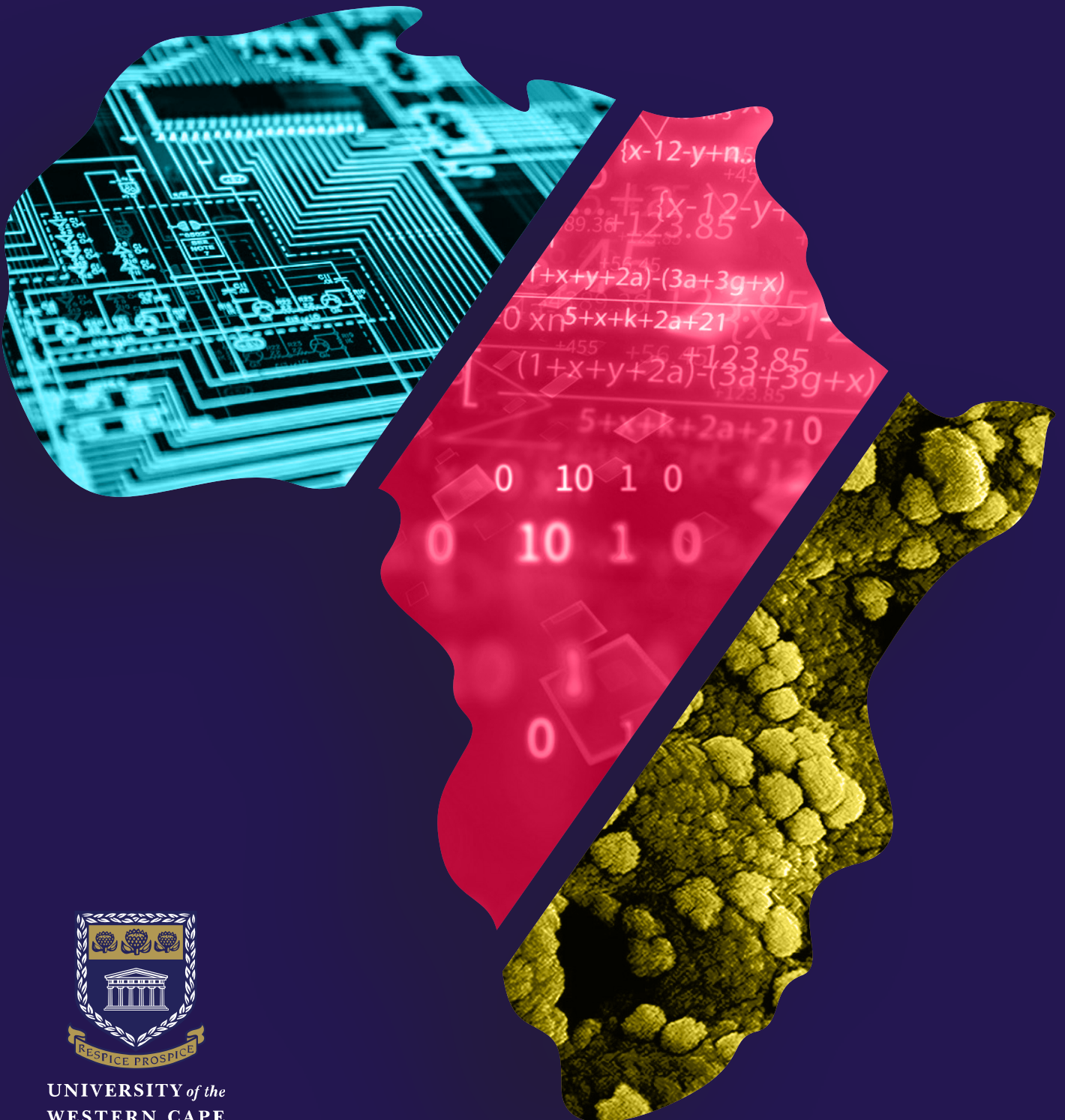




SANBI
South African National
Bioinformatics Institute

Annual Report 2021



UNIVERSITY of the
WESTERN CAPE

CONTENTS

ABOUT SANBI.....	2
DIRECTOR'S MESSAGE.....	4
YEAR IN REVIEW.....	5
STAFF.....	8
CAPACITY DEVELOPMENT.....	14
SANBI STUDENTS 2021.....	16
2021 SANBI GRADUATIONS.....	18
WORKSHOPS AND COURSES.....	19
COMPUTATIONAL RESOURCES.....	20
SANBI'S STAKEHOLDER ENGAGEMENT.....	23
RESEARCH OUTPUTS.....	25
RESEARCH LABORATORIES.....	31
COLLABORATIONS.....	48
FINANCIALS.....	50
ALUMNI.....	52
FUNDERS.....	53



ABOUT SANBI

WHO WE ARE

Bioinformatics is a specialist discipline straddling the fields of biology, mathematics and computer sciences, and it 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 that 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 bioinformatics tools. 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 specialised online resources for genomics and genome informatics;
- capacity development in genomics and bioinformatics in Africa; and
- the development and implementation of genome annotation methods.

OUR VISION

- To be a global leader in computational biology, achieving the highest level in biomedical research and education in the global, African and South African contexts.

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 promote 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...”.

National Research Foundation (NRF) Strategy 2025

The vision and mission of SANBI align with the NRF's Strategy 2025, specifically in “promoting global and competitive research and innovation”.

The research and development at SANBI align with the Department of Science and Innovation (DSI)'s White Paper on Science, Technology and Innovation 2019 by contributing to “the development of human capabilities, knowledge expansion and innovation performance...”.

The SA Medical Research Council (SAMRC) Act (Act 58 of 1991)

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

DIRECTOR'S MESSAGE

The COVID-19 pandemic continues to shape our public health response at SANBI. However, our research mandate extends beyond COVID-19, and it is exciting to see that staff and students have maintained their focus and contributed to knowledge generation in other areas besides infectious diseases.

In 2021, the tapering off of funding from SAMRC and the DSI/NRF Research Chairs Programme brought pivotal change to SANBI's support systems. In response to this shift in funding, in 2022 we have been exploring a range of external funding opportunities to mitigate any disruption to our research, training and development agenda. One of the initiatives highlighted in our report is the newly established World Health Organization / Regional Office for Africa (WHO-AFRO) centre of excellence for bioinformatics and genomic epidemiology based at SANBI.

Capacity development continues to be a priority at SANBI. In total, 30 postgraduate students (13 PhD and 17 MSc) and four postdoctoral fellows are supervised by the SANBI academic staff, comprising a cohort of culturally diverse and multi-talented researchers. Fifty-three percent of our postgraduate students are female and 62% are South African.

I trust the 2021 report will provide a glimpse of the exciting public health bioinformatics initiatives underway at SANBI, and the cutting-edge work being done by our postgraduate students.

Professor Alan Christoffels

PhD, M.ASSAf
Director & DSI/NRF Research Chair in Bioinformatics
and Health Genomics
Director of SA MRC Bioinformatics Unit

South African National Bioinformatics Institute
University of the Western Cape

“53% of the
postgraduate
students are **female**
and **62%** are
South African”



YEAR IN REVIEW

Over the course of 2021, SANBI staff and students have been an integral part of the global effort to find innovative solutions to the challenges of the COVID-19 pandemic. We remain committed to using sound science to counter disinformation about the pandemic and disseminate scientific information to local communities. As seen in the highlights below, we have cemented our position as a leader in translating bioinformatics research into public health advances.

In August 2021, Prof Alan Christoffels was appointed as a **senior advisor to the Africa Centres for Disease Control and Prevention (CDC)** on pathogen genomics and partnerships. During the past year, he has worked to spread awareness about the need for data governance on the African continent.



Africa CDC Mission to Botswana. From L-R: Prof Alan Christoffels with Dr Josephine Namboze, WHO Representative to Botswana; Dr Lul P Riek, Regional Coordinator for the Southern Africa Regional Collaborating Centre, Africa CDC; and Dr Francis Chikuse, Senior Technical Office-Public Health Policy, Africa CDC.

The **SANBI-SAMRC Bioinformatics Unit** had its five-year scientific review in August 2021 with a view to concluding a successful 20-year partnership with the SAMRC. The Bioinformatics Unit will wrap up its duties during 2022, but will continue to drive bioinformatics capacity development in South Africa as part of SANBI's constitutional mandate.

In September 2021 WHO-AFRICA partnered with SANBI to establish the **WHO-AFRO Regional Centre for Bioinformatics and Genomic Surveillance**. This centre specifically focuses on SADC countries, and its first event was a virtual training workshop on using bioinformatics techniques to analyse local SARS-CoV-2 genomic data. Peter van Heusden and Zipho Moshologu have been instrumental in supporting the centre's genomic surveillance work.

Our staff play a leadership role in the international programme, **Public Health Alliance for Genomic Epidemiology (PHA4GE)**, which brings bioinformatics solutions closer to public health. In her role as principal investigator and working group co-chair, Prof Nicki Tiffin has led the development of a framework that promotes ethical benefit sharing in health genomics research (<http://dx.doi.org/10.1136/bmjgh-2021-008096>).

Some of the **PHA4GE COVID-19 related work** is detailed below:

- Dr Dominique Anderson and Alan Christoffels were part of the PHA4GE working group which developed a data specification for maximising the use of metadata associated with SARS-CoV-2 biological specimens. The work defined the key metadata that should be captured when collecting SARS-CoV-2 biospecimens (<https://doi.org/10.1093/gigascience/giac003>).
- The COVID-19 pandemic has highlighted, yet again, the challenges involved in sharing genomic data. Anja Bedeker, Peter van Heusden and Nicki Tiffin have developed a framework to promote ethical benefit sharing in health research with implications for sharing data during a pandemic (<https://doi.org/doi:10.1136/bmjgh-2021-008096>).

SANBI contributed the following additional **COVID-19-related published research**:

- Dr Gordon Harkins was part of a team that characterised SARS-CoV-2 sequences from Uruguay and analysed the clinical impact of the COVID-19 outbreak in the country (<https://doi.org/10.1080/22221751.2020.1863747>).
- Dr Harkins also participated in the phylogenetic analysis that contributed to the understanding of the convergent evolution of the SARS-CoV-2 N501Y lineages (<https://doi.org/10.1016/j.cell.2021.09.003>).
- At UWC, we have been using a computational approach to identify predicted peptide-based drugs that can potentially inhibit the interaction of the SARS-CoV-2 spike protein with its target (<https://doi.org/10.1371/journal.pone.0245258>).

The impact of COVID-19 on the working environment

A collective effort was required for an effective response to the COVID-19 pandemic, which highlighted inequalities among students and staff in their access to the resources that facilitate remote working. SANBI supported its postgraduate students by providing data/internet bundles for remote work. As part of our teaching commitments at UWC, we have provided online lectures and prerecorded material for students to access; measures were also put in place to support online meetings.



Delegates attending the Regional Consultative Workshop on the Implementation of Public Health Pathogen Genomics in the Southern Africa Region on 25 October 2021, in Zambia.

2021 in Summary



R 31 383 413

Income



R 16 723 940

Expenditure

20 STAFF MEMBERS:



10 10
Male Female

18 South African
8 PhD
3 MSc
2 BSc
7 Dip

34 STUDENTS:



16 18
Male Female

21 South African
4 PostDoc
13 PhD
17 MSc

RESEARCH OUTPUTS:



23

Journals



1

Book Chapter



1

Software



6

Graduations



12

Keynotes/
Invited Talks



5

Workshops



4

Conference
Presentations



1

Course Module



17

Expert Panel/
Committee
Membership



44

Research
Projects



7

Awards/
Honours



49

Collaborations
in **13**
countries



51

Alumni
Countries

STAFF

As SANBI is a research institute, the Director reports to the Faculty of Natural Sciences at UWC. SANBI comprises a diverse group of research scientists and technical and administrative staff who all contribute to a dynamic, productive working environment.

SANBI academic staff are tasked with conducting research in their laboratories, securing funding, student graduations and capacity development.



Dominique Anderson, Dr

Senior Researcher

Funder: UWC



Hocine Bendou, Dr

Senior Researcher

Funder: UWC



Alan Christoffels, Prof

Director,
SARChI Bioinformatics,
SAMRC Unit Director

Funder: DSI/NRF Research
Chairs Programme



Ruben Cloete, Dr

Lecturer

Funder: UWC



Gordon Harkins, Dr

Senior Lecturer

Funder: UWC



Nicki Tiffin, Prof

Professor

Funder: UWC



Judit Kumuthini, Dr

Senior Researcher

Funder: H3A BioNetwork



Anja Bedeker

Research Associate

Funder: B&MGF

SANBI technical staff maintain and develop our computing infrastructure



Zipho Mashologu

Software Developer

Funder: SAMRC



Campbell Rae

Web Developer
(part time)

Funder: DSI/NRF Research
Chairs Programme



Peter van Heusden

Senior Systems Developer

Funder: DSI/NRF Research
Chairs Programme



Shadley Wentzel

Senior Systems
Administrator

Funder: SAMRC



**Rangarirai
Matima**

Communication Specialist
(PHA4GE)

Funder: B&MGF



Alecia Naidu

Technical Manager
(PHA4GE)

Funder: B&MGF



Jamie Southgate

Communications Officer
(PHA4GE)

Funder: B&MGF



Quinton Coert

Software Developer

Funder: Baobab LIMS

SANBI administrators ensure the smooth running of daily operations



Fungiwe Mpithi

PA/Administrator

Funder: SAMRC



Ferial Mullins

Finance Administrator

Funder: UWC



**Nawaal Nacerodien-
Weitz**

Administrator (PHA4GE)

Funder: B&MGF



Junita Williams

Marketing Administrator
(part time)

Funder: DSI/NRF Research
Chairs Programme

Awards and Honours

Alan Christoffels

Appointment as Senior Advisor to the Africa CDC in August.

Awarded by: Africa CDC

Dominique Anderson

Baobab LIMS, developed by SANBI, was nominated as a finalist for the NSTF innovation prize in October. The opensource project was recognised for its role in advancing data stewardship in laboratories to promote reproducible research.

Awarded by: National Science and Tech Forum (NSTF)

Nicki Tiffin

The Calestous Juma Science Leadership Fellowship recognises outstanding scientists tackling some of the greatest challenges in global health. The fellowship, to the value of \$980 000 over five years, supports career and leadership development for Prof Nicki Tiffin, as well as funding the development of an online platform to promote the equitable sharing of African data and biospecimens for future research.

Awarded by: Bill & Melinda Gates Foundation. (B&MGF), Calestous Juma Fellowship

Dominique Anderson

Baobab LIMS was pitched at the Virtual Start-up Sprint, a collaborative course between UWC and Mashauri held in July. The pitch competition included five entrepreneurs and was judged by a panel of experts. Baobab LIMS won the first prize of seed funding, and the team opted to pay it forward by sharing the prize money with the second runner-up.

Awarded by: UWC and Mashauri

Ruben Cloete

In October, the prestigious Fulbright fellowship was awarded to Ruben Cloete to attend the University of California San Francisco (UCSF) from August – December 2022 to establish new collaborations and networks between SANBI, UWC, and computational staff (Prof Andrej Sali, Prof Brian Shoichet and Matt Jacobson) at UCSF.

Awarded by: Fulbright Scholarship

Gordon Harkins

A grant to the value of R1.29 million was awarded for a collaborative project entitled "To predict and monitor T cell immune escape mutations at HLA-binding anchor motifs found in SARS-CoV-2 epitopes from genomes isolated in South Africa and sub-Saharan Africa".

Awarded by: SAMRC

SANBI academic staff are actively involved in translating research into policy. In 2021, half of the academic staff were appointed to national and international committees and expert panels that develop research frameworks to accelerate data science and biomedical research.

Expert Panel, Committee Membership and Conference Review Panels

Name	Institute	Role and Purpose
Dominique Anderson	ISCB Africa ASBCB 2021	Conference Chair; Conference Steering Committee; Scientific Program Committee Chair & Metaomics CoSI [Afrimetaomics]
	Academy of Science of South Africa	Standing Committee on POPIA – to guide the development of the POPIA Code of Conduct for Research
Alan Christoffels	African Society for Bioinformatics and Computational Biology	Governing Council member
	Academy of Science of South Africa	Member of Standing Committee on Biosafety and Biosecurity
	NHLS	Member of Biospecimen and Data Access Ethics Committee
	Sydney Brenner Institute for Molecular Bioscience, Wits	Research Advisory Committee
	Eastern Africa Network of Bioinformatics Training	Independent Scientific Advisory Board
	Global Emerging Pathogens Consortium	Member – hosting an annual conference on emerging pathogens as part of our mandate to create awareness of biosecurity threats across Africa, and to empower African scientists to respond to disease outbreaks
	ISCB Africa ASBCB 2021	Conference Steering Committee
	Applied Bioinformatics and Public Health Microbiology, Wellcome Genome Campus, UK Virtual Conference, 05–07 May	Scientific Programme Committee
Gordon Harkins	Centre for the AIDS Programme of Research in South Africa [CAPRISA]	Appointed Director of the Centre of Excellence in HIV Prevention at UWC

Name	Institute	Role and Purpose
Nicki Tiffin	University of Cape Town, Computational Biology Division	Honorary Professor
	International Common Disease Alliance	Member – Organising Committee; Co-Lead – Global Equity Working Group
	International 100k+ Cohorts Consortium	Member – Steering Committee; Co-Chair – Policy and Data Sharing Working Group
	PHA4GE Consortium	Member – Steering Committee; Chair – Ethics and Data Sharing Working Group
Tracy Calvert-Joshua	ISCB Africa ASBCB 2021	Scientific Program Committee Pathogen CoSI (Afripathogen)
Peter van Heusden	2021 Galaxy Community Conference [GCC2021]	Scientific Programme Chair
	ISCB Africa ASBCB 2021	Scientific Programme Chair

Journal Editing and Reviews

Hocine Bendou

Reviewer – Cancer Biomarkers

Alan Christoffels

Editor – Improper Scientist

Editorial Board Member – Data Journal

Reviewer – BMC Evolutionary Biology; BMC Genomics; BMC Bioinformatics; Bioinformatics; Journal of Parallel and Distributed Computing; PLoS ONE; Trends in Genetics; Genome Research

Nicki Tiffin

Journal Associate Editor – Nucleic Acids Research Genomics and Bioinformatics
Reviewer – Health Policy and Technology Journal

Further Development of Staff

Throughout the year, SANBI staff are encouraged to update their skills by attending relevant training interventions.

Hocine Bendou & Nasr Eshibona

Course Name and Benefit: Data Science with Python

Venue and Date: Get Smarter, UCT online [ongoing]

Peter van Heusden

Course Name and Benefit: Hosted by CABANA and GOBLET, with continued support from the ISCB, H3ABioNet, ELIXIR and APBioNET, a series of bioinformatics topics was presented.

Venue and Date: 2021 Bioinformatics Education Summit, 24–26 May, online

Theses Examined for Students at Other Institutions

Examiner: **Hocine Bendou**

Institution: University of Stellenbosch

Degree: MSc x 1

Examiner: **Alan Christoffels**

Institution: University of Stellenbosch

Degree: MSc x 2
PhD x 1

Examiner: **Nicki Tiffin**

Institution: University of Stellenbosch

Degree: PhD x 2

External Moderation

Alan Christoffels

Institution: University of Stellenbosch

Course: Bioinformatics module for third-year and Honours students

SANBI comprises a **diverse group** of research scientists and technical and administrative staff who all **contribute to a dynamic, productive working environment**

In 2021, half of the academic staff were **appointed to national and international committees and expert panels** that develop research frameworks **to accelerate data science and biomedical research.**

CAPACITY DEVELOPMENT

SANBI offers training programmes in keeping with its vision of becoming a centre of excellence in biomedical research and education in the global, African and South African contexts.

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)

Every year, the UWC undergraduate Bioinformatics module is taught to approximately 80 third-year Biotechnology students. During March and April 2021 this module was taught online. The SANBI faculty members Ruben Cloete, Gordon Harkins, Hocine Bendou and Alan Christoffels taught various Bioinformatics topics.

Internship Programme

As part of the DSI/NRF Research Chairs Programme, SANBI offers a one-year internship programme to students who have graduated with a BSc degree.

Honours Programme

Although SANBI does not have an Honours programme, students who obtain their undergraduate degree with a mark of >60% can include a bioinformatics component in their Honours project in the Biotechnology Department.

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.

Every year, the UWC undergraduate Bioinformatics module is taught to approximately **80** third-year students

Master's 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.

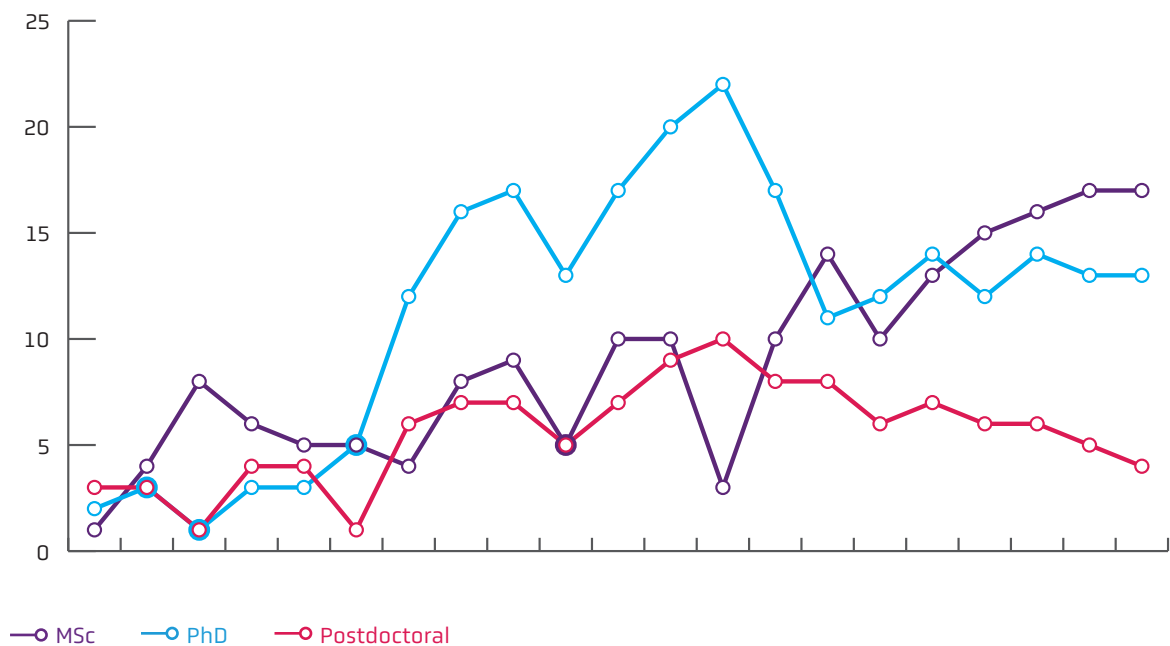
Doctoral Programme

Candidates should be in possession of a MSc degree in Bioinformatics or in a related scientific subject area, such as Computer Science, Mathematics, Biochemistry or Engineering. The PhD degree must be completed within five 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.

Figure 1: SANBI Postgraduate Registration Trends 2001–2021



SANBI STUDENTS 2021

SANBI students comprise a cohort of culturally diverse and multi-talented researchers. In total, 30 postgraduate students (13 PhD and 17 MSc) and four postdoctoral fellows are supervised by the SANBI academic staff. Fifty-three percent of the postgraduate students are female and 62% are South African.

A manuscript by SANBI PhD graduate Catherine Rossouw was selected as one of 11 papers in 2021 to have impacted the field of pathology and oncology (<https://doi.org/10.3389/pore.2021.622855>).

Postdoctoral Fellows



Ibrahim Ahmed

Supervisor:
Alan Christoffels



**Oladapo Michael
Olagbegi**

Supervisor:
Alan Christoffels



Wesley Williams

Supervisor:
Alan Christoffels



Catherine Rossouw

Supervisor:
Hocine Bendou

Doctoral Students



Anati Nkaule

Supervisor:
Alan Christoffels



Mmakamohelo Direko

Supervisor:
Alan Christoffels



Nikkita Pillay

Supervisor:
Alan Christoffels



Sarah DeRaedt

Supervisor:
Alan Christoffels



Souleymane Diallo

Supervisor:
Alan Christoffels



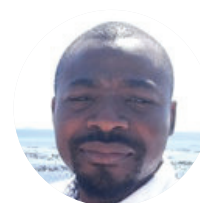
Tracey Calvert-Joshua

Supervisor:
Alan Christoffels



Abiola Abidemi Babjide

Supervisor:
Dominique Anderson



Abdulazeez Giwa

Supervisor:
Hocine Bendou



Nasr Eshibona

Supervisor:
Hocine Bendou



Bridget Langa

Supervisor:
Junaid Gamiieldien



Fanechka Esterhuysen

Supervisor:
Uljana Hesse



Yamkela Mgwatyu

Supervisor:
Uljana Hesse



Mohammed Hassan

Supervisor:
Ruben Cloete

Master's Students



Farzaana Diedericks

Supervisor:
Alan Christoffels



Peter Abiodun

Supervisor:
Alan Christoffels



Peter van Heusden

Supervisor:
Alan Christoffels



Ridaa Fredericks

Supervisor:
Alan Christoffels



Susan Alecia Fernol

Supervisor:
Alan Christoffels



Rudolph Serage

Supervisor:
Dominique Anderson



Siposetu Mazitshana

Supervisor:
Gordon Harkins



Nomlindelo Mfuphi

Supervisor:
Hocine Bendou /
Alan Christoffels



Wardah Jassiem

Supervisor:
Hocine Bendou



Darren Isaacs

Supervisor:
Ruben Cloete



Maryam Hassan

Supervisor:
Ruben Cloete



Rumbidzai Chitongo

Supervisor:
Ruben Cloete



Christianah Omotoso

Supervisor:
Samuel Egieyeh



Samson Oselusi

Supervisor:
Samuel Egieyeh



Lunathi Mhlali Ndlondlo

Supervisor:
Ruben Cloete



Keaghan Brown

Supervisor:
Ruben Cloete



Nkosazana Nyembezi

Supervisor:
Gordon Harkins

2021 SANBI GRADUATIONS

SANBI has great pleasure in presenting the 2021 graduands.

PhD

Souleymane Diallo

Supervisor: Alan Christoffels

Coding of tsetse repellents by olfactory sensory neurons: towards the improvement and the development of novel tsetse repellents.

Sophia Catherine Rossouw

Supervisor: Alan Christoffels

Optimisation of proteomics techniques for archival tumour blocks of a South African cohort of colorectal cancer.



Congratulations to Ruben Cloete's MSc Students Darren Isaacs and Maryam Hassan who both graduated in October 2021.

MSc

Oluwafemi Peter Abiodun

Supervisor: Alan Christoffels

Exploring the influence of organisational, environmental, and technological factors on information security policies and compliance at South African higher education institutions: implications for biomedical research.

Maryam Hassan

Supervisor: Ruben Cloete

Molecular modelling and simulation studies to prioritise sequence variants identified by whole-exome sequencing in a South African family with Parkinson's disease.

Darren Isaacs

Supervisor: Ruben Cloete

Molecular dynamic simulation studies of the South African HIV-1 Integrase subtype C protein to understand the structural impact of naturally occurring polymorphisms.

Jamie Southgate

Supervisor: Alan Christoffels

Establishing a framework for an African Genome Archive.

WORKSHOPS AND COURSES

Throughout the year, capacity development programmes are organised or hosted by the SANBI faculty to further develop human resources in the field of bioinformatics and computational biology. From time to time, SANBI faculty members are also invited to teach specific topics.

Figure 2: Online Workshops Organised/Taught By SANBI



SANBI Course Material Contributed

At the ASBCB 2019 conference in Ghana, Thoba Lose and Peter van Heusden participated in writing a tutorial on Mycobacterium tuberculosis variant analysis in Galaxy. The tutorial was then incorporated into the programme of Galaxy Smörgåsbord, a Galaxy training week organised by the Gallantries project in February 2021. As part of the preparation for that course, Peter van Heusden recorded a video walkthrough of the tutorial.

Early in 2022 the tutorial was incorporated into a larger course which is now [1] using it and Galaxy for other training components and [2] being shared freely online: (<https://gallantries.github.io/video-library/events/mtb-ngs/course.html>).

This is a great example of SANBI leading in an open-access space (with materials online under a licence that facilitates re-use) and working with Galaxy as a training and analysis platform for pathogen genomics.

The original tutorial was used again for Smörgåsbord 2 in March 2022, and it was completed by users from at least three continents (Africa, Europe and Asia).

COMPUTATIONAL RESOURCES

SANBI's IT and bioinformatics software development 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. The team has been focused on two projects in the past year: Baobab LIMS and COMBAT-TB.

BAOBAB LIMS

Baobab LIMS is an open-source laboratory information management system (LIMS) developed by SANBI (www.baobablims.org). To enable continuous development and improvement of Baobab LIMS, engagement with potential end-users is used to gain greater insight into operations in different laboratory and biobank environments. This provides the development team with an understanding of unique local LIMS functionality gaps, which enables the development of practical solutions to fit local needs. Several enhancements and customised modules have been built to strengthen the core capacity of Baobab LIMS, making the system applicable to sample quality control (QC) for a variety of biological collections. New modules have been introduced for sample preparation, include sample aliquoting, centrifugation and sample pooling.

Additional metadata-aware capability aimed at improving interoperability and harmonisation has been added. The most recent common structured vocabulary added to Baobab LIMS is based on the SARS-CoV-2 metadata standard, developed by a working group within the PHA4GE consortium (<http://www.pha4ge.org>). Using a combination of pre-coded options, as well as configurable content types, Baobab LIMS has broadened the scope of this module to ensure applicability to the collection of any viral sample.

Baobab LIMS analytics and quality control capability have also been strengthened by the addition of an audit logging functionality, Excel export functionality, as well as improvements in the audit reporting module. Importers, which include automatic control checks to ensure that data duplication does not occur, are a viable work-around for users with intermittent internet connectivity.

As the number of new functionalities in Baobab LIMS has increased, software verification to ensure new code changes have not affected existing functionality has been incorporated into the code base. This ensures that if any development errors occur, they are identified early on. To improve Baobab LIMS infrastructure deployment, repositories containing Terraform scripts and an Ansible playbook for the deployment of a production-ready installation have been made available, ensuring that users can successfully deploy Baobab LIMS on local servers, as well as on cloud infrastructure, such as Amazon Web Services (AWS).

Baobab LIMS undergoes continuous enhancement, and users can contact the team via a dedicated help desk (help@baobablims.org). User feedback is always welcomed by the Baobab LIMS team.

Baobab LIMS enables the development of **practical solutions to fit local needs**

COMBAT-TB NeoDB and WORKBENCH

The Computational Bacterial Analytical ToolKit for Tuberculosis (COMBAT-TB) (<https://combattb.org/>) is a SAMRC-funded project to develop bioinformatics tools and resources for understanding Mycobacterium tuberculosis (M. tb), the bacterium that causes tuberculosis (TB). Two components of this project have been developed at SANBI: the COMBAT-TB NeoDB and Workbench.

The NeoDB (<https://combattb.org/combat-tb-neodb/>) is an integrated M.tb knowledge base built on top of a database of genomic, protein and drug interaction data and publications. Users can either query the database directly or navigate it using the COMBAT-TB eXplorer web interface (<https://explorer.sanbi.ac.za/>). This work was published in the journal Bioinformatics, with Thoba Lose as lead author.

The second component is the COMBAT-TB Workbench, a web-based toolkit for routine M. tb bioinformatic analyses. This is built on top of the Integrated Rapid Infectious Disease Analysis (IRIDA) Platform, a project developed by the Public Health Agency of Canada's National Microbiology Laboratory (PHAC-NML). IRIDA has already been used to analyse thousands of bacterial samples. The COMBAT-TB Workbench makes the software easier to install and adds M. tb sample and phylogenetic analysis modules. The M. tb sample processing pipeline used in the Workbench is also the basis of a Galaxy tutorial that was included in the 2021 Galaxy Smörgåsbord online training event. A preprint on the COMBAT-TB Workbench was submitted in 2021, with publication planned for 2022.

GALAXY

SANBI remains an active part of the Galaxy Community. Peter van Heusden was the Scientific Programme Co-Chair of the 2021 Galaxy Community Conference, which took place entirely online. He also helped coordinate Galaxy Africa and Galaxy Public Health meetup events during that conference and has worked with Ziphozakhe Mashologu on numerous enhancements and bug fixes for the Galaxy software, Galaxy tools and Galaxy itself. Peter van Heusden was also an instructor during the Galaxy Smörgåsbord training week, held online in February 2021. The event saw more than 1000 attendees from around the world.

IT AND SYSTEMS SUPPORT

We welcomed Shadley Wentzel to fill the role of SANBI systems support engineer.

We continue to provide an HPC cluster and OpenStack computing cloud as well as a Ceph storage cluster for our users. Internet connectivity is provided by UWC, and SANBI is connected to SANReN, the South African Research and Education Network that provides 10Gb/s connections between South African universities and research centres.

SANBI HPC CLUSTER

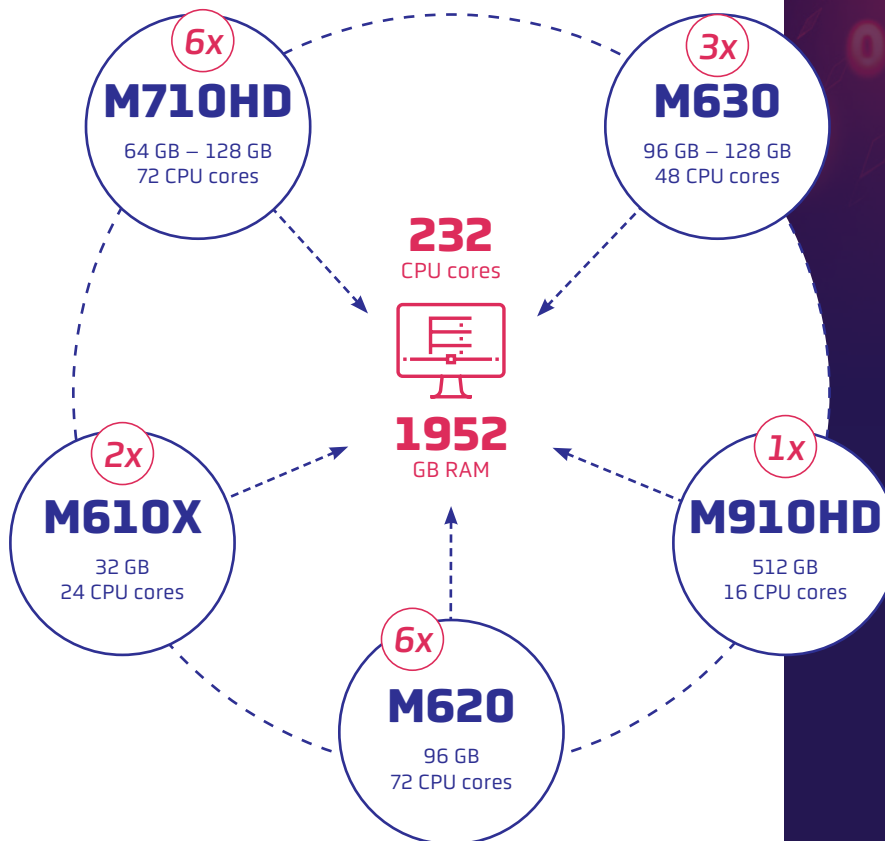
We have 232 CPU cores and 1952 GB of RAM on our Dell HPC cluster.

VIRTUAL MACHINE INFRASTRUCTURE

Our OpenStack cloud runs on Supermicro servers, which provide 64 CPUs and 340 GB of RAM.

CEPH STORAGE

SANBI storage runs on a Ceph cluster with 264 TB of RAW storage. This is provided by five Supermicro storage servers.



OUTREACH AND MENTORSHIP

SANBI supported UWC's "Team Parallelisers" in the 2021 Student Cluster Competition (SCC). The team performed exceptionally well and went on to win the competition. The four members of Team Parallelisers have gone on to form the core of South Africa's entry in the International Student Cluster Competition, which will be held alongside the ISC High Performance conference in Hamburg, Germany from 29 May 2022.

4 team members represented at Student Cluster Competition at **International Supercomputing Conference (ISC) 2021 Digital in Hamburg, Germany** with teams from China, Singapore, Taiwan, South Africa, Spain, and the United Kingdom.

SANBI'S STAKEHOLDER ENGAGEMENT

SANBI has used a range of platforms to communicate our research to relevant stakeholders and the public.

Public Science Engagement

Alan Christoffels was appointed as a senior advisor to the Africa CDC at the African Union, where he uses the platform to inform a range of public health stakeholders on the current genomics landscape and areas of impact for future public health applications.

Community Engagement

Ruben Cloete arranged a public seminar on COVID-19 vaccine myths in partnership with a local church in the Silversands area.

Nicki Tiffin was featured on a podcast discussing key skills for successful grant funding in health research, and the impact of funding opportunities on her career (https://twitter.com/mentor_podcast/status/1458370643255246848).

Nicki Tiffin was appointed as a board member for Adonis Musati Project (<https://www.adonismusatiproject.org/>), with the specific focus of providing data governance and ethics guidance to the programme, which works with refugees and asylum seekers in Cape Town.



SANBI students Darren Isaacs and Rumbidzai Chitongo assisted with administering COVID-19 vaccines at the Community Health Services in Bellville on 28 September 2021.

Media Engagement

13 August

Die Burger

UWK-prof nou raadgewer van Afrika-CDC

'n Navorsers van die Universiteit van Wes-Kaapland (UWK) is pas aangewys as senior raadgewer van die Afrika-sentrums vir die beheer en voorkoming van siektes (CDC).

Prof. Alan Christoffels, direkteur van die UWK se bio-informatika-instituut, Sanbi, wat ook die leerstoel vir bioinformatika en gesondheidsgenomika aan die UWK beklee, is aangestel as senior raadgewer vir patoogeen-genomika en vennootskappe by die Afrika-CDC, gesetel in die Afrika-Unie.

Sanbi werk reeds die afgelope twee jaar met die Afrika-CDC saam om stelsels vir die bestuur van siekte-uitbrekings te ontwikkel wat wyer toegepas sal kan word as die koronavirus-pandemie, sê die UWK in 'n verklaring. Christoffels sal leiding en tegniese bystand verskaf in die ontwikkeling en evaluering van strategiese projekte van die Afrika-CDC se instituut vir patoogeen-genomika.

"Hierdie werk fokus op die integrasie van patoogeen-genomika in die besluitnemingprosesse ten opsigte van openbare gesondheid op die vasteland," sê Christoffels.

"Daar is 'n dringende behoefte om eksperimentele laboratoriums wat DNS-volgordebepaling doen te steun sodat data-oplossings geskep word wat ons sal help om toekomstige uitbrekings te hanteer," sê hy.



Prof. Alan Christoffels

19 August

Research Professional News
(Christiaan van der Merwe and Linda Nordling)

South African to advise African disease centre on genomics data

[<https://www.research-professionalnews.com/rr-news-africa-south-2021-8-south-african-to-advise-african-disease-centre-on-genomics-data/>]

9 September

WHO Africa

Professor @alangchris takes the floor on the @WHOAFRO press conference announcing @SANBI_SA's role in the "Centre of Excellence for Genomic Surveillance" being launched today.

[<https://youtube.com/watch?v=WryyMtlJGcg>]

9 September

WHO Africa

COVID-19 variants prolong Africa's pandemic wave

[<https://www.afro.who.int/news/covid-19-variants-prolong-african-pandemic-wave>]

10 September

Times Live (Claire Keeton)

Africa gets major boost in efforts to track and trace COVID-19 changes

[<https://www.timeslive.co.za/news/south-africa/2021-09-10-africa-gets-major-boost-in-efforts-to-track-and-trace-covid-19-changes/>]

15 September

IOL (Anel Lewis)

#VaxxFacts: UWC partners with WHO on a new genomic surveillance centre to monitor disease outbreaks beyond COVID-19 in Africa.

[<https://www.iol.co.za/news/covid19/vaxxfacts-uwc-partners-with-who-on-a-new-genomic-surveillance-centre-to-monitor-disease-outbreaks-beyond-covid-19-in-africa-22900852-cc9e-409a-9fc3-2bc6e45ac0ed>]

30 September

WHO Regional Office for Africa
Scaling up genomic sequencing in Africa

[<https://www.afro.who.int/news/scaling-genomic-sequencing-africa>]

1 October

Africa-Newsroom (APO Group)
Intensify genomic sequencing in Africa

[<https://nnn.ng/intensify-genomic-sequencing-in-africa/>]

24 October

Political Economist Nigeria
Lagos to host seventh African Conference on Biosecurity, OneHealth

[<https://www.politicaeconomistng.com/lagos-host-seventh-african-conference-biosecurity-one-health/>]

9 November

Twitter

The Calestous Juma Science Leadership Fellowship recognises outstanding scientists tackling some of the greatest challenges in global health. Congratulations to the inaugural 2021 cohort!

[<https://twitter.com/gatesfoundation/status/1458117470879879171>]

RESEARCH OUTPUTS

SANBI's profile of excellence is reflected in its ability to publish high-impact scientific articles in international publications. Bioinformatics is an interdisciplinary field, as evidenced by our collaborative publication outputs. SANBI staff and students occupy leading roles in at least 40% of our outputs for the review period.

Journal Publications

Authors and Title	Publication Details	Date
Egieyeh S, Egieyeh E, Malan S, Christoffels A , Fielding B. Computational drug repurposing strategy predicted peptide-based drugs that can potentially inhibit the interaction of SARS-CoV-2 spike protein with its target (humanACE2).	PLoS ONE 16(1): e0245258. (https://doi.org/10.1371/journal.pone.0245258)	8 January
Elizondo V, Harkins GW , Mabvakure B , Smidt S, Zappile P, Marier C, et al. SARS-CoV-2 genomic characterization and clinical manifestation of the COVID-19 outbreak in Uruguay.	Emerging Microbes & Infections 10:1, 51-65 (https://doi.org/10.1080/2221751.2020.1863747)	15 January
Shamsi S, Anjum H, Shahbaaz M , Khan MS, Ataya FS, Alamri A, Alhumaydhi FA, Husain FM, Rehman MT, Mohammad T, Islam A, et al. A computational study on active constituents of Habb-ul-aas and Tabasheer as inhibitors of SARS-CoV-2 main protease.	Journal of Biomolecular Structure and Dynamics. (https://doi.org/10.1080/07391102.2021.1900920)	24 March
Sebate B, Cuttler K, Cloete R , Britz M, Christoffels A , Williams M, et al. Prioritization of candidate genes for a South African family with Parkinson's disease using in-silico tools.	PLoS ONE 16(3): e0249324. (https://doi.org/10.1371/journal.pone.0249324)	26 March
Yousuf M, Shamsi A, Queen A, Shahbaaz M , Khan P, Hussain A, et al. Targeting cyclin-dependent kinase 6 by vanillin inhibits proliferation of breast and lung cancer cells: combined computational and biochemical studies.	Journal Cellular Biochemistry 2021, 122:897-910. (https://doi.org/10.1002/jcb.29921)	8 April
Diallo S , Shahbaaz M , Makwatta JO, Muema, JM, Masiga D, Christoffels A , Getahun MN. Antennal Enriched Odorant Binding Proteins Are Required for Odor Communication in Glossina f. fuscipes.	Biomolecules 2021, 11, 541. (https://doi.org/10.3390/biom11040541)	8 April
Mikasi SG, Isaacs D , Chitongo R , Ikomey GM, Jacobs GB, Cloete R . Interaction analysis of statistically enriched mutations identified in Cameroon recombinant subtype CRF02_AG that can influence the development of Dolutegravir drug resistance mutations.	BMC Infectious Diseases (2021), 21:379 (https://doi.org/10.1186/s12879-021-06059-x)	23 April

Authors and Title	Publication Details	Date
Adams R, Adeleke F, Anderson D , Bawa A, Branson N, Christoffels A , et al. POPIA Code of Conduct for Research.	South African Journal of Science 2021;117(5/6), Art. #10933. [https://doi.org/10.17159/sajs.2021/10933]	03 May
Rossouw SC , Bendou H , Blignaut RJ, Bell L, Rigby J and Christoffels A . Evaluation of Protein Purification Techniques and Effects of Storage Duration on LC-MS/MS Analysis of Archived FFPE Human CRC Tissues.	Pathology & Oncology Research 27:622855. [https://doi.org/10.3389/pore.2021.622855]	03 May
Anand K, Vadivalagan C, Joseph JS, Singh SK, Gulati M, Shahbaaz M , Abdellattif MH, Prasher P, Gupta G, Chellappan DK, Dua KA. Novel nano therapeutic using convalescent plasma derived exosomal [CPExo] for COVID-19: A combined hyperactive immune modulation and diagnostics.	Chemico-Biological Interactions 344, 1, 109497 [https://doi.org/10.1016/j.cbi.2021.109497]	19 May
Dellicour S, Hong SL, Vrancken B, Chaillon A, Gill MS, Maurano MT, Ramaswami S, Zappile P, Marier C, Harkins GW , Baele G, Duerr R, Heguy A. Dispersal dynamics of SARS-CoV-2 lineages during the first epidemic wave in New York City.	PLoS Pathogens 17(5): e1009571. [https://doi.org/10.1371/journal.ppat.1009571]	20 May
Oselusi SO, Egieyeh SA, Christoffels A . Cheminformatic Profiling and Hit Prioritization of Natural Products with Activities against Methicillin-Resistant Staphylococcus aureus [MRSA].	Molecules 2021, 26, 3674. [https://doi.org/10.3390/molecules26123674]	16 June
Oselusi SO, Christoffels A , Egieyeh SA. Cheminformatic Characterization of Natural Antimicrobial Products for the Development of New Lead Compounds.	Molecules 2021, 26, 3970. [https://doi.org/10.3390/molecules26133970]	29 June
Ramesh M, Anand K, Shahbaaz M and Abdellattif MH. Current Perspectives in the Discovery of Newer Medications Against the Outbreak of COVID-19.	Frontiers in Molecular Biosciences 8:648232. [doi: 10.3389/fmolb.2021.648232]	12 July
Badmus KO, Wewers F, Al-Abri M, Shahbaaz M , Petrik L. Synthesis of Oxygen Deficient TiO₂ for Improved Photocatalytic Efficiency in Solar Radiation.	Catalysts 2021, 11, 904. [https://doi.org/10.3390/catal11080904]	26 July
Martin DR, Sibuyi NR, Dube P, Fadaka AO, Cloete R , Onani M, Madiehe AM, Meyer M. Aptamer-Based Diagnostic Systems for the Rapid Screening of TB at the Point-of-Care.	Diagnostics 2021, 11,1352. [https://doi.org/10.3390/diagnostics11081352]	28 July

Authors and Title	Publication Details	Date
Desmirean M, Richlitzki C, Pasca S, Teodorescu P, Petrushev B, Rauch S, Steinheber J, Iluta S, Liu J, Dima D, Grewal R , Ma W, Qian L, Tomuleasa C. Correlation between the prevalence of T-cell lymphomas and alcohol consumption.	Medicine and Pharmacy Reports Vol. 94, No. 3, 2021: 298 – 306. (https://doi.org/10.15386/mpr-1777)	28 July
Vatlin AA, Shitikov EA, Shahbaaz M , Bespiatykh DA, Klimina KM, Christoffels A , Danilenko VN and Maslov DA. Transcriptomic Profile of Mycobacterium smegmatis in Response to an Imidazo[1,2-b][1,2,4,5] Tetrazine Reveals its Possible Impact on Iron Metabolism.	Frontiers in Microbiology 12:724042. (doi: 10.3389/fmicb.2021.724042)	04 August
Martin DP, Weaver S, Tegally H, San JE, Shank SD, Wilkinson E, Lucaci AG, Giandhari J, Naidoo S, Pillay Y, Singh L, Lessells RJ, NGS-SA415COVID-19 Genomics UK [COG-UK], Gupta RK, Wertheim JO, Nekturenko A, Murrell B, Harkins GW , et al. The emergence and ongoing convergent evolution of the SARS-CoV-2 N501Y lineages.	Cell 2021, Vol 184, Issue 20, P5189-5200. (https://doi.org/10.1016/j.cell.2021.09.003)	7 September
Abdellattif MH, Shahbaaz M , Arief MMH, Hussien MA. Oxazinethione Derivatives as a Precursor to Pyrazolone and Pyrimidine Derivatives: Synthesis, Biological Activities, Molecular Modeling, ADME, and Molecular Dynamics Studies.	Molecules 2021, 26, 5482. (https://doi.org/10.3390/molecules26185482)	9 September
Cloete R, Shahbaaz M , Grobbelaar M, Sampson SL, Christoffels A . In silico repurposing of a Novobiocin derivative for activity against latency associated Mycobacterium tuberculosis drug target nicotinate-nucleotide adenyllyl transferase [Rv2421c].	PLoS ONE 16(11): e0259348. (https://doi.org/10.1371/journal.pone.0259348)	2 November
Giwa A, Rossouw SC, Fatai A, Gamielien J, Christoffels A, Bendou H . Predicting amplification of MYCN using CpG methylation biomarkers in neuroblastoma.	Future Oncology, 17(34) 4769-4783 (https://doi.org/10.2217/fon-2021-0522)	9 November
Rossouw S, Bendou H , Bell L, Rigby J, Christoffels A . Effect of polyethylene glycol 20 000 on protein extraction efficiency of formalin-fixed paraffin-embedded tissues in South Africa.	African Journal of Laboratory Medicine 10(1), a1122. (https://doi.org/10.4102/ajlm.V10i1.1122)	17 December

Book Chapter Publications

Title	Author/s	ISBN	Publishers	Date
Biobanking in Low and Middle Income Countries: Relevance, Setup and Management	Dominique Anderson , Hocine Bendou , Bettina Kipperer, Kurt Zatloukal, Heimo Müller and Alan Christoffels	ISBN-10: 3030876365 ISBN-13: 978-3030876364	Springer International Publishing	December

Software Resources/Computational Tools Developed

Name of Resource	Developed by	Description and Purpose/Usage
COMBAT-SARS-CoV-2 Africa CDC catalogue (Proof of principle)	Quinton Coert, Ziphozakhe Mashologu, Peter van Heusden, Dominique Anderson	Catalogue tool developed to pull information from Baobab LIMS databases for early surveillance through independent virus sample data capture from laboratories using the LIMS.

Keynotes or Invited Talks

Open Meeting 2021
The intersection between bioinformatics and public health

Evolution of Baobab LIMS: coding in metadata standards for infectious disease

Dr Dominique Anderson, SANBI

Date: 30 March 2021 • 4-7pm CAT Zoom Registration: bit.ly/3b2XSSG

Presenter	Conference Name	Date	Title
Dominique Anderson	Women in Data Science Africa (WiDS)	8 March	Invited panellist for the “Effects of COVID on Women in Data Science” session
Dominique Anderson	PHA4GE Open Meeting	30 March	Evolution of Baobab LIMS: Coding in metadata standards for infectious disease
Dominique Anderson	POPIA Public Consultation Forum	3 May	Progress on the POPIA Code of Conduct for Research
Peter van Heusden	Bioinformatics Hub of Kenya	13 May	Contributions of Bioinformatics to COVID-19 research in Africa. (https://www.youtube.com/watch?v=IWBIGac7JEI)
Alan Christoffels	GET Biosecurity Virtual Summit	24 June	Invited Talk: Understanding the use of big data and genomics in solving emerging biosecurity threats

BIOSECURITY VIRTUAL SUMMIT

SPEAKERS



PROF AKIN ABAYOMI
Honourable Commissioner for Health, Lagos State

Topic: Enhancing public private partnership in addressing the biosecurity threats



DR. AYODOTUN BOBADOYE
Chief Operating Officer, GET Consortium

Topic: Understanding emerging biosecurity threats in Nigeria



PROF. CHUKWUMERIJE OKEREKE
Director for Centre for Climate Change and Development in Alex Ekwueme Federal University Ebonyi State

Topic: Climate Crisis Our Crisis: Finding solutions to impacts of climate change in Nigeria



PROF. ALAN CHRISTOFFELS
Director of the South African National Bioinformatics Institute (SANBI)

Topic: Understanding use of big data and genomics in solving emerging biosecurity threats

Theme:
Trans-disciplinary approach to addressing biosecurity threats in Nigeria

Thursday, 24th June, 2021

10:00am WAT

Zoom

Registration Link:

<http://bit.ly/GETBiosecuritySummit>

[@getconsortium](https://www.getconsortium.org)
www.getafrica.org

Alan Christoffels	WHO, Virtual international meeting	15 July	Invited WHO panel: Pathogen sequence data sharing for emergency preparedness and response: developing consensus for 21st century sharing and analysis platforms
Peter van Heusden	International Center for Journalists seminar	12 August	Invited Talk: What to Know When Reporting on the Delta Variant – for journalists reporting on SARS-CoV-2 Variants of Concern (https://www.icfj.org/news/what-know-when-reporting-delta-variant)
Alan Christoffels	Global Alliance for Genomics and Health (GA4GH) 9th Plenary Meeting, Virtual Conference	28-29 September	Keynote: The role of PHA4GE in response to barriers in global (pathogen) data sharing
Alan Christoffels	Health Data Governance Summit, WHO	30 September	Invited session: What needs to change to implement good data governance practices?

Alan Christoffels	7th African Conference on OneHealth and Biosecurity, Civic Centre, Lagos, Nigeria	27–29 October	Invited talk: Importance of genomics in the management of infectious diseases and epidemics Theme: Universal approach to addressing biosecurity threats: genomics intelligence and vaccines
Peter van Heusden	Nnamdi Azikiwe University Genomics and Bioinformatics Consortium seminar: Bioinformatics in the Nigerian Landscape: Opportunities, Limitations and the Way Forward	25–27 November	Invited talk: Travelling the Road Towards Genomic Surveillance in Africa
Nicki Tiffin	United States–Mali Research Ethics Training Program Workshop: Ethical considerations in biospecimen sharing across borders.	1–3 December	Invited speaker: Capacity Building in resource-limited settings to overcome challenges to sample and data sharing

Seminars Presented

Presenter	Conference Name	Date	Title	Type of presentation
Peter van Heusden	ASCB Pathogen CoSI Virtual Event	04 March	Mycobacterium tuberculosis bioinformatics and analysing Mycobacterium tuberculosis pathogen samples with Galaxy	Seminar
Dominique Anderson, Alan Christoffels	African Society for Laboratory Management – LabCoP Extended ECHO Session Virtual Event	25 March	Leveraging Laboratory Information Management Systems in the COVID-19 Pandemic: a case study of Baobab LIMS	Seminar

Conference Participation

Presenter	Conference Name	Date	Title	Type of presentation
Oluwafemi Peter Abiodun	PathRed Congress 2021	June	Exploring the influence of organisational, environmental, and technological factors on information security policies and compliance at South African higher education institutions: implications for biomedical research	Oral
Ruben Cloete	ISCB–Africa ASBCB Conference on Bioinformatics Virtual Event	07–10 June	Virtual screening of three SARS–CoV–2 coronavirus proteins to identify novel inhibitors.	Oral
Dominique Anderson	ESBB Africa Conference	7–9 September	‘Biobanking and me’ – a bilingual speaking book for effective community engagement	Oral
Nicki Tiffin	United States–Mali Research Ethics Training Program Webinar	23 November	Ethics in data and biospecimen sharing in collaborative research	Oral

RESEARCH LABORATORIES

Dr Dominique Anderson

Research projects

My research group is investigating several areas involved in dental metagenomics, data security and privacy, biomedical data governance, informatics of biobanking and OneHealth. I have an ongoing role in a team focused on informatics solutions for biobanking with continued development, enhancement, and training in the open-source Baobab laboratory information management system (LIMS).

Information management and quality management remains a core area of activity in my research group, with hopes of expanding the collaborative network of developers and researchers. The pandemic highlighted the need for the LIMS to become more metadata aware, and so we focused on incorporating viral contextual metadata into the system, developed by the PHA4GE consortium. A recent collaboration has been established with the NHLS to evaluate data quality and metadata standards with a view to improving data linkage in data marts in the public health setting.

With regard 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. In addition to this work, the UWC Dentistry faculty and SANBI have an ongoing collaborative effort to increase research into dental genetics, particularly the analysis of microbiomes, including both wet-bench and dry-bench approaches. I am also currently working towards establishing networks within the agricultural, veterinary, medical, and environmental sphere to intensify research in the area of OneHealth and machine learning, and am also involved in community engagement and effective science communication and training.

Research Collaborations

1. Data governance and the POPI Act

Collaborating parties:

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

Nature and purpose:

To investigate the impact of local and international regulations on the sharing of biomedical research data in Africa.

Drafting committee member for the POPIA code of conduct for research

Output in the last 12 months:

Adams R, Adeleke F, Anderson D, Bawa A, Branson N, Christoffels A, et al. POPIA Code of Conduct for Research. S Afr J Sci. 2021;117(5/6), Art. #10933. [<https://doi.org/10.17159/sajs.2021/10933>]

2. Dental genetics

Collaborating parties:

Drs Manogari Chetty and Anthea Jetha – Faculty 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.
Impact of dental health in disease.



RESEARCH THEMES

- Informatic Tools
- Biomedical Data Governance
- Dental Genetics
- One Health

Highlights

Committee member of the draft POPI code of conduct for research

1 Book chapter and 1 journal publication contributed

Supervisor of 2 PhD students and 2 MSc students

Co-supervisor of 1 PhD student, 1 MSc student

3. Biobank informatics and data quality management

Collaborating parties:

Prof Alan Christoffels – SANBI, UWC
NHLS and University of the Witwatersrand
Africa CDC
PHA4GE Consortium

Nature and purpose:

Enhancement and customisation of Baobab LIMS for LMICs.
Standardisation of e-infrastructure.
Evaluating the impact of data standardisation in public health. Data linkage algorithms in data warehouses.

Output in the last 12 months:

Africa CDC funding.
One MSc graduate.
Book chapter: Dominique Anderson, Hocine Bendou, Bettina Kipperer, Kurt Zatloukal, Heimo Müller and Alan Christoffels. Biobanking in Low and Middle Income Countries: Relevance, Setup and Management. Springer International Publishing.

Future direction:

New collaborations within Africa as well as training initiatives on the continent. Maintenance and enhancement of Baobab LIMS. Entrepreneurship focused on hybrid social and sustainability models for QM software.

4. OneHealth

Collaborating parties:

Seeking collaborative partners

Nature and purpose:

Multidisciplinary approaches to understanding infectious diseases.
Development of metadata standards, novel AI-based tools for surveillance and species cross-over and translational research in OneHealth.

“I am also currently working towards establishing networks within the **agricultural, veterinary, medical, and environmental sphere** to intensify research in the area of **OneHealth** and **machine learning**.”

– Dr Dominique Anderson

Dr Hocine Bendou

Highlights

In August 2021, my PhD student, Abdulazeez Giwa, submitted for examination his thesis entitled, “Computational analysis of multi-omic data for the elucidation of molecular mechanisms of neuroblastoma”.

In 2021, my research group identified a methylation CpG signature capable of predicting the amplification of the gene MYCN, a predictor of poor prognosis, in the early stages of neuroblastoma. This could help clinicians design better treatment for patients who show MYCN amplification. The work was published in the journal Future Oncology.

Nasr Eshibona, one of my PhD students, identified three genes that may be responsible for the poor prognosis associated with the FLT3-ITD mutation in neuroblastoma patients. A research manuscript on this work is currently under review.

A collaborative research project with Dr Carmen Pheiffer from the South African Medical Research Council was launched to study the effect of physical training on the expression of miRNAs in gluteal adipose tissue in women with obesity. An NIH grant application with my postdoctoral student Dr Catherine Rossouw was also submitted, aimed at the early-stage development of informatics technologies for cancer research and management.

Research Collaborations

1. Molecular mechanisms of neuroblastoma

Collaborating parties:

Dr Azeez Fatai – Department of Biochemistry, Lagos State University, Lagos, Nigeria
Abdulazeez Giwa – SANBI, UWC

Nature and purpose:

Elucidate the molecular mechanisms of neuroblastoma in high-risk patients using multi-omics data and machine learning approaches.

Output in the last 12 months:

Publication article in Future Oncology (3.4 impact factor): Predicting amplification of MYCN using CpG methylation biomarkers in neuroblastoma.

2. Molecular biology and genetics of oesophageal cancer

Collaborating parties:

Prof Iqbal Parker – University of Cape Town

Nature and purpose:

Provide bioinformatics expertise and PhD student co-supervision.

Output in the last 12 months:

Whole genome sequencing analysis on paired tumour-normal samples from 60 participants.
Prepare and publish a manuscript.

Future direction:

Wet lab work for validation of the identified somatic mutations.



RESEARCH THEMES

- Precision medicine in paediatric cancer
- Pipeline development
- Gene regulatory networks
- Whole genome sequencing
- miRNA analysis

3. Effect of physical training in miRNA expression in gluteal adipose tissue in women with obesity

Collaborating parties:

Dr Carmen Pheiffer – South African Medical Research Council

Nasr Eshibona – SANBI UWC

Nature and purpose:

Provide bioinformatics expertise

Output in the last 12 months:

Manuscript submitted to an international journal with a high impact factor.

Professor Alan Christoffels

Highlights

During 2021, we maintained an active research programme. Our work on COMBAT-TB has allowed us to leverage funding from Africa CDC to develop a data analytics platform to analyse SARS-CoV-2 sequencing data. The lab continues to work with the Africa CDC to conceptualise a data sharing platform. In October 2021, we formed a partnership with WHO-AFRICA regional office (WHO-AFRO) to establish a regional centre for bioinformatics and genomic surveillance that would support and strengthen the SADC region in response to disease outbreaks, including COVID-19.

Members of our group are active in the global PHA4GE consortium. Specifically, our efforts have been directed at strengthening national public health labs in resource-limited settings in the data analytics space. I have engaged various international forums to promote our public health bioinformatics work.

A total of two PhD and two MSc students completed their degrees, and another two students submitted their theses in 2021. The hard work of my students has resulted in a total of 14 publications. The research of one of my PhD graduates, Catherine Rossouw, was selected as one of the top 11 papers to have impacted the field of clinical research and pathology in 2021.

Research Projects

The projects below underpin our translational work:

Tuberculosis

A comprehensive research programme is underway that includes investigating genetic determinants in both host (human) and pathogen (*Mycobacterium tuberculosis*) to understand drug resistance, and protein structure determination to inform patient-centric drug design. These findings inform the development of a scientific workflow management system to support reproducible high throughput computational experiments. These workflows form the basis of our COMputational Bacterial analytical toolkit for tuberculosis research (COMBAT-TB) (www.combattb.org) initiative to deploy analytic tools across the African continent. This analytics platform has been customised for SARS-CoV-2 analysis.

Biobanking

In collaboration with Dr Dominique Anderson at SANBI, we continue to strengthen our biobank software capabilities so that genetics data and biospecimens can be curated and stored.

Medicinal plant target prioritisation

The medicinal benefits of plant extracts generate a plethora of data. In collaboration with the UWC School of Pharmacy, we are extending our predictive analytics capability to prioritise candidate molecules for drug-resistant bacteria. This builds on our work on antimalarial targets.

Applications of machine learning methods to protein-protein interactions

Recently, we used quadruplet amino acids as a feature in a training model for protein interactions. We are extending this work to improve on the accuracy of our models with small datasets.



RESEARCH THEMES

- Pathogen genomics
- Pathogen surveillance
- Drug discovery

RESEARCH COLLABORATIONS

1. Developing a biobank informatics management system

Collaborating parties:

Dr Dominique Anderson – SANBI, UWC
Africa Centres for Disease Control 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:

Added a freezer monitoring schedule. Submitted a book chapter on biobank software for low- and middle-income countries.

2. Analysing genetic networks in Mycobacterium 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 is currently benchmarked against two international methods.

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

3. Identification of novel drug targets for drug resistant tuberculosis

Collaborating parties:

Prof Samantha Sampson and Prof Rob Warren – University of Stellenbosch
Dr Ruben Cloete – SANBI, UWC
Prof Sarel Malan – School of Pharmacy, UWC
Prof Valery Danilenko and Dr Dmitry Maslov – Vavilov Institute of General Genetics, Russia

Nature and purpose:

Current tuberculosis drugs are more than 30 years old and have unacceptable efficacy and safety profiles, emphasising the need for new drugs. We published a paper on kinase inhibitors in Mycobacterium tuberculosis, and efflux pump inhibitors.

Output in the last 12 months:

We continue to refine our computational screening system. And submitted a funding application via the BRICS-country funding mechanism.

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

Collaborating parties:

Prof Sarel Malan & Dr Samuel Egieyeh – School of Pharmacy, UWC.

Nature and purpose:

At least 1000 compounds were identified in plant extracts, showing antiplasmodial activity. In the past we have successfully used a range of computational tools to compare plant extracts that show antiplasmodial activity compounds with currently registered antimalarial drugs to identify common signatures and ultimately prioritise the list of compounds identified in plant extracts. We are now utilising a similar strategy for drug resistant bacteria.

Output in the last 12 months:

One MSc student in the Pharmacy department has completed his thesis on "Hit prioritisation of natural products with activities against methicillin-resistant staphylococcus aureus (MRSA)".

5. Computational bacterial analytical toolkit for tuberculosis (COMBAT-TB)

Collaborating parties:

Prof Samantha Sampson and Prof Rob Warren – University of Stellenbosch
Peter van Heusden – SANBI-UWC

Nature and purpose:

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

Output in the last 12 months:

Added further analysis tools and a user manual. Our manuscript was published in mSphere Journal.

6. Identification of SNPs implicated in rare diseases

Collaborating parties:

Prof Manogari Chetty – Dentistry Faculty, UWC.

Nature and purpose:

We utilise bioinformatics tools to support the dental faculty to analyse the genetic basis of some of the syndromes observed in dental clinics, including Osteogenesis Imperfecta.

Output in the last 12 months:

Completed the MSc thesis for a project to analyse whole genome sequencing data from patients with Osteogenesis Imperfecta III.

7. 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 establish 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 South African universities with a view to propose guidelines for data management that impact biomedical researchers.

8. 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:

PHA4GE data standards working group developed a metadata standard for COVID-19 biospecimen collection.

9. African genomics archive

Collaborating parties:

Africa CDC

Nature and purpose:

To establish a data platform in Africa to manage disease surveillance data with a view to strengthen public health institutes.

Output in the last 12 months:

A series of workshops to gather information on various use cases.

“The research of one of my PhD graduates (Catherine Rossouw) was selected as **one of the top 11 papers to have impacted the field** of clinical research and pathology in 2021.”

“The hard work of my students has resulted in **a total of 14 publications.**”

– Prof Alan Christoffels

Dr Ruben Cloete



RESEARCH THEMES

- Pathogen resistance
- Drug discovery
- Human disease genomics
- Machine learning approaches
- Diagnostic tool development

Highlights

My UWC MSc students, Darren Isaacs and Maryam Hassan, graduated in 2021.

I currently supervise one PhD student, Mohammed Hassan, and co-supervise two students, Katelyn Kuttler from Stellenbosch University and Darius Martin from the Biotechnology Department, UWC.

Two new MSc students joined my group in 2021: Keaghan Brown (University of Stellenbosch) and Lunathi Mhlahi Ndondlo (UWC).

Grants/Funding Awarded: Co-Principal Investigator on a SAMRC RCDI grant with Prof Megan Shaw for 2021-2022.

Research Projects

My primary research interests focus on understanding drug resistance in Mycobacterium tuberculosis and Human Immunodeficiency Virus (HIV-1), and identifying causal variants in Parkinson's disease development.

Tuberculosis and HIV-1 drug resistance is a huge problem in South Africa, and requires the identification of new 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 and to screen for drugs that inhibit enzyme targets and can be 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 useful in determining whether a drug remains within an enzyme's active pocket in a mutated protein, and if the mutation results in reduced binding or no binding. The results of this research may provide improved treatment regimens for South African populations to combat infectious diseases, as well as non-communicable diseases like Parkinson's.

Recently, my group also embarked upon research to identify novel drugs to treat SARS-CoV-2 coronavirus infections using computational methods. Other collaborative projects involve studying drug transporter proteins associated with diabetes mellitus and investigating structural differences between human coronavirus envelope proteins to understand differences in pathogenicity between virulent and non-virulent strains.

Research Collaborations

1. Novel drug discovery against Mycobacterium tuberculosis

Collaborating parties:

Prof Alan Christoffels – SANBI, UWC
Prof Samantha Samson and Dr Melanie Grobbelaar – 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: Cloete R, Shahbaaz M, Grobbelaar M, Sampson SL, Christoffels A. In silico repurposing of a Novobiocin derivative for activity against latency associated Mycobacterium tuberculosis drug target nicotinate-nucleotide adenyllyl transferase (Rv2421c). PLoS One. 2021 Nov 2;16(11):e0259348. Doi: 10.1371/journal.pone.0259348. PMID: 34727137.

Future direction:

Draft two manuscripts, one targeting Mycobacterium tuberculosis drug target Rv2196, and a follow-up article on Rv2421c to identify novel drug molecules to treat tuberculosis infections.

2. Structural impact of resistance associated mutations in the South African HIV-1C integrase protein

Collaborating parties:

The late Dr Graeme Jacobs – Stellenbosch University [deceased]

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:

Mikasi SG, Isaacs D, Chitongo R, Ikomey GM, Jacobs GB, Cloete R. Interaction analysis of statistically enriched mutations identified in Cameroon recombinant subtype CRF02_AG that can influence the development of Dolutedegravir drug resistance mutations. *BMC Infectious Diseases*. 2021;21(1):1–12.

Future direction:

Two manuscripts in preparation, one focusing on the effect of Raltegravir resistant mutations on the HIV-1C integrase structure and another on the development of an automated pipeline to investigate the effects of mutations on HIV-1C and HIV-1 AG recombinant integrase structure.

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:

1) Sebate B, Cuttler K, Cloete R, Britz M, Christoffels A, Williams M, et al. Prioritization of candidate genes for a South African family with Parkinson's disease using in silico tools. *PLoS One*. 2021;16(3):e0249324.

2) Cuttler K, Hassan M, Carr J, Cloete R, Bardien S. Emerging evidence implicating a role for neurexins in neurodegenerative and neuropsychiatric disorders. *Open Biol*. 2021 Oct;11(10):210091. Doi: 10.1098/rsob.210091. Epub 2021 Oct 6. PMID: 34610269; PMCID: PMC8492176.

Future direction:

Two manuscripts under preparation for publication following up on neurexin and neuroligin interaction associated with Parkinson's disease and another on a novel gene candidate identified in a South African family with Parkinson's disease using exome sequencing.

4. Haplotype variation within a South African Xhosa population and its effect on diabetic treatment

Collaborating parties:

Prof Mongi Benjeddou – Biotechnology Department, UWC

Nature and purpose:

Understand the effect of haplotype variation on SLCAA2 transporter protein and the binding of diabetic drugs to SLCAA2 within a South African Xhosa population.

Output in the last 12 months:

None

Future direction:

Manuscript submitted for publication.

5. Structural studies of more and less virulent coronavirus envelope proteins to understand human host interaction and severity of disease

Collaborating parties:

Prof Burtram Fielding – Department of Medical Biosciences, UWC

Nature and purpose:

Perform topology predictions, structural modelling as well as simulation studies to understand the structural differences between four coronavirus envelope proteins and their interaction with the human host PALS1 protein.

Output in the last 12 months:

None

Future direction:

Manuscript in preparation.

6. The identification and characterisation of DNA aptamers for application in diagnosis of infectious diseases

Collaborating parties:

Prof Mervin Meyer – Biotechnology Department, UWC

Nature and purpose:

The aim of the research is to develop multiplex lateral flow devices (LFDs) for the detection of serum human biomarker proteins for tuberculosis and Ebola diagnosis using DNA aptamers.

Output in the last 12 months:

Martin DR, Sibuyi NR, Dube P, et al. Aptamer-Based Diagnostic Systems for the Rapid Screening of TB at the Point-of-Care. *Diagnostics* (Basel, Switzerland). 2021 Jul;11(8). DOI:

10.3390/diagnostics11081352. PMID: 34441287; PMCID: PMC8391981.

[<https://doi.org/10.3390/diagnostics11081352>]

Future direction:

To draft a research article on the identification of potential human biomarkers that can be targeted by DNA aptamers for tuberculosis diagnosis. Complete an in silico research paper on the development of novel DNA aptamers against Ebola virus nucleoproteins as well as drafting a review article on Ebola virus diagnostic tools.

Dr Gordon Harkins

My research focuses on the evolution and molecular epidemiology of DNA and RNA infectious viral pathogens, where I seek to determine the evolutionary underpinnings of the emergence and spread of the numerous viral diseases that seriously threaten the health and food security of Africa and the rest of the developing world.

In early 2020, in response to the emergence and rapid early spread of COVID-19, we immediately shifted our research focus to tackling the unprecedented volume of SARS-CoV-2 genomic and spatial data that were being generated and shared with the scientific community to gain real-time insights into the virus transmission during a viral pandemic.

Highlights

Awarded South African Medical Research Council (SAMRC) funding to predict and monitor T cell immune escape mutations at HLA-binding anchor motifs found in SARS-CoV-2 epitopes from genomes isolated in South Africa and sub-Saharan Africa. Principal investigator: Nobubelo Ngandu.

Appointed as Director of the Centre of Excellence (CoE) in HIV Prevention at UWC funded by the South African Department of Science and Innovation (DSI) and the National Research Foundation (NRF).

Five manuscripts were submitted for publication in 2021, three of which were published within the calendar year and the fourth in 2022. Included as co-authors on one of the 2021 articles were my Honours student from 2020, Sabine Smidt, and a previous MSc student of mine, Dr Batsirai Mabvakure.

The following two papers have been written. The first has been accepted with minor corrections in the journal *Virus Evolution* and the second is under review in the journal *Phytopathology*.

“Prediction and monitoring of T cell immune escape mutations at HLA-binding anchor motifs found in SARS-CoV-2 epitopes from genomes isolated in South Africa and sub-Saharan Africa”.

“Virion-associated nucleic acid-based metagenomics: a decade of advances in molecular characterisation of plant viruses”.

RESEARCH PROJECTS

A summary of some of the research projects that my laboratory has been involved in during 2021 is presented below.

SARS-CoV-2

The characterisation and clinical manifestation of the SARS-CoV-2 outbreak in Uruguay

Although South America was mostly spared in the early months of the SARS-CoV-2 pandemic, it was severely hit with the arrival of the autumn season in the Southern Hemisphere, and the virus has since continued to ravage Latin America. In contrast, Uruguay, a small country located south of Brazil, has become known for curtailing

SARS-CoV-2 exceptionally well. In order to attain a better understanding of disease patterns and regional aspects of the pandemic in Latin America, we combined spatial and genetic analyses to infer the global historical dispersal dynamics of the causative virus SARS-CoV-2. We showed that most virus introductions into Uruguay originated from other South American countries, with the earliest seeding of the virus occurring weeks before the borders were closed to all non-citizens and a partial lockdown implemented.



RESEARCH THEMES

- SARS-CoV-2
- SARS-CoV-2 Phylogeography
- SARS-CoV-2 Dynamics
- SARS-CoV-2 Evolution
- HIV-1 Latent Viral Reservoir Dynamics
- Phytovirus Species Richness
- Viral Metagenomics

The evolutionary dynamics of SARS-CoV-2 lineages during the first epidemic wave in New York City.

During the first phase of the COVID-19 epidemic in the United States, New York City rapidly became an epicentre of the pandemic. In order to gain insights into the dispersal history and transmission dynamics of SARS-CoV-2 during the first months of the New York City epidemic, my collaborators and I analysed the genomic and spatial data using a Bayesian phylogeographic approach to infer the dispersal history and dynamics of viral lineages at the state and city levels. Our findings showed that peripheral samples likely correspond to distinct dispersal events originating from the central city area, and that the borough of Queens was a relatively important transmission hub associated with higher local circulation and dispersal of viral lineages to the surrounding boroughs.

The role of natural selection in the emergence and ongoing evolution of the N501Y lineages (V1-V3) of SARS-CoV-2.

The independent emergence and rapid rise in prevalence of three SARS-CoV-2 "501Y" lineages since October 2020 has raised renewed concerns about the evolutionary capacity of SARS-CoV-2 to adapt to both public health interventions such as vaccines and social distancing, and rising population immunity. In collaboration with researchers from the South African SARS-CoV-2 genomics task force, we analysed the genomic and spatial data for the three existing 501Y lineages and found that they had likely independently acquired multiple unique and convergent mutations and that a major change in the selective forces acting on immunologically important SARS-CoV-2 lineages coincided with the emergence of the 501Y lineages.

Human immunodeficiency virus (HIV-1)

Investigation of the viral dynamics involved in HIV-1 subtype C latent reservoir formation, maintenance and evolution.

The HIV-1 latent reservoir is established during acute infection, and viral variants from both early infection time points, as well as time points proximal to ART initiation, can be identified in the cellular long-lived reservoir. However, it is unknown if different distinct compartments in the body contribute to the latent reservoir in different ways and to what extent the

reservoir contributes to viral evolution during viremia by reseeding of infection. We are currently generating the necessary data to elucidate the contribution of viral variants from samples collected from the blood and cervix to the long-lived latent reservoir in a longitudinal sample of subtype C infected women from the CAPRISA 002 cohort. Our analyses will provide estimates of the timing of the establishment of the reservoir and the most probable source, along with the dates, magnitude and direction of the independent movements between the different tissue compartments within each individual over the course of infection.

Reservoir size determination and timing of entry of HIV-1 variants into the latent reservoir

Understanding the key determinants of HIV-1 latent reservoir establishment, size and maintenance is imperative to designing appropriate HIV cure interventions. We are making use of longitudinal next-generation sequence data from a population of subtype C infected women from the CAPRISA 002 cohort to elucidate viral evolution patterns and generate a pre-treatment evolutionary timeline with which to date viruses isolated from the long-lived latent reservoir. These analyses will aid in better understanding the establishment of viral reservoirs in individuals who initiate therapy. The proposed work will contribute to a larger study investigating reservoir dynamics that will also include measurement of reservoir size and viral promoter function and nef gene function.

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, Arizona State University and UCT, we received funding from the French National Research Agency (ANR) in 2019 for a project that had three primary objectives: (1) to test whether plant species richness influences phytovirus species richness in natural and cultivated areas; (2) to experimentally study the effect of plant communities on phytovirus species richness; and (3) to search for evolutionary footprints associated with emergence within phytovirus genomes.

RESEARCH COLLABORATIONS

1. The characterisation and clinical manifestation of the SARS-CoV-2 outbreak in Uruguay

Collaborating parties:

Victoria Elizondo, Victoria Perez – Natalia Mazza Carolina Beloso, Silvana Ifran, Mariana Fernandez, Andrea Santini, Veronica Perez, Veronica Estevez, Matilde Nin, Gonzalo Manrique, Leticia Perez, Fabiana Ross, Susana Boschi, Maria Noel Zubillaga, Raquel Ballestea – Laboratorio de Biología Molecular, Asociación Española Primera en Salud, Montevideo, Uruguay. Batsirai Mabvakure – Department of Medicine, Johns Hopkins School of Medicine, Baltimore, USA.

Paul Zappile Christian Marier – Genome Technology Center, Office for Science and Research, NYU Langone Health, New York, USA. Victoria Perez – Departamento de Desarrollo Biotecnológico, Instituto de Higiene, Facultad de Medicina, Udelar, Montevideo, Uruguay. Carolina Beloso – Departamento de Biodiversidad y Genética. Instituto de Investigaciones Biológicas Clemente Estable, Montevideo, Uruguay.

Matthew T. Maurano – Institute for Systems Genetics, NYU Grossman School of Medicine, New York, USA. Simon Dellicour – Spatial Epidemiology Lab. (SPELL), Université Libre de Bruxelles, Brussels, Belgium; Department of Microbiology, Immunology and Transplantation, Rega Institute, Leuven, Belgium. Adriana Heguy, Ralf Duerr – Department of Pathology, NYU Grossman School of Medicine, New York, USA.

Nature and purpose:

To identify SARS-CoV-2 introductions into Uruguay and their subsequent transmission dynamics to better understand disease patterns and regional aspects of the pandemic in Latin America.

Output in the last 12 months:

One paper was submitted in 2020 (and published in January 2021).

Future direction:

This is an ongoing collaboration that is now focused on generating the genomic sequence data from the second wave of the pandemic in Uruguay. Our priority will be detection of the presence of potential vaccine escape variants and other potentially problematic variants of concern.

2. The evolutionary dynamics of SARS-CoV-2 lineages during the first epidemic wave in New York City

Collaborating parties:

Simon Dellicour – Spatial Epidemiology Lab (SPELL), Université Libre de Bruxelles, Brussels, Belgium. Bram Vrancken, Samuel L. Hong, Guy Beale – Department of Microbiology, Immunology and Transplantation, Rega Institute, Leuven, Belgium. Matthew T. Maurano – Institute for Systems Genetics, NYU Grossman School of Medicine, New York, USA. Antoine Chaillon – Division of Infectious Diseases and Global Public Health, University of California San Diego, USA. Sitharam Ramaswami, Paul Zappile, Christian Marier – Genome Technology Center, Office for Science and Research, NYU Langone Health, New York, USA. Adriana Heguy, Ralf Duerr – Department of Pathology, NYU Grossman School of Medicine, New York, USA.

Nature and purpose:

Phylogeographic investigation to elucidate the circulation of viral lineages during the first months of the New York outbreak – the epicentre of the COVID-19 epidemic in the United States.

Output in the last 12 months:

One paper was published in PLoS Pathogens in 2021.

Future direction:

This is an ongoing collaboration that will shift its focus to the COVID-19 pandemic in other areas in the USA and beyond.

3. Investigating the role of natural selection in the emergence and ongoing evolution of the N501Y lineages (V1-V3) of SARS-CoV-2

Collaborating parties:

Darren P Martin – Institute of Infectious Diseases and Molecular Medicine, Division Of Computational Biology, Department of Integrative Biomedical Sciences, University of Cape Town, South Africa.

Steven Weaver, Stephen D Shank, Sergei L Kosakovsky Pond – Institute for Genomics and Evolutionary Medicine, Department of Biology, Temple University, Pennsylvania, USA.

Houryiah Tegally, Emmanuel James San, Eduan Wilkinson, Jennifer Giandhari, Sureshnee Naidoo, Yeshnee Pillay, Lavanya Singh, Richard J Lessells, Tulio De Oliveira – KwaZulu-Natal Research Innovation and Sequencing Platform (KRISP), School of Laboratory Medicine & Medical Sciences, University of KwaZulu-Natal, Durban, South Africa NGS-SA – (http://www.krisp.org.za/ngs-sa/ngs-sa_network_for_genomic_surveillance_south_africa/).

COVID-19 Genomics UK (COG-UK) – (<https://www.cogconsortium.uk>).

Ravindra K Gupta – Clinical Microbiology, University of Cambridge, Cambridge, UK; Africa Health Research Institute, KwaZulu-Natal, South Africa. Joel O Wertheim – Department of Medicine, University of California San Diego, La Jolla, CA 92093, USA. Anton Nekturenko – Department of Biochemistry and Molecular Biology, The Pennsylvania State University, Pennsylvania, USA.

Ben Murrell – Department of Microbiology, Tumor and Cell Biology, Karolinska Institutet, Stockholm, Sweden. Philippe Lemey – Department of Microbiology, Immunology and Transplantation, Rega Institute, KU Leuven, Leuven, Belgium. Oscar A MacLean, David L Robertson – MRC-University of Glasgow Centre for Virus Research, Scotland, UK.

Nature and purpose:

To examine temporal patterns of natural selection within protein coding sequences of the N501Y viral lineages since the emergence of SARS-CoV-2 in 2019.

Output in the last 12 months:

One paper was published in the journal Cell in 2021 .

Future direction:

Surveillance of the rapidly growing pool of SARS-CoV-2 genomic sequence data for the presence of potential vaccine escape variants and other potentially problematic variants of concern in South Africa and beyond will be continued.

4. Selection analysis identifies unusual clustered mutational changes in Omicron lineage BA.1 that likely impact spike function.

Collaborating parties:

Darren P Martin, Arne De Klerk – Institute of Infectious Diseases and Molecular Medicine, Division Of Computational Biology, Department of Integrative Biomedical Sciences, University of Cape Town, South Africa.

Philippe Lemey – Department of Microbiology, Immunology and Transplantation, Rega Institute, KU Leuven, Leuven, Belgium.

Steven Weaver, Stephen D Shank, Sergei L Kosakovsky Pond – Institute for Genomics and Evolutionary Medicine, Department of Biology, Temple University, Pennsylvania, USA.

Houryiah Tegally, Emmanuel James San, Eduan Wilkinson, Jennifer Giandhari, Richard J Lessells, Anton Nekrutenko, Tulio De Oliveira – KwaZulu-Natal Research Innovation and Sequencing Platform (KRISP), School of Laboratory Medicine & Medical Sciences, University of KwaZulu-Natal, Durban, South Africa NGS-SA.

Ben Murrell – Department of Microbiology, Tumor and Cell Biology, Karolinska Institutet, Stockholm, Sweden. Carolyn Williamson – Division of Medical Virology, Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Cape Town, South Africa.

Spyro Lytras, Oscar S MacLean, David L Robertson, Richard J Orton – MRC-University of Glasgow Centre for Virus Research, University of Glasgow, Glasgow G61 1QH, UK.

Alexander G Lucaci – Institute for Genomics and Evolutionary Medicine, Department of Biology, Temple University, Philadelphia, PA 19122, USA.

Wolfgang Maier, Bjorn Gruning – Bioinformatics Group, Department of Computer Science, University of Freiburg, Freiburg, Germany.

Maciej F Boni – Center for Infectious Disease Dynamics, Department of Biology, Pennsylvania State University, University Park, PA, USA.

Cathrine Scheepers, Jinal N Bhiman – National Institute for Communicable Diseases (NICD) of the National Health Laboratory Service, (NHLS), Johannesburg, South Africa, SA MRC Antibody Immunity Research Unit, School of Pathology, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa.

Josie Everatt, Daniel G Amoako – National Institute for Communicable Diseases (NICD) of the National Health Laboratory Service, (NHLS), Johannesburg, South Africa.

James Emanuel San, Jennifer Giandhari – KwaZulu-Natal Research Innovation and Sequencing Platform (KRISP), School of Laboratory Medicine & Medical Sciences, University of KwaZulu-Natal, Durban, South Africa.

Alex Sigal, Ravindra K Gupta – Africa Health Research Institute, Durban, South Africa. Nei-yuan Hsiao – Division of Medical Virology, University of Cape Town

and National Health Laboratory Service, Cape Town South Africa. Anne von Gottberg – National Institute for Communicable Diseases (NICD) of the National Health Laboratory Service, (NHLS), Johannesburg, South Africa. Robert W Shafer – Division of Infectious Diseases, Department of medicine, Stanford University, Stanford, USA. Robert J Wilkinson – Wellcome Center for Infectious Diseases Research in Africa, Institute of Infectious Disease and Molecular Medicine and Department of Medicine, University of Cape Town, South Africa. Brian Trevor Sewell – Structural Biology Research Unit, Department of Integrative Biomedical Sciences, Institute for Infectious Diseases and Molecular Medicine, University of Cape Town, South Africa. Allison J Greaney, Tyler N Starr, Jesse D Bloom – Basic Sciences Division and Computational Biology Program, Fred Hutchinson Cancer Research Center, Seattle; Department of Genome Sciences & Medical Scientist Training Program, University of Washington, Seattle, Howard Hughes Medical Institute, Seattle, USA.

Nature and purpose:

To determine what drives the mutational changes in the Omicron variant of concern (VOC).

Output in the last 12 months:

One paper published in 2022 in the Journal Molecular Biology and Evolution.

Future direction:

This is an ongoing collaboration.

5. Conserved recombination in coronavirus subgenera

Collaborating parties:

Darren P Martin, Arné de Klerk, Phillip Swanepoel, Mpumelelo Zondo, Isaac Abodunran – Institute of Infectious Diseases and Molecular Medicine, Division Of Computational Biology, Department of Integrative Biomedical Sciences, University of Cape Town. Rentia Lourens – Division of Neurosurgery, Neuroscience Institute, Department of Surgery, University of Cape Town. Spyros Lytras, Oscar A MacLean, David Robertson – MRC–University of Glasgow Centre for Virus Research, University of Glasgow, Glasgow. Sergei L Kosakovsky Pond, Jordan D Zehr – Institute for Genomics and Evolutionary Medicine, Department of Biology, Temple University, Philadelphia, USA. Venkatesh Kumar, Ben Murrell – Department of Microbiology, Tumor and Cell Biology, Karolinska Institutet, Stockholm, Sweden. Michael J. Stanhope – Department of Population and Ecosystem Health, College of Veterinary Medicine, Cornell University, Ithaca, New York, USA.

Nature and purpose:

To determine whether recombination is non-random in coronavirus subgenera.

Output in the last 12 months:

One article has been accepted in the journal *Virus Evolution*.

Future direction:

This is an ongoing collaboration.

6. T cell immune escape mutations at HLA-binding anchor motifs found in SARS-CoV-2 epitopes from genomes isolated in South Africa and sub-Saharan Africa

Collaborating parties:

Nobubelo Ngandu – SAMRC. Burtram Fielding – University of the Western Cape. Wendy Burgers – University of Cape Town. Zabrina Brumme – Simon Fraser University, Canada.

Nature and purpose: Prediction and monitoring of T cell immune escape mutations at HLA-binding anchor motifs found in SARS-CoV-2.

Output in the last 12 months: None

Future direction:

This is an ongoing collaboration funded by SAMRC for 2022.

7. HIV-1 latent viral reservoir dynamics

Collaborating parties:

Melissa-Rose Abrahams, Lynn Tyers, David Matten, Deelan Doolabh, Colin Anthony, Carolyn Williamson – Division of Medical Virology, Institute of Infectious Disease and Molecular Medicine, University of Cape Town. Salim Abdool Karim – Centre for the AIDS Programme of Research in South Africa, University of KwaZulu-Natal. Andrew Redd – Johns Hopkins University, School of Medicine in Baltimore, USA. Sipo sethu Matzishana, Nkosazama Nyembezi – SANBI.

Nature and purpose:

To study the viral dynamics involved in HIV-1 subtype C latent reservoir formation maintenance and evolution to better understand the determinants thereof.

Output in the last 12 months:

None.

Future direction:

This project is funded by the National Institutes of Health (NIH) USA and the South African Medical Research Council and will run between 2020 and 2025.

8. Phytovirus diversity

Collaborating parties:

Philippe Roumagnac, Denis Filloux, Charlotte Julian and Emmanuel Fernandez – CIRAD, Montpellier, 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 Benoit 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 (1) to develop an integrated approach to reveal the phytoviral species richness of plant communities within several unmanaged and managed ecosystems, (2) to experimentally study the effect of controlled plant communities on phytovirus species richness, and (3) 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:

Plant samples were first collected in South Africa in 2019 but the field sampling trips planned for 2020 and 2021 were postponed until October 2022 and scheduled annually thereafter until 2025, pandemic conditions permitting. Efforts to detect phytoviruses in our plant samples are currently being performed in the laboratory of Philippe Roumagnac in Montpellier, France.

Future direction:

This is an ongoing project that is funded until 2023.

“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.”

– Dr Gordon Harkins

DR ULJANA HESSE

[Senior Lecturer, UWC Department of Biotechnology]

South Africa is home to an exceptional biological resource – the unique flora of the Cape Floristic Region, which 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 (*Aspalathus linearis*), 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 recognised as a potential phytopharmaceutical, as the species produces a wide range of phenolic compounds, which are associated with diverse medicinal properties of the plant (including anti-diabetic, cardioprotective, antispasmodic and anti-aging effects).

Rooibos 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 Program aims to:

- locally establish all methodologies essential for medicinal plant genome analysis.
- generate a high-quality assembly of the rooibos genome with extensive functional annotation.
- identify genes involved in medicinal compound production and plant stress tolerance.

Collaboration with SANBI: I have a longstanding partnership with SANBI that dates back to my postdoctoral fellowship at the Institute (2011–2013). Since the initiation of the Rooibos Genomics Program, all postgraduate students of my research team have received training in bioinformatics through the one-month course offered by SANBI. Establishment of biocomputational data analysis procedures is conducted or initiated at SANBI.

My research team includes PhD students, Fanechka Esterhusen and Yamkela Mgwatyu, MSc students, Casey Lee and Tanweer Beckett and Honours student Cailyn Joy Fortuin.

Highlights

In 2021, my research team has:

Optimised third generation sequencing procedures for the analysis of plant genomic DNA using MinION from Oxford Nanopore (my research team is the first at UWC to use this technology).

Generated long read (2–112kb) sequencing data for the rooibos genome (30x genome coverage).

Evaluated assembly procedures for joint analysis of Illumina and MinION sequencing data and generated a publishable genome assembly for rooibos.

Established procedures for plant organelle genome assembly and assembled the chloroplast and mitochondrial genomes of rooibos.

Established a biocomputational pipeline for comparative transcriptome analysis and identified first rooibos genes involved in drought stress tolerance.

Investigated convolutional neural networks towards establishing machine learning algorithms for plant protein classification.

RESEARCH THEMES

- Rooibos genomics

PROFESSOR NICKI TIFFIN

I rejoined the SANBI faculty in November 2021, bringing my expertise in working with health data. My work includes the integration and analysis of routine health data, epidemiological, clinical and genomic data, as well as the ethics and governance of working with health data in research. I lead the public health node of the H3Africa Informatics Network (H3ABioNet), leading the genetic epidemiology project; I co-lead the ICDA Global Equity working group, co-chair the IHCC Policy and Systems working group and chair the PHA4GE Ethics and Data Sharing working group. I also collaborate with the Provincial Health Data Centre at the Western Cape Department of Health, working as a data analyst and epidemiologist to support this routine health data platform. I am a fellow of the African Academy of Sciences and in 2021 was awarded a Calestous Juma Science Leadership Fellowship by the Bill & Melinda Gates Foundation.



RESEARCH THEMES

- Health data integration
- Facilitating sharing of African data and biospecimens
- Ethics and governance for research in Africa

RESEARCH COLLABORATIONS

1. Multimorbidity in the Western Cape

The focus of this project is the management, integration and analysis of health data from health care clients attending government health facilities in the Western Cape.

Optimising data linkage of African health client data: When integrating individualised data from a variety of sources, different linkage algorithms can be used to ensure the most accurate linkage and deduplication of records that originate from the same individual. Most of these algorithms, however, have been developed using Eurocentric datasets and do not necessarily perform as well in African contexts. We are also working on a systematic approach to improve data linkage for routine health data.

Researcher: Themba Mutemaringa

A virtual Khayelitsha cohort: We are analysing anonymised and perturbed routine health data from a virtual cohort of all health care clients who attended any government health facility in the Khayelitsha sub-district in Cape Town during 2016 and 2017. Analysis of these data provides insights into multimorbidity in this population group, which suffers from a high burden of both infectious and non-communicable diseases.

Researchers: Richard Osei-Yeboah, Tsaone Tamuhla

Piloting a virtual genotyped population cohort linking genotype and routine health data: With informed consent from health care clients, it is possible to link individuals' genotype data with their routine health data in order to establish a genotyped virtual cohort with clinical phenotype data that can be updated into the future using routine health records. We are piloting this approach by establishing the informed consent process, sample and data workflows to create, in the first instance, a virtual genotyped cohort for patients with diabetes in Cape Town.

Researcher: Tsaone Tamuhla

Recent publications

Dave JA, Tamuhla T, Tiffin N et al. (2021) Risk factors for COVID-19 hospitalisation and death in people living with diabetes: A virtual cohort study from the Western Cape Province, South Africa. *Diabetes Res Clin Pract.* Jun 21;177:108925. doi: 10.1016/j.diabres.2021.108925

Tamuhla T, Dave JA, Raubenheimer P, Tiffin N (2021) Diabetes in a TB and HIV-endemic South African population: Analysis of a virtual cohort using routine health data. *PLoS ONE* 16(5): e0251303. [<https://doi.org/10.1371/journal.pone.0251303>]

Osei-Yeboah R, Tamuhla T, Ngwenya O, Tiffin N (2021) Accessing HIV care may lead to earlier ascertainment of comorbidities in health care clients in Khayelitsha, Cape Town. *PLOS Glob Public Health* 1(12): e0000031. [<https://doi.org/10.1371/journal.pgph.0000031>]

Allie T, et al. [Senior author Tiffin N], (2021). TBDBT: A TB DataBase Template for collection of harmonised TB clinical research data in REDCap, facilitating data standardisation for inter-study comparison and meta-analyses. *PLoS ONE* 16(3): e0249165. [<https://doi.org/10.1371/journal.pone.0249165>]

2. Ethics and governance to promote ethical and equitable health research in Africa

In order to undertake ethical research in Africa and other regions of the Global South, there are multiple considerations around risks, benefits, informed consent and equitable research. We explore different ways to better do research and to ensure that health research in Africa is ethical and equitable through the development of tools and frameworks. I also participate in multiple ethics and governance working groups of international consortia including PHA4GE [Chair: Ethics and Data Sharing Working Group], ICDA [co-Lead: Global Equity Working Group] and IHCC [co-Chair: Systems and Policy Working Group]. In 2021 I initiated a forum in which the ELSI working groups from the many different large health genomics research networks and consortia are able to meet and share information about their activities, challenges and resources.

3. Facilitating ethical and equitable sharing of African data and biospecimens: the African Data and Biospecimen Exchange ADBEx

Towards the end of 2021 I established a programme to build an online platform to facilitate ethical and equitable secondary sharing of African data and biospecimen resources. The queryable platform will store metadata about sample and data collections, submitted by researchers who hold those resources, and will enable direct sharing partnerships to be established between resource providers and resource consumers without centralisation of samples or data. It will also provide online infrastructure for drawing up and recording data-sharing, benefit-sharing and collaborative agreements as well as building a repository of data and biospecimen guidelines and governance resources.

Students

I currently supervise two full-time PhD students [Tsaone Tamuhla and Richard Osei-Yeboah], two part-time PhD students [Themba Mutemaringa and Florence Phelanyane] and one MPH student [Pierre Dane] who all registered for their degrees under my supervision at the University of Cape Town before I moved to SANBI. I also co-supervise a PhD student at the University of Pretoria [Fatima Barmania], and at the University of Cape Town [Ryan Aylward]. These students have a variety of projects working with large routine health datasets as well as conducting pilot studies to link genomic and routine health data.

Funding

In November 2021, the Bill & Melinda Gates Foundation launched the Calestous Juma Science Leadership Fellowship awards, providing five years of funding for fellows as well as opportunities for networking and training in non-scientific skills. These fellowships aim to establish a collaborative cohort of African fellows with a wide range of expertise and research interests. My Calestous Juma Fellowship in the data science track will fund the creation of the African Data and Biospecimen Exchange [ADBEx]. In addition, we are entering the final year of H3Africa funding for the Public Health node of H3ABioNet, which continues to fund the contributions of Tsaone Tamuhla as she completes her PhD, and myself.

COLLABORATIONS

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

CANADA

- University of British Columbia
- BC Centre for Disease Control
- Simon Fraser University

USA

- USA CDC
- Broad Institute
- Johns Hopkins School of Medicine
- Genome Technology Center, Office for Science and Research
- NYU Grossman School of Medicine
- University of California San Diego
- Temple University
- Pennsylvania State University
- Stanford University
- Fred Hutchinson Cancer Research Center
- University of Washington
- Howard Hughes Medical Institute
- College of Veterinary Medicine, Cornell University
- Arizona State University

GLOBAL CONSORTIA

- Africa CDC
- GET Africa Global Consortium on Emerging Infectious Diseases
- PHA4GE Consortium
- H3Africa

URUGUAY

- Laboratorio de Biología Molecular, Asociación Española Primera en Salud
- Departamento de Desarrollo Biotecnológico, Instituto de Higiene, Facultad de Medicina, Udelar
- Instituto de Investigaciones Biológicas Clemente Estable



UNITED KINGDOM

- School of Law, Middlesex University
- Oxford University
- University of Birmingham
- COVID-19 Genomics UK (COG-UK)
- University of Cambridge
- MRC-University of Glasgow Centre for Virus Research

BELGIUM

- Université Libre de Bruxelles Rega Institute

SWEDEN

- Karolinska Institutet

GERMANY

- University of Freiburg

FRANCE

- CIRAD, Montpellier
- CIRAD, UMR PVBMT, Réunion Island
- INRA, UMR BFP, Villenave d'Ornon
- ANSES
- La Tour du Valat
- UMR Evo-Eco-Paleo

RUSSIA

- Vavilov Institute of General Genetics

SOUTH AFRICA

- National Institute for Communicable Diseases
- National Health Laboratory Service
- South African Medical Research Council
- University of Cape Town
- University of Stellenbosch
- University of the Western Cape
- University of the Witwatersrand
- KwaZulu-Natal Research Innovation and Sequencing Platform
- Africa Health Research Institute
- Centre for the AIDS Programme of Research in South Africa

NIGERIA

- Department of Biochemistry, Lagos State University

AUSTRALIA

- University of Melbourne

FINANCIALS

SANBI's income and expenditure trends for 2021 are shown in this section.

Figure 3: Research income received from all sources

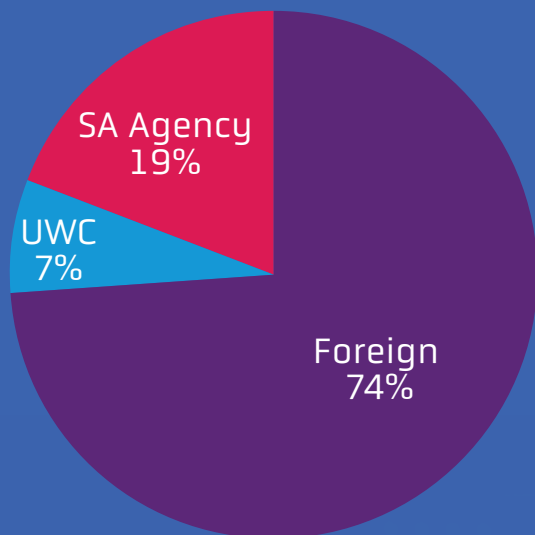


Figure 4: Income received from SA sources

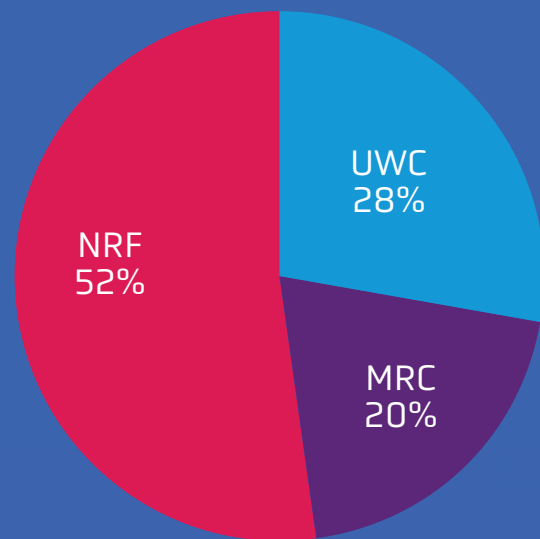


Figure 5: Expenditure in 2021

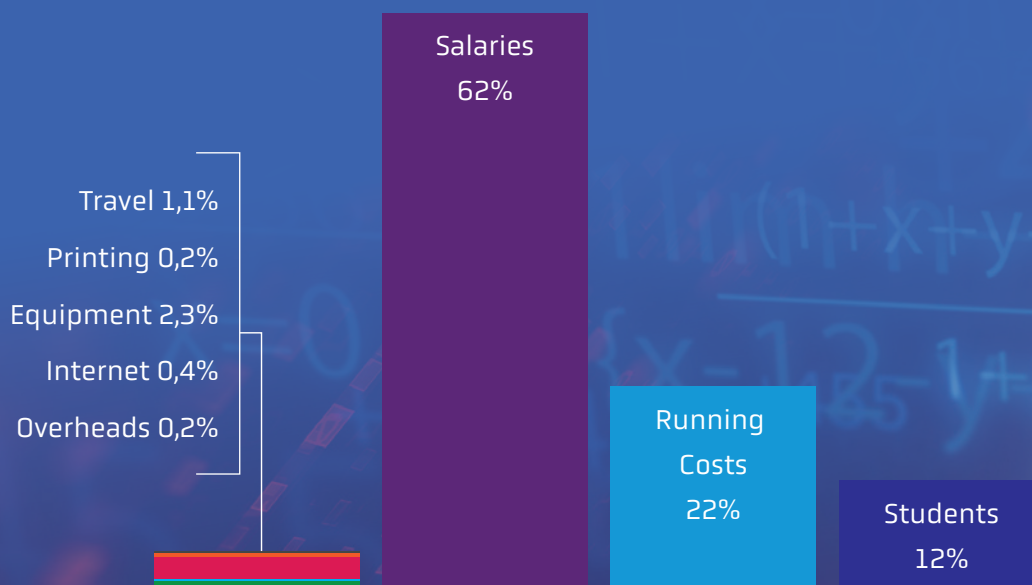
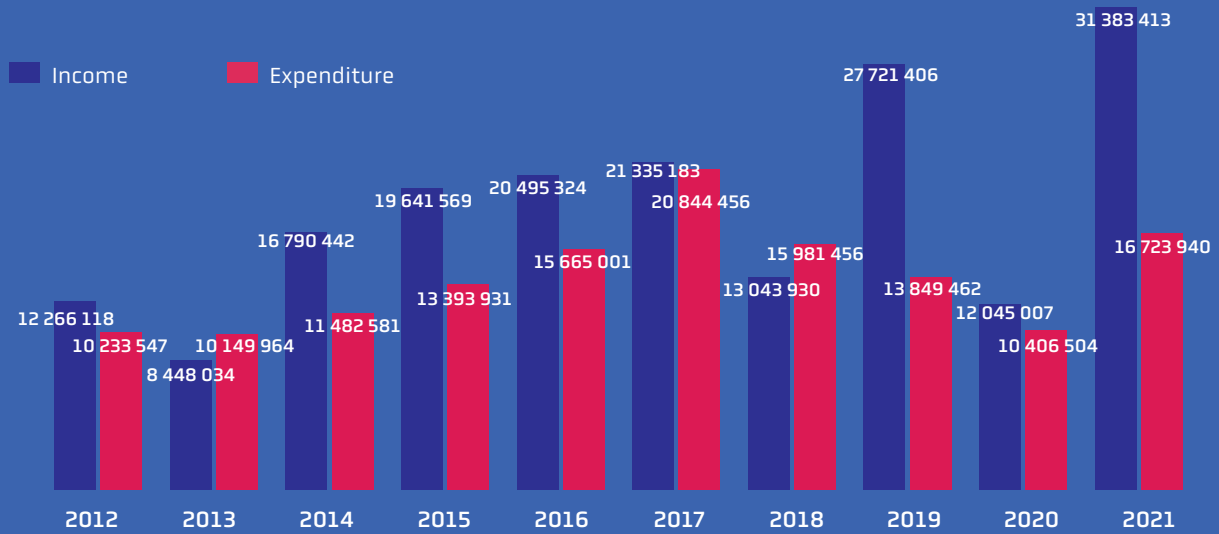


Figure 6: Income versus Expenditure 2012–2021



ALUMNI

Figure 7: SANBI has produced many alumni over the years who have taken their research to institutions and corporates all over the world.

SOUTH AFRICA

- University of the Western Cape
- University of KwaZulu-Natal
- University of Cape Town
- University of Stellenbosch
- University of Pretoria
- University of South Africa
- University of the Free State
- University of the Witwatersrand
- SA Medical Research Council
- National Institute of Communicable Diseases
- Centre for Proteomic and Genomic Research
- Centre for High Performance Computing
- H3ABioNet

INDUSTRY

- Hyrax Biosciences
- IBM
- Roche
- B&M Scientific
- GENE diagnostics
- City of Cape Town
- World Bank
- SA Government Administration
- KappaBiosystems
- Takealot
- Konga.com

- Monash
- Entersekt

KENYA

- Egerton University
- International Livestock Research Institute

GHANA

- University of Ghana

NIGERIA

- Lagos State University
- University of Nairobi

SAUDI ARABIA

- King Abdullah University of Science and Technology

GERMANY

- Institute for Inorganic Chemistry, RWTH Aachen University

UNITED STATES

- Harvard School of Public Health
- Princeton University
- Washington University
- Jackson State University
- University of California San Diego
- Wayne State University

- Pacific Northwest Diabetes Research Institute
- Beth Israel Deaconess Medical Centre, Harvard Medical School
- Johns Hopkins University

CANADA

- University of Western Ontario
- University of British Columbia

IRELAND

- National University of Ireland

SWEDEN

- Karolinska Institute

SWITZERLAND

- Syngenta Crop Protein AG

RUSSIA

- Institute of Cytology and Genetics

SINGAPORE

- Nanyang Technology University
- University of Singapore

MALAYSIA

- Novocraft Technologies Research



FUNDERS



Research Institute at UWC since 1997



National Research Foundation (NRF) funding since 1998



South African Medical Research Council (SAMRC) Bioinformatics Unit funding since 2002



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



H3Africa Bionetwork Node (H3ABioNet Node) 2020–2022



Funding since 2014



Funding since 2016



Funding since 2019



Funding since 2019

Funding since 2021:



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

+27 (0)21 959 3645
info@sanbi.ac.za
(www.sanbi.ac.za)
@SANBI_SA

