%
2. UWC 30%
3. SAMRC 21%
4. TIA 5%

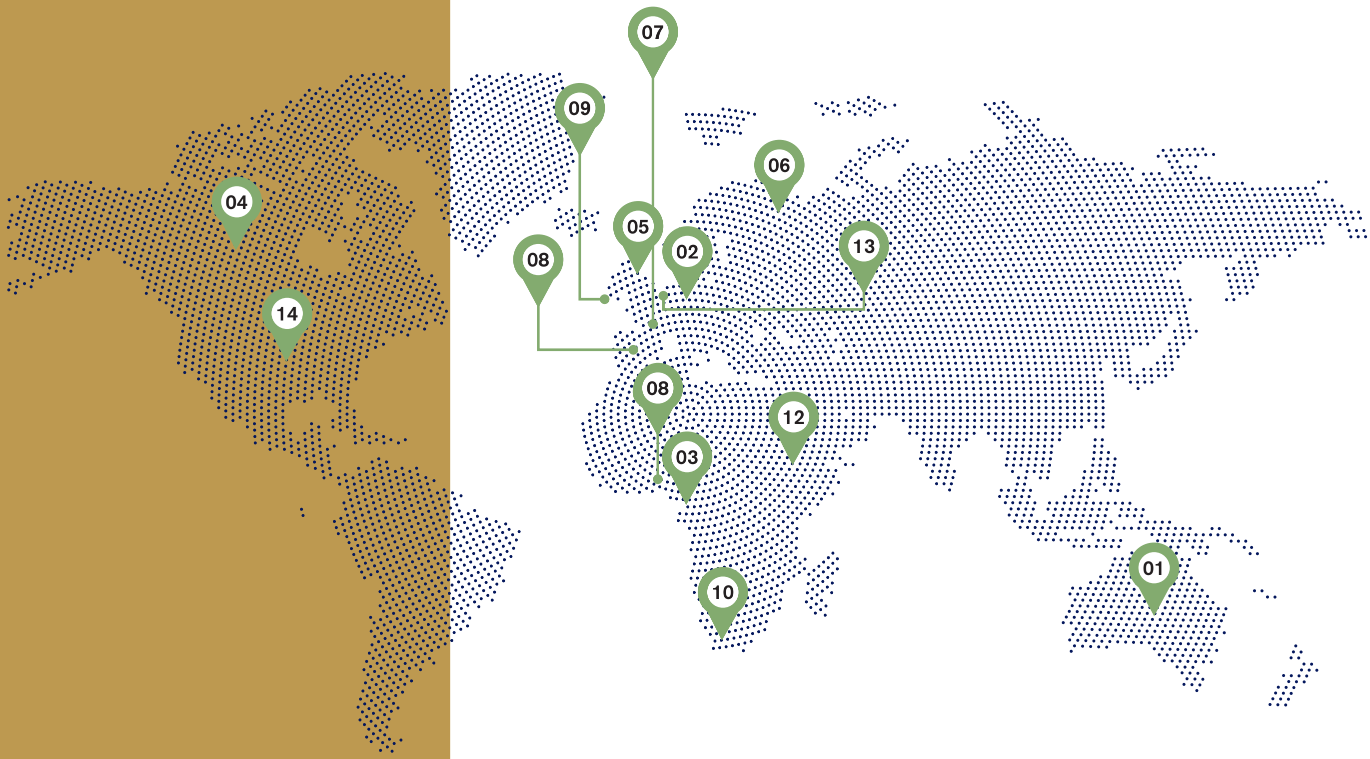


1. Salaries 60%
2. Students 21%
3. Running Costs 9%
4. Travel 5%
5. Equipment 3%
6. Internet 1%
7. Printing 1%
8. Telecomms 0,09%

Income vs. Expenditure 2011 – 2019



COLLABORATIONS



01 Australia

University of Melbourne
University of Queensland

02 Switzerland

Basel University, Basel
World Health Organisation

03 Cameroon

University of Yaoundé, Yaoundé

04 Canada

University of British Columbia
BC Center for Diseases Control
Centre de Recherche du CHUM
Université de Montréal, Montréal, QC
Simon Fraser University
British Columbia Centre for Excellence in HIV/AIDS
St. Paul's Hospital, Vancouver
McGill University, Montréal

05 United Kingdom

School of Law, Middlesex University
Oxford University
University of Birmingham

06 Finland

University of Helsinki, Finland

07 France

CIRAD, Montpelier, France.
CIRAD, UMR PVBMT, Réunion Island
CIRAD-INRA-SupAgro, Montpellier
INRA, UMR Villenave d'Ornon Cedex
INRA, URGI, Versailles
INRA, UMR BFP, Villenave d'Ornon
CIRAD, UMR AGAP, Guadeloupe
ANSES France
La Tour du Valat
CBN CPIE Mascarin
UMR Evo-Eco-Paleo

08 Nigeria

Global Consortium on Emerging Infectious Diseases

09 Republic of Ireland

Maynooth University

10 South Africa

University of the Western Cape
University of Cape Town
University of Stellenbosch
Cape Peninsula University of Technology
University of the Witwatersrand
University of Pretoria
University of KwaZulu-Natal
NHLS/ Stellenbosch University Tygerberg
National Institute for Communicable Diseases (NICD)
Centre for the AIDS Programme of Research in South Africa (CAPRISA)
H3Africa

11 Spain

Unidad de Sanidad Vegetal Zaragoza

12 Ethiopia

Africa CDC

13 Belgium

Katholieke Universiteit

14 United States

USA CDC
Broad Institute
Pacific North West Research Institute, Seattle
University of Washington, Seattle
Cleveland Clinic Lerner College of Medicine
New York University School of Medicine
Johns Hopkins University
Stanford University
Arizona State University
Icahn School of Medicine at Mount Sinai, New York
National Institutes of Health-National Institute of Allergy and Infectious Diseases, Bethesda, Maryland
Rocky Mountain Laboratories, Hamilton, Montana
Veterans Affairs New York Harbor Healthcare Systems, New York.
Arthur Nádas New York University School of Medicine, New York
Scripps Research Institute, La Jolla, California
Thermo Fisher Scientific
University of Georgia

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
Heikki Lehvaslaiho	Senior Research Scientist, Computational Bioscience Research Centre, King Abdullah University of Science and Technology
Tulio de Oliviera	Director: KRISP - KwaZulu-Natal Research and Innovation Sequencing Platform, UKZN, Durban, South Africa
Nicky Mulder	Head Computational Biology Group, Institute of Infectious Disease and Molecular Medicine, UCT
Cathal Seoighe	Stokes Professor of Bioinformatics, School of Mathematics, Statistics and Applied Mathematics, National University of Ireland, Galway
Dale Gibbs	IT Consultant
Samantha Alexander	UCT Faculty of Commerce, Administrative Assistant
Mario Jonas	Data Manager, UCT Computational Biology Group
Nicki Tiffin	Researcher, UCT School of Public Health and Family Medicine
Natasha Schoeman	UWC Student Administration
Long Yi	DevOps Manager, Konga.com
Maryam Salie	UWC Community Health Services, Administrative Assistant
Margaret Khumalo	Unknown
Thoba Lose	Development Operations at Entersekt

POSTDOCTORAL FELLOWS

Name	Completed	Currently
Soraya Bardien-Kruger	2002	Associate Professor, US
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	Senior Lecturer, Biotechnology, UWC
Sumir Panji	2012	H3ABioNet Project Manager, UCT
Samson Muyanga	2012	Researcher, Monash South Africa

John Pool	2012	University KwaZulu-Natal
Uljana Hesse	2013	Senior Researcher, UWC Biotechnology
Barbara Picone	2013	Associate Editor, Journal of Zoology
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
Rajan Sharma	2018	Postdoctoral Fellow, CPUT

PhD

Name	Completed	Currently
Alan Christoffels	2001	Director and Professor SANBI; DST/NRF Research Chair
Ekow Oppon	2002	Zambia
Junaid Gamieldien	2002	Associate Professor, 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	USA
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	Pretoria
Darlington Mapiye	2016	IBM
Rosaline Macharia	2016	Lecturer, University of Nairobi
Colleen Saunders	2016	Lecturer, UCT
Emil Tanov	2018	UCT

MSc

Name	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	South African MRC Research Scientist
Saleem Adam	2011	Unknown
Firdous Khan	2012	PhD UWC Biotechnology Department
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
Warren Jacobus	2015	Takealot
Batsirai Mabvakure	2015	Post-doctoral Fellow, Johns Hopkins University, Dept Pathology
Stephanie Pitts Muller	2016	IBM Research, Pretoria
Eugene Madzokere	2016	Zimbabwe
Toluwaleka Ademuyiwa	2018	UWC IT Services
Gratia Willemse	2018	Molecular Scientist, GENEdiagnostics
Olabode Ajayi	2019	Bioinformatics Business Intelligence Developer, CPGR

HONOURS

Name	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
Chantal De Long	2018

OBITUARY: ON THE PASSING OF PROFESSOR VLADIMIR BAJIC (1952-2019)

It is with sadness that I heard of the passing of a friend, and colleague, Professor Vladimir Bajic. Vlad (as he was known to many) had a prolific scientific career. Born in Slovenia, Vlad completed his studies at the University of Belgrade in Serbia and the University of Zagreb in Croatia. His career spanned academic or research posts in Serbia, South Africa, Singapore, and Saudi Arabia.

Vlad was an elected member of the Academy of Nonlinear Sciences in Russia for work done in stability theory of singular differential systems. His research in artificial intelligence and modeling resulted in the development of internationally acclaimed promoter prediction tools and a knowledge extraction platform called Dragon.

Staff and students at the South African National Bioinformatics Institute (SANBI) and the broader UWC community were fortunate to have interacted with him. Vlad joined UWC in 2006 as a Professor of Bioinformatics at SANBI. His training in engineering and mathematics shaped his insightful contribution to computational biology. Over the past decade, he has developed analytical tools to assist in the interpretation and understand of biological data. Among his prolific number of publications – more than 400 – is a suite of software tools that includes promoter prediction tools to understand gene regulation (**http://vladbajic.org/**). Vlad meticulously went about building modules to identify the DNA signals that trigger genetic messages through a process called "transcription". For those of us that knew his work, it was clear that he symbolized interdisciplinary research-using engineering principals to improve our understanding of mechanistic biology.
In 2007 Vlad was among the first DST/NRF Research Chairs (SARChI) awarded in South Africa. He was appointed the SARChI in Bioinformatics in Human Health and was based at SANBI. Here is cemented his methods for predicting the start of a genetic message i.e., promoter prediction, and various biological resources to analyze data related to ovarian cancer and hepatitis C.

I worked with Vlad between 2006-2009 before he took up a post as the director of Computational Bioscience Research Center at the King Abdullah University of Science and Technology (KAUST) in Jeddah. For the three years that I had worked alongside Vlad at SANBI, I observed an academic who was concerned for the wellbeing of students and staff. He would generously assist students with funding to complete their studies. He would give of his time to any student who was struggling to complete their thesis even if he was not formally associated with their graduate studies. Many postdoctoral students and graduate students would recall the time spent in his home reviewing data for a publication. No-one would leave his home until a draft manuscript had been written.

On a personal note, I was fortunate to have interacted with Vlad between 2001-2005 when both of us lived in Singapore. As fellow South Africans living abroad, he felt a responsibility to look out for me. Over time I saw him as a mentor. I fondly recall Saturday mornings trying to sleep late and Vlad would phone me and ask why postdocs are not in the lab doing research. He would report that he has been in the office since 8.30 am and he expected to see me also working. I would reluctantly get out of bed and go to campus on a Saturday morning. But within one hour of interacting with him on a scientific topic, I found myself engulfed in my research and he would remind me in the evening that its time to have dinner.

Vlad will be missed by the many students and colleagues whom he has interacted with across the globe. And certainly, by me – a former postdoc whose scientific philosophy has in part been shaped by the many hours of interaction around a computer or a pasta and bruschetta (his favourite dish after a hard day's work).

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

Funders

Research Institute at UWC
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

H3ABioNet Node 2012 – 2017

B3 Africa funding since 2015

Bill & Melinda Gates Foundation
funding since 2014

Technology Innovation Agency
funding since 2016

Poliomyelitis Research Foundation
Funding since 2019

Agriprotein funding since 2019

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

TELEPHONE: +27 (0)21 959 3645
FACSIMILE: +27 (0)21 959 2512
EMAIL: info@sanbi.ac.za
WWW.SANBI.AC.ZA

SANBI
South African National Bioinformatics Institute

RESPICE PROSPICE
UNIVERSITY of the WESTERN CAPE