



UNIVERSITY of the
WESTERN CAPE

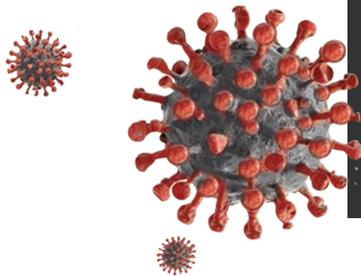


SANBI
South African National
Bioinformatics Institute

ANNUAL REPORT 2020
SOUTH AFRICAN NATIONAL BIOINFORMATICS INSTITUTE

CONTENTS

01	ABOUT SANBI
03	DIRECTOR'S MESSAGE
04	YEAR IN REVIEW
06	STAFF
09	CAPACITY DEVELOPMENT
16	COMPUTATIONAL RESOURCES
19	SANBI'S STAKEHOLDER ENGAGEMENT
21	RESEARCH OUTPUTS
30	RESEARCH PROJECTS THEMES
32	RESEARCH LABORATORIES:
	• DR DOMINIQUE ANDERSON
	• DR HOCINE BENDOU
	• PROF ALAN CHRISTOFFELS
	• DR RUBEN CLOETE
	• DR RAVNIT GREWAL
	• DR ULJANA HESSE
	• DR GORDON HARKINS
50	SANBI COLLABORATIONS
51	FINANCIALS
52	ALUMNI
53	FUNDERS



ABOUT SANBI

WHO WE ARE

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

to accelerate their genomics data analyses. SANBI aims to heighten awareness of bioinformatics in South Africa and to assist the country in making optimal use of 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 online specialised resources for genomics and genome informatics;
- capacity development in genomics and bioinformatics in Africa; and
- the development and implementation of genome annotation methods.

OUR VISION

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

OUR MISSION

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

OUR GOALS

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

POLICY MANDATES

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

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

National Research Foundation Strategy 2020

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

The R&D at SANBI aligns with the Department of Science and Technology'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 (MRC) Act (Act 58 of 1991)

As an extramural unit of the SAMRC, SANBI falls under the legislative mandates of the SAMRC. At Section 3, this Act states that the Legislative Mandate of the SAMRC is:

“**...through research, development and technology transfer, to promote the improvement of the health and quality of life of the population of the Republic, and to perform such functions as may be assigned to the SAMRC by or under this Act...**”.

DIRECTOR'S MESSAGE

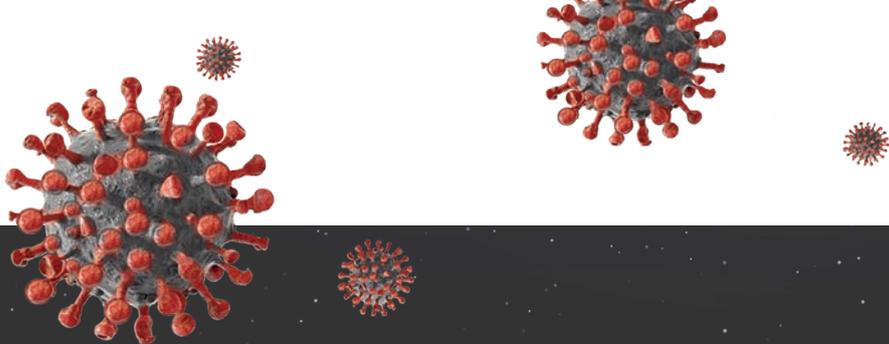


The year 2020 has been marked globally by unprecedented challenges as a result of the COVID-19 pandemic. Many journalists have commented on the view that the “COVID-19 vaccine would not be the panacea we hoped for”. Many of us are aware that we need multiple approaches including non-pharmaceutical interventions in order to stop the spread of SARS-COV-2.

During 2020, SANBI staff and students have persevered in spite of intense mental, physical and psychological strain. Throughout this reporting period we have prioritised our deliverables to address the immediate needs of the COVID-19 pandemic while being true to other research commitments. SANBI was identified as one of three supranational reference laboratories for the World Health Organisation (WHO) on the African continent to respond to COVID-19 and other disease outbreaks. Our annual report for 2020 summarises the projects that our staff and students embarked on in response to the COVID-19 pandemic, and the benefit of global partnerships to accelerate research and development.

As we await the completion of a country-wide vaccination initiative in 2021, we are mindful of many countries whose anticipated COVID-19 vaccination programme is scheduled in 2022 and later. As an institute, we continue to engage in activities that can assist South Africa and beyond as we intensify our research efforts on re-emerging diseases.

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



YEAR IN REVIEW

The highlights for 2020 reflect a completion of existing milestones and new opportunities that have arisen as a result of the prevailing COVID-19 pandemic which provided opportunity to align our infectious disease analytics platform to the needs of COVID-19 analysis, as countries attempted to scale up their laboratory capabilities.

RESEARCH HIGHLIGHTS

SANBI was selected as one of three WHO reference laboratories to support the response to emerging infectious diseases, including COVID-19.

This announcement was preceded by our work to assist the National Institute of Communicable Diseases (NICD) to sequence and analyse the first SARS-COV-2 genome in South Africa.

We secured funding to the order of US\$120,000 from the Africa CDC (Center for Diseases Control). The objective is to build an analytics platform for SARS-COV-2 analysis so that African countries can carry out in-country data analysis. This platform will support the African Union's drive to strengthen regional analytical capability.

In collaboration with researchers from the Institute of Infectious Diseases and Molecular Medicine at UCT and Johns Hopkins University, School of Medicine in Baltimore, USA, we secured funding from the National Institutes of Health, USA (NIH) and the SA MRC to the amount of R13.7million (UWC sub-award R4.5million) over the next five years to characterise the dynamics of the HIV-1 subtype C latent viral reservoir and determinants thereof in African populations.

As a Member of the WHO Strategic Advisory Group of experts on In Vitro Diagnostics (SAGE IVD), Ravnit Grewal reviewed the haematology related diagnostics tests of the “*WHO Technical Report Series 1031. The selection and use of essential in vitro diagnostics.*” This initiative was started to assist lower to middle income countries in setting up diagnostic tools that can guide clinicians on which tests to use in different settings based on their availability in the various countries.

Our regional collaborations continue to expand. Recently we applied our drug discovery platform to Tsetse repellents in collaboration with the International Center for Insect Physiology and Entomology in Nairobi, Kenya. In this work the molecular analysis of waterbuck repellents was described in a publication by our PhD student, Souleymane Diallo in *Frontiers in Cellular Neuroscience*.

Biobanking research remains a cornerstone of development in Africa. Formalin-embedded tumor blocks are a rich source of clinical material that is collected over many years at hospitals. A 6-year study at SANBI examined the integrity of these archival biospecimens spanning a 10-year period. The data added empirical evidence to the integrity of protein signatures in the archival material. This study adds value for many pathology departments that store clinically relevant biospecimens.

Funding from the Technology Innovation Agency (TIA) was used in 2020 to develop a sustainable business model for our open-source biobank LIMS software. This output provided an investor deck that we use to pitch the software to potential clients. We have presented the software at the BioAfrica Convention in August 2020.

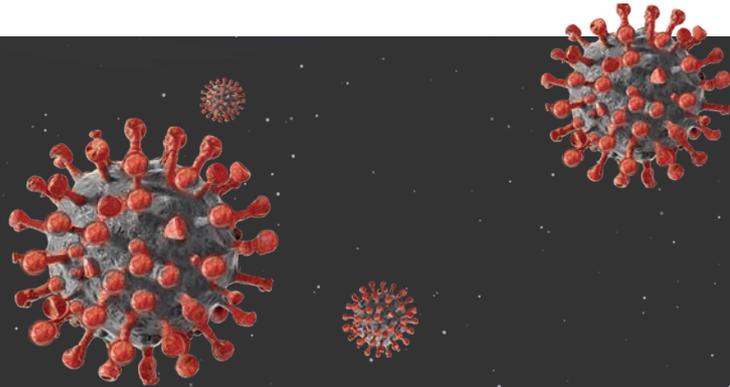
On 22 October 2020, our trademark application for Baobab LIMS (2018/21323-24) was successfully registered.

Our work on cancer genomics resulted in a first author publication by PhD candidate Abdulazeez Giwa in Oncotarget entitled "*Identification of novel prognostic markers of survival time in high-risk neuroblastoma using gene expression profiles*".

Our research on the structural impact of resistance associated mutations in the South African HIV-1C integrase protein yielded four publications, of which two are student first author papers.

PHA4GE

The secretariat for the Public Health Alliance for Genomic Epidemiology (PHA4GE) is hosted at SANBI.



SOFTWARE DEVELOPMENT

The computational tools developed at SANBI, have yielded software code that is used by the international community.

Baobab LIMS highlights:

- nearly 400 views of the code online and 64 unique downloads (github)
- customisation of the LIMS for the Africa CDC to support lab strengthening in African countries
- completion of the LIMS customisation for the Abidjan biobank that was funded through the West African Health Organisation.

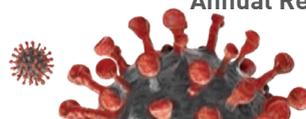
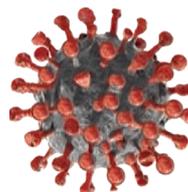
COMBAT-TB highlights:

- tool wrappers created by SANBI are present in the GALAXY Toolshed (<https://github.com/SANBI-SA/tools-sanbi-uwc>).
- the graph database (Neo4j) in COMBAT-TB (Lose et al., 2019) has had 114 views; 77 users of our system distributed across China (14 users), South Africa (13), Japan(10), USA (6), Belgium (5), Italy (5), Germany (3), India (3), Argentina (2) and Canada (2).
- we deposit our opensource code to github for the scientific community to use (<https://github.com/COMBAT-TB>).

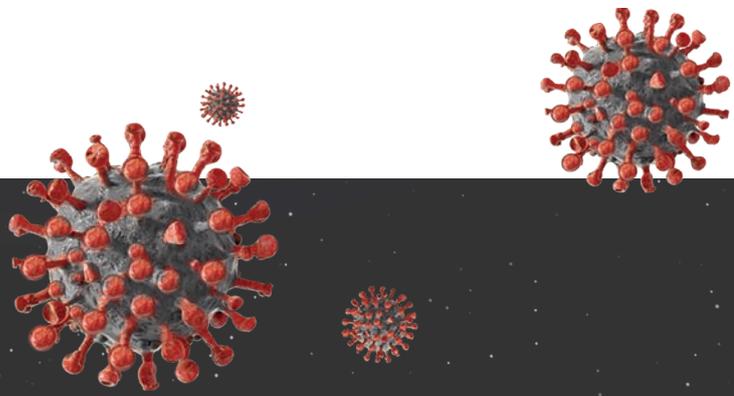
RESEARCH OUTPUTS

SANBI has yielded a large number of publications in the last year:

- 31 journal publications
- 2 book chapters
- 1 conference proceedings
- 1 trademark registered
- 1 journal commentary



STAFF



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

ACADEMIC STAFF are tasked with conducting research in their laboratories, securing funding, student graduations and capacity development.



Dominique Anderson, Dr
Senior Researcher
University of the Western Cape



Hocine Bendou, Dr
Senior Researcher
University of the Western Cape



Alan Christoffels, Prof
Director, SARChI
Bioinformatics,
SA MRC Unit Director
DST/NRF Research Chairs
Programme



Ruben Cloete, Dr
Lecturer
University of the Western Cape



Ravnit Grewal, Dr
Senior Lecturer
SA Medical Research Council



Gordon Harkins, Dr
Senior Lecturer
University of the Western Cape



Judit Kumuthini, Dr
Senior Researcher
H3A BioNetwork



Anja Bedeker
Research Associate
Bill and Melinda Gates Foundation

TECHNICAL STAFF maintain and develop the computing infrastructure.



Zipho Mashologu
Software Developer
 SA Medical Research Council



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



Peter Van Heusden
Senior Systems Developer
 DST/NRF Research Chairs Programme



Shadley Wentzel
Senior systems Administrator
 SA Medical Research Council



Rangarirai Matima
Communication Specialist (PHA4GE)
 Bill and Melinda Gates Foundation



Alecia Naidu
Technical Manager (PHA4GE)
 Bill and Melinda Gates Foundation



Jamie Southgate
Communications Officer (PHA4GE)
 Bill and Melinda Gates Foundation



Lunga Baliwe
Software Developer
 Baobab LIMS



Quinton Coert
Software Developer
 Baobab LIMS

SANBI ADMINISTRATORS ensure the smooth running of daily operations.



Duane Kellerman
Receptionist
 SA Medical Research Council



Fungiwe Mpithi
PA/Administrator (PHA4GE)
 SA Medical Research Council



Ferial Mullins
Finance Administrator (PHA4GE)
 UWC



Nawaal Nacerodien-Weitz
Administrator (PHA4GE)
 BMGF



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

FURTHER DEVELOPMENT OF STAFF

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

Attended by	Venue and Date	Course Name and Benefit
Dominique Anderson	UWC, March 2020	POPIA workshop: This half-day interactive workshop, presented by Era Gunning of the law firm ENSafrica, offered an introduction to POPI, guidelines for compliance, and an opportunity to discuss the challenges that POPI poses for processing different kinds of research data and information sources (including archives, special collections, data repositories, and other institutional records).
Anja Bedeker	The Global Health Network, September 2020	Research in global health emergencies: ethical issues (eLearning course)
	The Steve Biko Centre for Bioethics, University of the Witwatersrand, September 2020	The annual Steve Biko bioethics lecture: Towards a post-COVID-19 future with a more human face.
	Africa Centres for Disease Control and Prevention, October 2020	Virtual Launch of the Africa Pathogen Genomics Initiative
	University of the Witwatersrand, October – November 2020	Wits Pandemic Pangolin: Systems, Science and Society (webinar series) 1. The human face of vaccine trials: Biopolitics, trust, science and communication 2. Pandemic Lessons from History 3. Pandemic Innovations 4. The Endangered Pangolin: The True Story 5. Pandemics and Society – community engagements 6. The Human Resource in a time of pandemic: Vulnerabilities and Strengths
	The Global Health Network	Ethics in epidemics, emergencies and disasters: Research, surveillance and patient care (eLearning course).
	Academy of Science of South Africa	Academy of Science of South Africa Open Access Week (webinar series) 1. The POPI Act Forum 2. The Open Science Agenda & Framework for South Africa 3. The African Scientists Directory
	World Health Organisation (WHO)	Inequities in COVID-19 infection and mortality: socioeconomic risk factors and populations at risk (webinar)
	Dullah Omar Institute, UWC	Webinar on Privacy, Digital Technology, and COVID-19

CAPACITY DEVELOPMENT

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

UNDERGRADUATE TRAINING PROGRAMME

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

Bioinformatics Module (BTN 315)

Each year the UWC undergraduate Bioinformatics Module is taught to approximately 85 third-year Biotechnology students. During April and May 2020 the SANBI faculty, Ruben Cloete, Gordon Harkins, Hocine Bendou and Alan Christoffels presented various Bioinformatics topics.

Internship Programme

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

Honours Programme

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

Masters Programme

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

Doctoral Programme

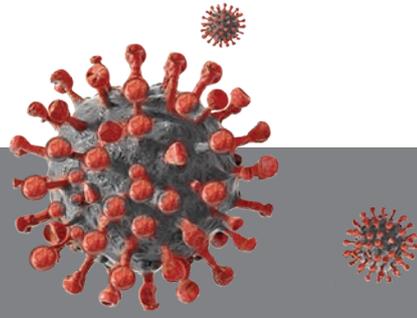
Candidates should be in possession of a Masters degree in Bioinformatics or in a related scientific field subject area such as Computer Science, Mathematics, Biochemistry and Engineering. The PhD degree must be completed within five years.

Postdoctoral Programme

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

SANBI Postgraduate Registration Trends 2001-2020.





SANBI STUDENTS 2020

The SANBI student complement comprises a cohort of diverse and multi-talented researchers.

POSTDOCTORAL FELLOWS



Ibrahim Ahmed
Alan Christoffels*



**Oladapo Micheal
Olagbegi**
Alan Christoffels*



Mohd Shahbaaz
Alan Christoffels*



Wesley Williams
Alan Christoffels*



Sunday Vodah
Alan Christoffels*

DOCTORAL STUDENTS



**Abiola Abidemi
Babjide**
Dominique Anderson*



**Tracey
Calvert-Joshua**
Alan Christoffels*



Sarah DeRaedt
Alan Christoffels*



Souleymane Diallo
Alan Christoffels*



**Mmakamohelo
Direko**
Alan Christoffels*

* supervisor

DOCTORAL STUDENTS



Nasr Eshibona
Hocine Bendou*



Fanechka Esterhuysen
Uljana Hesse*



Abdulazeez Giwa
Hocine Bendou*



Bridget Langa
Junaid Gamiieldien*



Yamkela Mgwatyu
Uljana Hesse*



Anati Nkaule
Alan Christoffels*



Nikkita Pillay
Alan Christoffels*



Catherine Rossouw
Alan Christoffels*



MASTERS STUDENTS



Lynley Abdoll
Alan Christoffels*



Peter Abiodun
Alan Christoffels* and
Dominique Anderson*



Cleo-Rose Barendse
Hocine Bendou*



Rumbidzai Chitongo
Ruben Cloete*



Farzaana Diedericks
Alan Christoffels*

* supervisor

MASTERS STUDENTS



Susan Alecia Fernol
Alan Christoffels*



Ridaa Fredericks
Alan Christoffels*



Maryam Hassan
Ruben Cloete*



Darren Isaacs
Ruben Cloete*



Wardah Jassiem
Hocine Bendou*



Siposetu Mazitshana
Gordon Harkins*



Nomlindelo Mfuphi
Hocine Bendou*



Christianah Omotoso
Samuel Egieyeh*



Samson Oselusi
Samuel Egieyeh*
(School of Pharmacy)



Jamie Southgate
Alan Christoffels*

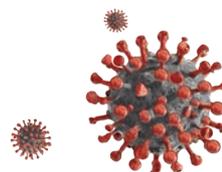
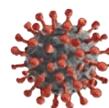
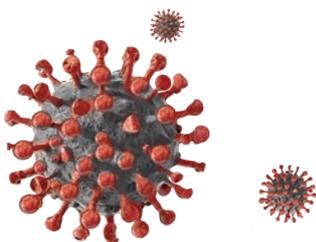


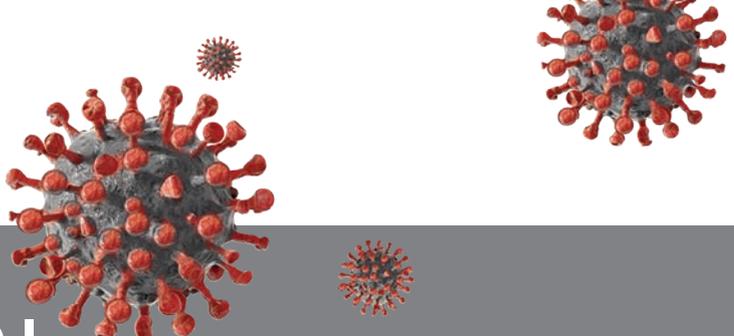
Rudolph Serage
Dominique Anderson*



Peter van Heusden
Alan Christoffels*

* supervisor

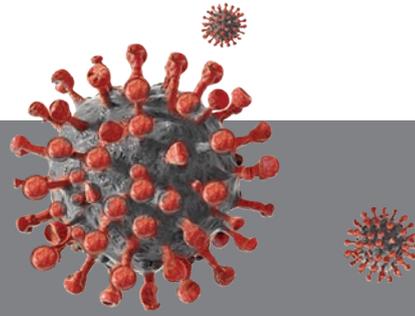




2020 SANBI THESIS COMPLETIONS

Student Name	Degree	Supervisor	Thesis Title
Catherine Rossouw	PhD	Alan Christoffels	Optimisation of proteomics techniques for archival tumour blocks of a South African cohort of colorectal cancer
Souleymane Diallo	PhD	Alan Christoffels	Coding of tsetse repellents by olfactory sensory neurons: towards the improvement and development of novel tsetse repellents
Jamie Southgate	MSc	Alan Christoffels	Establishing a Framework for an African Genome Archive
Peter Abiodun	MSc	Alan Christoffels	Exploring the influence of Organisational, Environmental and Technological factors on information securities policies and compliance on biomedical data at South African higher institutions
Rumbidzai Chitongo	MSc	Ruben Cloete	Investigating the structural effect of known Raltegravir resistance associated mutations on the South African HIV-1 integrase subtype C protein

WORKSHOPS AND COURSES



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

Bioinformatics Training Course

This annual Western Cape course ran from 17 February to 18 March 2020. The first week aimed to get students from different backgrounds to a similar level to start the modules such that biologists could learn basic programming and programmers could learn the fundamentals of molecular biology. The rest of the course was aimed at providing an overview of a range of bioinformatics topics so that students are exposed to this broad field.

There were 25 participants in attendance from different universities and scientific backgrounds.

Lectures were presented by Peter van Heusden, Ibrahim Ahmed, Dominique Anderson, Ruben Cloete, Gordon Harkins, Hocine Bendou (all from SANBI) and Nicki Tiffin and Verena Ras (from UCT).

Participants of the 2020 Bioinformatics Course held at SANBI. (prior to COVID-19).



WORKSHOPS ORGANISED BY SANBI

Presenter	Date	Venue	Course Name and Purpose
Dominique Anderson	January 2020	Online	Baobab LIMS demo sessions for NIH and CDC.
Dominique Anderson	March 2020	Online	Baobab LIMS demo sessions for the Provincial Veterinary Department, Stellenbosch University.
Dominique Anderson	November 2020	Online	Baobab LIMS demo sessions for Biobanks across Sweden.
Dominique Anderson	December 2020	Online	Baobab LIMS demo sessions for Hitech Health.

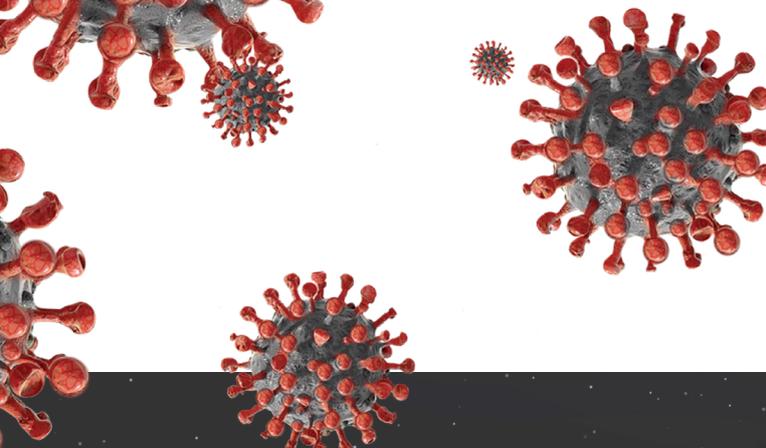
EXTERNAL LECTURES BY SANBI STAFF

Presenter	Date	Venue	Course Name and Purpose
Dominique Anderson	February 2020	Haematology Department, Tygerberg Hospital	Molecular biology course for Pathology. Introduction to proteomics and bioinformatics. Delegates are clinical researchers.
Dominique Anderson	February 2020	IMBM, UWC	Protein expression workshop for students and staff at IMBM.
Dominique Anderson	April 2020	Online	Bioinformatics for Clinicians. Delegates were from the Faculty of Dentistry, UWC.
Ruben Cloete	April 2020	Division of Medical Virology, Stellenbosch University. Virtual online teaching	Honours course Protein analysis (basic science) lecture and Protein analysis (structural biology) lecture.
Alan Christoffels	May 2020	Online	COVID-19 Diagnostics and Testing online teaching material for African Society of Laboratory Medicine (ASLM). Also offered under Future Learn short courses. https://www.futurelearn.com/courses/COVID-19-diagnostics-and-testing

PATHOGEN GENOMICS GRANT WRITING WORKSHOP

This workshop was held in Cape Town 2-7 March 2020 where experts from the Africa CDC and African Society for Laboratory Medicine gathered with SANBI staff to formulate a strategy for pathogen analysis and handling of potential future pandemics.





COMPUTATIONAL RESOURCES

The IT team at SANBI supports the work of the institute through software development and management of our inhouse research cloud, data storage and High Performance Computing (HPC) facilities.

SANBI IT and Bioinformatics Software Development

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



COMBAT-TB NeoDB and Workbench

The Computational Bacterial Analytical ToolKit for Tuberculosis (<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 knowledgebase built on top of a Graph database of genomic, protein, drug interaction 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 (lead author), who left SANBI at the start of 2020, 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 IRIDA, 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. A paper on this work will be submitted in 2021.



Baobab LIMS

Baobab LIMS (www.baobablms.org) is an open-source laboratory information management system developed by SANBI. Key to continuous development and improvement of Baobab LIMS includes engagement with potential end-users to gain greater insight into the overall heterogeneity of operations in different laboratory and biobank environments. This provides the development team with an understanding of the unique LIMS functionality gaps to promote 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 for sample quality control for a variety of biological collections. New modules for sample preparation include sample aliquoting, centrifugation and sample pooling.



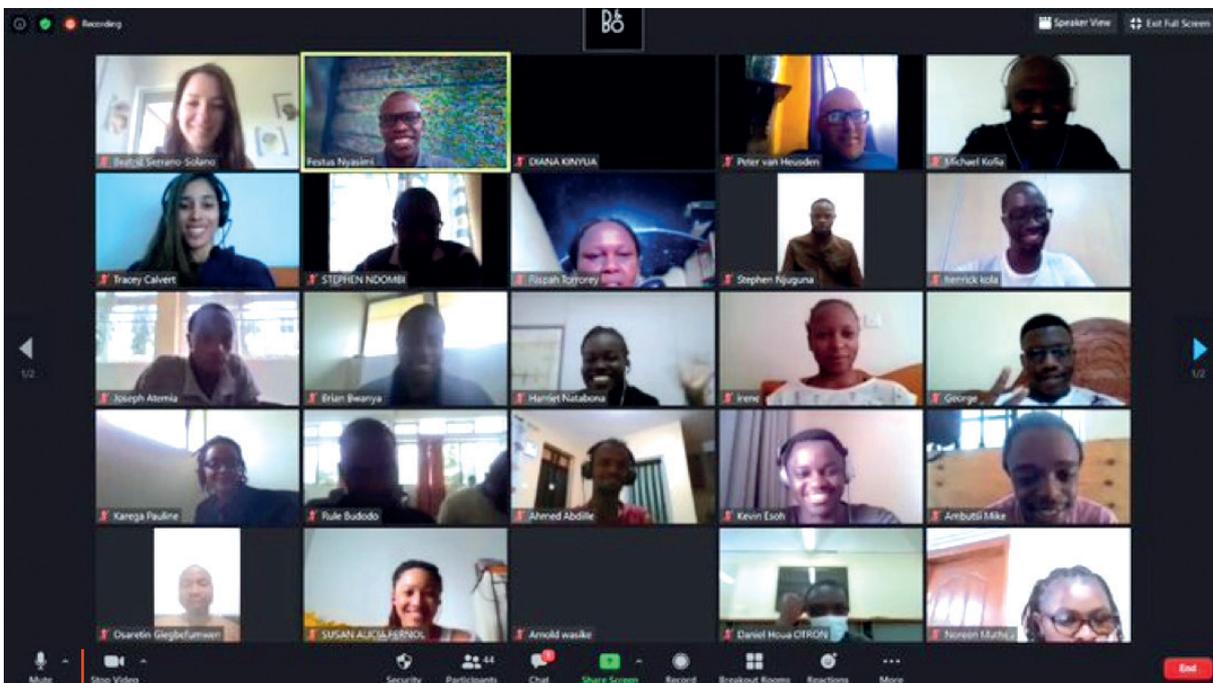
Additional meta-data aware capability aimed at improving interoperability and harmonisation has been added. The most recent common structured vocabulary added to Baobab LIMS is based on the SARS-COV-2 metadata standard, developed by a working group within the PHA4GE consortium (<http://www.pha4ge.org>). 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 QC capability has 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, is 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 did not affect existing functionality has been incorporated into the code base. This ensures that if any development error occurred, they are identified early on. To improve Baobab LIMS infrastructure deployment, repositories containing Terraform scripts and an Ansible playbook for 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 AWS.

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



Our work on Galaxy and M. tuberculosis sequence analysis was presented at a Bioinformatics Hub of Kenya online event in July 2020. The Bioinformatics Hub of Kenya is a forum created by Kenyan students and early career researchers in the field of bioinformatics.



Galaxy

SANBI remains an active part of the Galaxy Community. Peter van Heusden was the Scientific Programme Chair of the Galaxy Stream in the 2020 Bioinformatics Community Conference, which, like so many events, took place entirely online. He also helped coordinate a Galaxy Africa meetup event during that conference and has worked with Ziphozakhe Mashologu on numerous enhancements and bug fixes in the Galaxy software and Galaxy tools.

Outreach and Mentorship

SANBI provided assistance to the National Institute for Communicable Diseases (NICD) in their sequencing of the first SARS-CoV-2 genome sequenced in South Africa and has provided advice and assistance to groups in Nigeria and Zambia in their similar efforts.

We supported two teams of UWC students that took place in the 2020 Student Cluster Competition (SCC). Unfortunately due to increased workload, neither team could complete the final round of the SCC.

IT and Systems Support

We welcomed Shadley Wentzel to fill the role of SANBI systems support engineer. With some sadness we bade farewell to Eugene de Beste, who left SANBI to join the Center for High Performance Computing (CHPC)'s Advanced Computer Engineering (ACE) lab.

We continue to provide a HPC cluster and OpenStack compute 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 Communication 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.

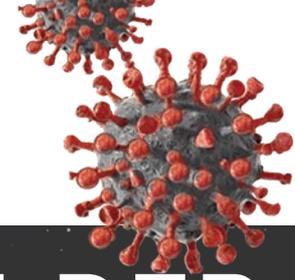
Server Type	Server Count	RAM	CPU cores
M710HD	6	64 GB - 128 GB	72
M910HD	1	512 GB	16
M620	6	96 GB	72
M610X	2	32 GB	24
M630	3	96 GB - 128 GB	48

Virtual Machine Infrastructure

Our OpenStack cloud runs on Supermicro servers providing 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 5 Supermicro storage servers.



SANBI'S STAKEHOLDER (PUBLIC) ENGAGEMENT

Journal Commentary

CAREFUL GOVERNANCE OF AFRICAN BIOBANKS

Christoffels A, Abayomi A, January 2020

"In a World Report, 54gene (a start-up genomics company) was featured as the first pan-African biobank that plans to collect 40 000 biospecimens from ten hospitals in Nigeria by the end of 2019. The World Report has subsequently been reproduced in the media. In a world where media reports are dominated by fake news, clarification of African biobank initiatives is imperative. These initiatives have been active for years and have delivered tangible interventions that affect Africans who donate biospecimens for research and empower the researchers who are resident in Africa..."

The Lancet (0140-6736) [https://doi.org/10.1016/S0140-6736\(19\)32624-8](https://doi.org/10.1016/S0140-6736(19)32624-8)

Online Publication

SA SCIENTISTS MAKE VITAL COVID-19 DISCOVERY

By Kelly Turner / April 6, 2020

"CAPE TOWN- Scientists around the world have been tirelessly working to find a cure for COVID-19. A group of South African scientists and researchers have contributed to the cause and have made a vital discovery..."

<https://www.coronavirusmonitor.co.za/treatment/sa-scientists-make-vital-COVID-19-discovery/>

Radio Discussion

RADIO 702 PROGRAMME BREAKFAST WITH BONGANI BINGWA

7 April 2020

Peter Van Heusden chats to Radio702 about the team's successful genome sequencing of the first local sample of SARS-Cov-2, the virus which causes COVID-19.

<https://omny.fm/shows/the-breakfast-show-702/sa-s-first-sars-cov-2-genome-sequencing-to-find-out>

TV Discussion

Alan Christoffels appeared on tv news on Newzroom Afrika on 19 May 2020. He emphasised that an alignment of academics with policymakers and implementers in government is crucial to be able to track the spread of COVID-19.

<https://www.youtube.com/watch?v=CdKTEc5RZnl>



Online TV Show

WIM TV

On 17 November, Peter Van Heusden joined a panel with Profs Lynn Morris (NICD) and Glenda Grey (MRC), the late journalist Karima Brown and WHO Epidemiologist Dr Otim Patrick Ramadan. Hosted by the online show WIM TV, the discussion was helpful in spreading a positive message about vaccines during the COVID-19 global pandemic.

<https://www.youtube.com/watch?v=uR5TNesuQg4&feature=youtu.be>



Public Science Engagement

THE CONVERSATION

Africa joins the race to trace COVID-19 with genomics

by Peter Van Heusden, 19 April 2020

“One important aspect of managing emerging infections is identifying chains of transmission and assigned cases to clusters of infection. A case in point is South African trade union leader Zwelinzima Vavi, who spent a few days in hospital with the new coronavirus disease. He stated that he “had no idea” where he may have contracted the virus and was scrutinising his travel history for clues...”

NRF SPECIAL EDITION OF “SCIENCE MATTERS”

SA’s first SARS-COV-2 genome sequence and analysis - REFLECTIONS

by Prof Alan Christoffels, May 2020

“As citizens we applaud the speed with which the South African Government has responded to the COVID-19 pandemic. Our national response teams are ramping up the much needed national control measures in response to COVID-19, including health screening, COVID-19 testing and associated contact tracing and provision for self-isolation of vulnerable communities...”

<https://www.nrf.ac.za/sites/default/files/documents/03%20NRF%20SMM%20SPECIAL%20ISSUE%20Reflections%20SA%E2%80%99s%20first%20SARS-COV-2%20genome%20sequencing.pdf>

Print Media

DAILY MAVERICK CITIZEN OP-ED

Glimmer of hope: Africa CDC launches guidelines to make COVID-19 tests more affordable and accessible

Peter van Heusden, 18 December 2020

“On 14 December 2020 the Western Cape health department restricted access to COVID-19 testing for the second time in the year, excluding most of those under 45 and all of those without symptoms from access to public-sector testing (see here)....”

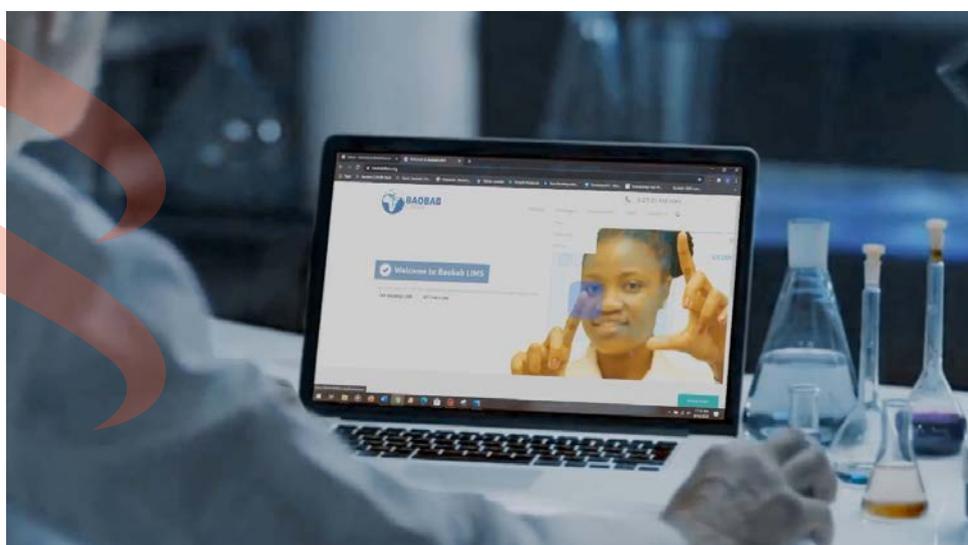
<https://www.dailymaverick.co.za/article/2020-12-18-glimmer-of-hope-africa-cdc-launches-guidelines-to-make-COVID-19-tests-more-affordable-and-accessible/>

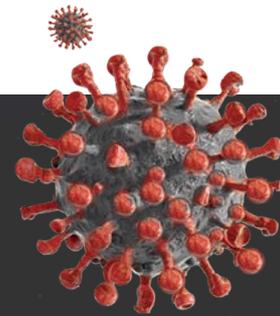
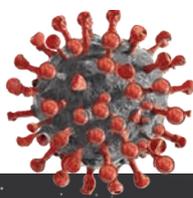
Industry Engagement

BIO AFRICA 2020

A Baobab LIMS promotional video was launched at the BIO Africa 2020 Digital Convention held 24-26 August.

https://baobablms.org/wp-content/uploads/2020/12/BaobabLIMS_BIOAfrica_2020_Video-profile.mp4





RESEARCH OUTPUTS

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

JOURNAL PUBLICATIONS

#	Publication (SANBI contributors)	Published Date	Impact Factor
1	Symptom evolution following the emergence of maize streak virus Monjane Aderito, Dellicour Simon, Hartnady Penelope, Oyeniran Kehinde, Owor Betty, Bezeidenhout Marion, Linderme Daphne, Syed Rizwan, Donaldson, Lara, Murray, Shane, Rybicki, Edward, Kvarnheden, Anders, Yazdkhasti Elhman, Lefeuve Pierre, Froissart, Rémy, Roumagnac, Philippe, Shepherd, Dionne, Harkins Gordon , Suchard Marc, Martin Darren. <i>eLife Sciences</i> 9 DOI: 10.7554/eLife.51984	January 2020	7.080
2	Exploring new genetic variants within COL5A1 intron 4-exon 5 region and TGF-6 family with risk of anterior cruciate ligament ruptures. Mary-Jessica N. Laguette, Kelly Barrow, Firzana Firfirey, Senanile Dlamini, Colleen J. Saunders , Collet Dandara, Junaid Gamielidien , Malcolm Collins and Alison V. September. <i>Journal of Orthopaedic Research</i> 2020; 38: 1856-1865. https://doi.org/10.1002/jor.24585	10 January 2020	3.14
3	Differential Diagnosis of Malignant Lymphadenopathy Using Flow Cytometry on Fine Needle Aspirate: Report on 269 Cases Carla Griesel, Minodora Desmirean, Tonya Esterhuizen, Sergiu Pasca, Bobe Petrushev, Cristina Selicean, Andrei Roman, Bogdan Fetica, Patric Teodorescu, Carmen Swanepoel, Ciprian Tomuleasa and Ravnit Grewal . <i>Journal of Clinical Medicine</i> 2020, 9, 283 doi:10.3390/jcm9010283	20 January 2020	5.583
4	Protection of Personal Information Act 2013 and data protection for health research in South Africa Ciara Staunton, Rachel Adams, Dominique Anderson , Talishiea Croxton, Dorcas Kamuya, Marianne Munene, Carmen Swanepoel <i>International Data Privacy Law</i> , ipz024, https://doi.org/10.1093/idpl/ipz024	24 January 2020	4.162
5	COMBAT-TB-NeoDB: fostering tuberculosis research through integrative analysis using graph database technologies Thoba Lose, Peter van Heusden and Alan Christoffels <i>Bioinformatics</i> , 36(3), 2020, 982–983 doi: 10.1093/bioinformatics/btz658	01 February 2020	5.610
6	Rooibos (<i>Aspalathus linearis</i>) Genome Size Estimation Using Flow Cytometry and K-Mer Analyses Yamkela Mgwaty, Allison Anne Stander , Stephan Ferreira, Wesley Williams and Uljana Hesse. <i>Plants</i> 2020, 9, 270 doi:10.3390/plants9020270	18 February 2020	2.632

7	A population model for the 2017/18 listeriosis outbreak in South Africa. Witbooi PJ, Africa C, Christoffels A , Ahmed IHI <i>PLoS ONE</i> 15(3): e0229901. https://doi.org/10.1371/journal.pone.0229901	12 March 2020	2.740
8	Drug Resistance Mutations Against Protease, Reverse Transcriptase and Integrase Inhibitors in People Living With HIV-1 Receiving Boosted Protease Inhibitors in South Africa. Obasa AE, Mikasi SG, Brado D, Cloete R , Singh K, Neogi U and Jacobs GB (2020) <i>Frontiers in Microbiology</i> . 11:438 doi: 10.3389/fmicb.2020.00438	20 March 2020	4.076
9	HIV-1 Integrase Diversity and Resistance-Associated Mutations and Polymorphisms Among Integrase Strand Transfer Inhibitor-Naive HIV-1 Patients from Cameroon Sello Given Mikasi, Josiah Otwoma Gichana, Cheri Van der Walt, Dominik Brado, Adetayo Emmanuel Obasa, Duncan Njenda, Martha Messembe, Emilia Lyonga, Okomo Assoumou, Ruben Cloete , George Mondinde Ikomey, and Graeme Brendon Jacobs. <i>AIDS Research and Human Retroviruses</i> , Vol. 36, No. 5 https://www.liebertpub.com/doi/10.1089/AID.2019.0264 DOI: 10.1089/aid.2019.0264	30 April 2020	1.805
10	Molecular dynamic simulations to investigate the structural impact of known drug resistance mutations on HIV-1C Integrase-Dolutegravir binding. Rumbidzai Chitongo, Adetayo Emmanuel Obasa, Sello Given Mikasi, Graeme Brendon Jacobs, Ruben Cloete <i>PLoS ONE</i> 15(5):e0223464. https://doi.org/10.1371/journal.pone.0223464	7 May 2020	2.740
11	Ellagic Acid Controls Cell Proliferation and Induces Apoptosis in Breast Cancer Cells via Inhibition of Cyclin-Dependent Kinase 6. Mohd Yousuf, Anas Shamsi, Parvez Khan, Mohd Shahbaaz , Mohamed F. AlAjmi, Afzal Hussain, Gulam Mustafa Hassan, Asimul Islam, Qazi Mohd Rizwanul Haque and Md. Imtaiyaz Hassan. <i>International Journal of Molecular Sciences</i> 2020, 21, 3526; doi:10.3390/ijms21103526	15 May 2020	4.556
12	Tomorrow's privacy: information security at South African universities—implications for biomedical research Dominique Anderson, Oluwafemi Peter Abiodun, Alan Christoffels <i>International Data Privacy Law</i> ipaa007 https://doi.org/10.1093/idpl/ipaa007	27 May 2020	4,162
13	Functional and Structural Analysis of Predicted Proteins Obtained from Homo sapiens' Minisatellite 33.15-Tagged Transcript pAKT-45 Variants Mohd Shahbaaz , Awad Saeed Al-Samghan, Arshi Malik, Sarah Afaq, Afaf S. Alwabli, Irfan Ahmad, Mostafa A. Hussien, Mohamad Zubair, Farha Fatima, Shamina Begum, Othman Alzahrani, and Mohammed Tarique. <i>BioMed Research International</i> Volume 2020 https://doi.org/10.1155/2020/2562950	27 May 2020	2.583
14	Cellular and Molecular targets of Waterbuck Repellent Blend (WRB) odours in antennae of Glossina fuscipes fuscipes, Newstead, 1910 Souleymane Diallo, Mohd Shahbaaz , Baldwin Torto, Alan Gilbert Christoffels , Daniel Masiga, Merid Negash Getahun. <i>Frontiers in Cellular Neuroscience</i> https://doi.org/10.3389/fncel.2020.00137	03 June 2020	4,555
15	Identification of a novel WAS mutation in a South African patient presenting with atypical Wiskott-Aldrich syndrome: a case report. Brigitte Glanzmann, Marlo Möller, Mardelle Schoeman, Michael Urban, Paul D. van Helden, Lisa Frigati, Ravnit Grewal , Hermanus Pieters, Ben Loos, Eileen G. Hoal, Richard H. Glashoff, Helena Cornelissen, Helena Rabie, Monika M. Esser & Craig J. Kinnear <i>BMC Medical Genetics</i> 21, 124 (2020). https://doi.org/10.1186/s12881-020-01054-6	05 June 2020	1.585

16	HIV-1 Drug Resistance Mutation Analyses of Cameroon-Derived Integrase Sequences. Mikasi Sello, Isaacs Darren , Ikomey George, Henerico Shimba, Cloete Ruben , Jacobs Graeme. <i>AIDS Research and Human Retroviruses</i> . 10.1089/AID.2020.0022.	July 2020	1.805
17	Genome sequencing of a severe acute respiratory syndrome coronavirus 2 isolate obtained from a South African patient with coronavirus disease 2019. Allam M, Ismail A, Khumalo ZTH, Kwenda S, van Heusden P , Cloete R , Wibmer CK, Mtshali PS, Mnyameni F, Mohale T, Subramoney K, Walaza S, Ngubane W, Govender N, Motaze NV, Bhiman JN, on behalf of the SA-COVID-19 response team. <i>Microbiology Resource Announcements</i> 9:e00572-20. https://doi.org/10.1128/MRA.00572-20 .	02 July 2020	0.785
18	Unravelling the unfolding pathway of human Fas-activated serine/threonine kinase induced by urea Khalid A. Alamry, Saurabha Srivastava, Mohd Shahbaaz , Parvez Khan, Preeti Gupta, Sunayana Begum Syed, Naved Azum, Abdullah M. Asiri, Asimul Islam, Faizan Ahmad and Md. Imtaiyaz Hassan <i>Journal of Biomolecular Structure and Dynamics</i> https://doi.org/10.1080/07391102.2020.1790423	14 July 2020	4.986
19	Classification and functional analyses of putative virulence factors of <i>Mycobacterium tuberculosis</i>: A combined sequence and structure based study Mohd Shahbaaz , Vladimir Potemkin, Krishna Bisett, Imtaiyaz Hassan, Mostafa A. Hussien. <i>Computational Biology and Chemistry</i> Volume 87, 107270	August 2020	n/a
20	Multistage antiplasmodial activity of hydroxyethylamine compounds, in vitro and in vivo evaluations Neha Sharma, Yash Gupta, Meenakshi Bansal, Snigdha Singh, Prateek Pathak, Mohd Shahbaaz , Raman Mathur, Jyoti Singh, Mohammad Kashif, Maria Grishina, Vladimir Potemkin, Vinoth Rajendran, Poonam, Prakasha Kempaiah, Agam Prasad Singh and Brijesh Rathi <i>Royal Society of Chemistry Advances</i> 2020, volume 10, page 35516-35530 DOI: 10.1039/d0ra03997g	11 August 2020	3.119
21	Structural Comparison of Diverse HIV-1 Subtypes using Molecular Modelling and Docking Analyses of Integrase Inhibitors. Isaacs Darren , Mikasi Sello, Obasa Adetayo, Ikomey George, Shityakov Sergey, Cloete Ruben , Jacobs Graeme. <i>Viruses</i> 2020, 12(9), 936 https://doi.org/10.3390/v12090936	26 August 2020	3.816
22	Diverse genomoviruses representing twenty-nine species identified associated with plants. Fontenele Rafaela, Roumagnac Philippe Richet, Cécile, Kraberger Simona, Stainton Daisy, Aleamotu'a Maketalena, Filloux Denis, Bernardo Pauline, Harkins Gordon , et al. <i>Archives of Virology</i> 1-11 DOI: 10.1007/s00705-020-04801-5	06 September 2020	2.261
23	HIV-1 diversity and the implementation of integrase strand-transfer inhibitors as part of combination antiretroviral therapy. Mikasi Sello, Ikomey George, Obasa Adetayo, Cloete Ruben , Jacobs Graeme. <i>South African Medical Journal</i> 2020;110(9):827. https://doi.org/10.7196/SAMJ.2020.v110i9.14848	September 2020	1.500
24	Transcriptomics of the Rooibos (<i>Aspalathus linearis</i>) Species Complex Emily Amor Stander , Wesley Williams, Yamkela Mgwatyu , Peter van Heusden , Fanie Rautenbach, Jeanine Marnewick, Marilize Le Roes-Hill and Uljana Hesse . <i>BioTech</i> 2020, volume 9, issue 4 doi:10.3390/biotech9040019	23 September 2020	3.912

25	Multistage antiplasmodial activity of hydroxyethylamine compounds, in vitro and in vivo evaluations Sharma N, Gupta Y, Bansal M, Singh S, Pathak P, Shahbaaz M, Mathur R, Singh J, Kashif M, Grishina M, Potemkin V, Rajendran V, Poonam Kempaiah, P, Singh AP, Rathi B. <i>RSC Advances</i> , Volume 10, Issue 58, Pages 35516–35530	25 September 2020	3.070
26	Capacity building for whole genome sequencing of <i>Mycobacterium tuberculosis</i> and bioinformatics in high TB burden countries Emmanuel Rivière, Tim H. Heupink, Nabila Ismail, Anzaan Dippenaar, Charlene Clarke, Gemeda Abebe, Peter van Heusden , Rob Warren, Conor J. Meehan and Annelies Van Rie. <i>Briefings in Bioinformatics</i> , 00(00), 2020, 1–10 doi: 10.1093/bib/bbaa246	03 October 2020	8.990
27	Inhibiting CDK6 activity by Quercetin is an attractive strategy for cancer therapy. Yousuf M, Khan P, Shamsi A, Shahbaaz M , Hasan GM, Haque QMR, Christoffels A , Islam A, Hassan MI <i>ACS Omega</i> , 5, 42 (27480–27491)	27 October 2020	2.870
28	Identification of novel prognostic markers of survival time in high-risk neuroblastoma using gene expression profiles Giwa A, Fatai A, Gamielien J, Christoffels A, Bendou H. <i>Oncotarget</i> Volume 11, Issue 46, Pages 4293–4305	November 2020	3.765
29	Evaluation of Binding of Rosmarinic Acid with Human Transferrin and Its Impact on the Protein Structure: Targeting Polyphenolic Acid-Induced Protection of Neurodegenerative Disorders Shamsi A, Anwar S, Shahbaaz M , Mohammad T, Alajmi MF, Hussain A, Hassan I, Ahmad F, Islam A, Gil G. <i>Oxidative Medicine and Cellular Longevity</i> Volume 2020, 2020, Article number 1245875	05 November 2020	5.076
30	Electronic properties investigation of human dihydrofolate reductase complexes with ligands, Vladislav Naumovich, Maria Grishina, Jurica Novak, Prateek Pathak, Vladimir Potemkin, Mohd Shahbaaz & Magda H. Abdellattif <i>Journal of Biomolecular Structure and Dynamics</i> DOI: 10.1080/07391102.2020.1861985	05 December 2020	4.986
31	Structural analyses and classification of novel isoniazid resistance coupled mutational landscapes in <i>Mycobacterium tuberculosis</i>: a combined molecular docking and MD simulation study Mohd Shahbaaz , Sameer H. Qari, Magda H. Abdellattif & Mostafa A. Hussien <i>Journal of Biomolecular Structure and Dynamics</i> 21;1–10 doi: 10.1080/07391102.2020.1861986	21 December 2020	4.986

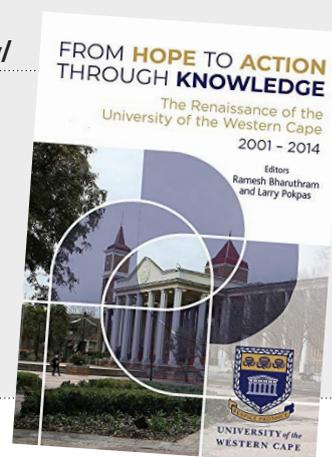
CONFERENCE PROCEEDINGS

- 5th African Conference on Emerging Infectious Diseases and Biosecurity, Abuja, Nigeria, 7 – 9 August 2019.
Volume Editor: **Alan Christoffels**
Proceedings 2020, 45, 2



CHAPTERS IN BOOKS

No.	Chapter Title in Book
1	<p>Immune Correlates of Disease Progression in Linked HIV-1 Infection. Tuen Michael, Bimela Jude, Banin Andrew, Ding Shilei, Harkins Gordon, Weiss Svenja, Itri Vincenza, Durham Allison, Porcella Stephen, Soni Sonal, Mayr Luzia, Meli Josephine, Torimiro Judith, Tongo Passo Marcel, Wang Xiaohong, Kong Xiang-Peng, Nadas Arthur, Kaufmann Daniel, Brumme Zabrina, Dürr Ralf. In: Prime Archives in Immunology Ajmal Khan, editor ISBN: 978-81-945175-9-7 Hyderabad, India: Vide Leaf. February 2020 https://videleaf.com/product/prime-archives-in-immunology/</p>
2	<p>Africa's Bioinformatics Roots are grounded at UWC: O'Connell's Vision for Innovation Alan Christoffels In: From Hope to Action through Knowledge: The Renaissance of the University of the Western Cape, 2001-2014 R Bharuthram, L Pokpas Published by UWC Press and African Sun Media 27 October 2020 ISBN: 978-1-990995-00-2 https://doi.org/10.18820/9781990995019</p>



CONFERENCE PARTICIPATION

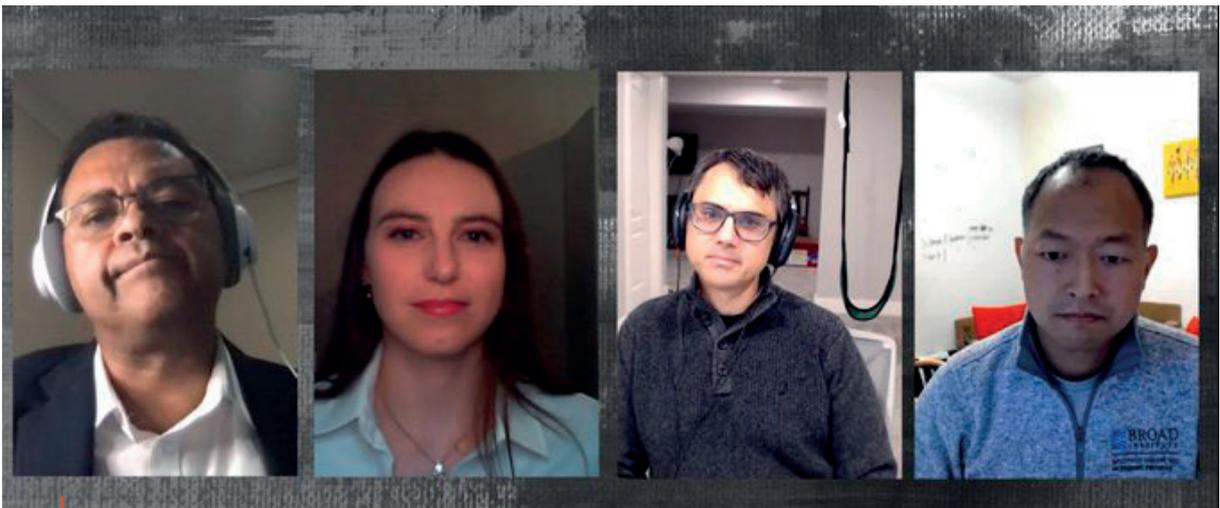
Presenter	Date	Conference Name and Venue	Title	Type of Participation
Alan Christoffels	24 June 2020	H3ABioNet Seminars series -online	Public Health Alliance for Genomic Epidemiology (PHA4GE): Bioinformatics for Public Health	Seminar
Dominique Anderson	24 – 27 August 2020	BIO Africa 2020 Digital Convention – online	Baobab LIMS	Profile Video
Christianah Taye Omotoso	November 2020	Cold Spring Harbor Laboratory Biological Data Science (Virtual) Conference	Molecular and mechanistic insight into the varied rate of metabolism by allele of CYP3A4 enzyme (Cold Spring Harbor Laboratory Biological Data Science (Virtual) Conference)	Poster
Katelyn Cuttler	November/ December 2020	14 th CHPC National Conference	Molecular dynamic simulation of a variant found in a South African family with Parkinson's disease	Poster Won Third Prize
Yamkela Mgwatyu, Allison Anne Stander	December 2020	CHPC National (Virtual) Conference 2020	Rooibos-specific polyphenol biosynthesis genes – a genome-wide analysis	Oral
Dominique Anderson	3 December 2020	International FAIR Convergence Symposium	SARS-CoV2- Metadata standard	Oral

CONFERENCE ATTENDANCE

Attended by:	Date	Conference Name and Venue	Purpose and Benefit
Dominique Anderson	17 – 18 June 2020	London Calling. Oxford Nanopore conference.	London Calling is a conference dedicated to scientific research using nanopore DNA/ RNA sequencing.
Christianah Omotoso	4 – 6 November 2020	Cold Spring Harbour Laboratories -Biological Data Science Virtual Conference	The scope of this meeting was the infrastructure, software, and algorithms needed to analyse large data sets in biological research.
Dominique Anderson	24 – 27 November 2020	SA-Sweden University Forum digital conference	60 Parallel Workshops, providing a flexible, engaging platform to share and present research and collaborations between Sweden and South Africa.

INVITED TALKS

Presenter	Date	Venue	Title
Alan Christoffels	19-21 October 2020	Grand Challenges Annual Meeting. Virtual Meeting.	Invited Session Moderator: Leveraging pathogen genetic sequencing.
	15 November 2020	American Society for Tropical Medicine and Hygiene. Virtual Meeting.	Towards the standardisation of genomic tools to support detection and control of MDR Tuberculosis.
Dominique Anderson	29 September 2020	Online	UWC's Research Week: Biobanking educational book



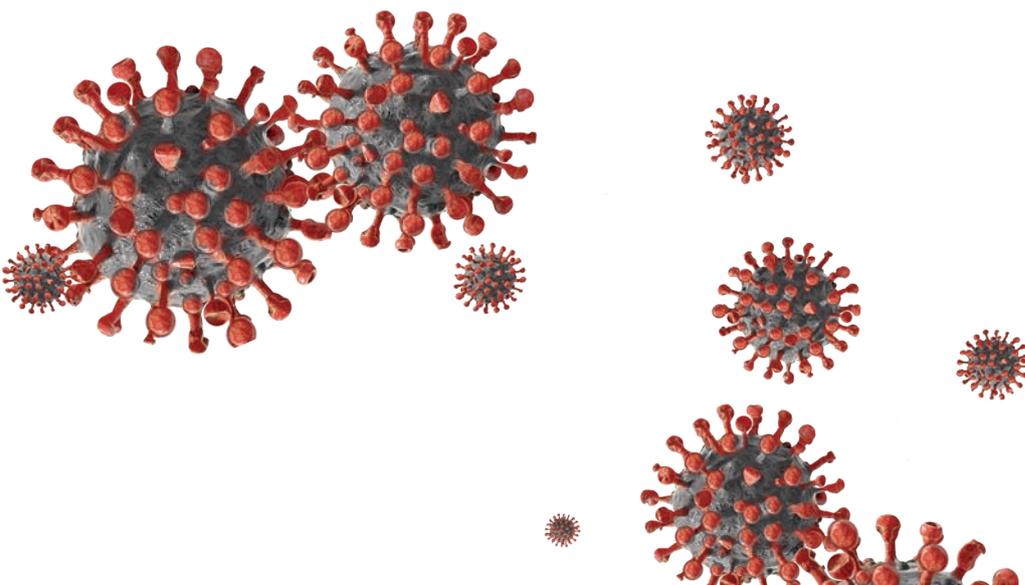
Alan Christoffels (far left) moderating a panel of speakers at the online Grand Challenges Annual Meeting 2020. 2nd from Left: Dr Emma Hodcroft, Nextstrain and University of Basel, Switzerland; Dr Anders Goncalves, University of Melbourne, Australia; Dr Daniel Park, The Broad Institute, USA.



Left to Right: Prof Erik Bongcam-Rudloff, Swedish University of Agricultural Sciences; Dr Carmen Swanepoel, Stellenbosch University; Dr Dominique Anderson, University of the Western Cape; Roxana Merino Martinez, Karolinska Institute; Prof Alan Christoffels, University of the Western Cape.

HOSTING OF VISITING ACADEMICS

Visitor Name	Date	Meeting Name	Purpose of meeting + outcomes
Erik Bongcam-Rudloff (Swedish University of Agricultural Sciences), Roxana Merino (Karolinska University), Carmen Swanepoel (Stellenbosch University)	03 - 06 March 2020	SA-Sweden collaboration meeting	South Africa Sweden partnership and opportunities. Writing and submitting the STINT grant to develop international programmes. Discussions of other potential partnership opportunities.
Elizabeth Neuhaus (US-CDC)	17 February 2020	Baobab LIMS integration	Meeting to discuss the integration of Baobab LIMS into a larger CDC e-infrastructure. Various follow-up meetings have resulted in successful alpha testing.



EXPERT PANEL OR COMMITTEE MEMBERSHIP

Name	Membership and Purpose	Role
Alan Christoffels	African Society for Bioinformatics and Computational Biology http://www.asbcb.org .	President
	TBGenAfrica: H3Africa funded TB consortium	Scientific Advisory Board Member
	PHINDAccess: EU-funded project in Tunisia to build capacity in Host-pathogen interaction omic analysis	Scientific Advisory Board Member
	East African Bioinformatics Network: Fogarty Funded Bioinformatics Network in East Africa	Advisory Board Member
	Global Emerging Pathogens Consortium: to create awareness of biosecurity threats across Africa, and to empower African scientists to respond to disease outbreaks	Co-Chair the annual conference and develop training modules
	African Academy of Sciences: to develop guidelines for biospecimen and data management in Africa	Biospecimen and Data Governance Expert Panel
	NRF Rating Panel	Assessor
	Department of Science and Technology Genomic and Precision Medicine Think Tank: to develop a strategic framework for personalised medicine in South Africa	Member
	National Health Laboratory Services: biospecimen and data access ethics	Review guidelines for data sharing
	Ravnit Grewal	The Selection and use of essential in vitro diagnostics. WHO Technical Report Series 1031. Report of the third meeting of the WHO Strategic Advisory Group of Experts (SAGE) on In vitro Diagnostics, 2020.
		
Peter van Heusden	PHA4GE Infrastructure Working Group	Chair

JOURNAL EDITING AND REVIEWS

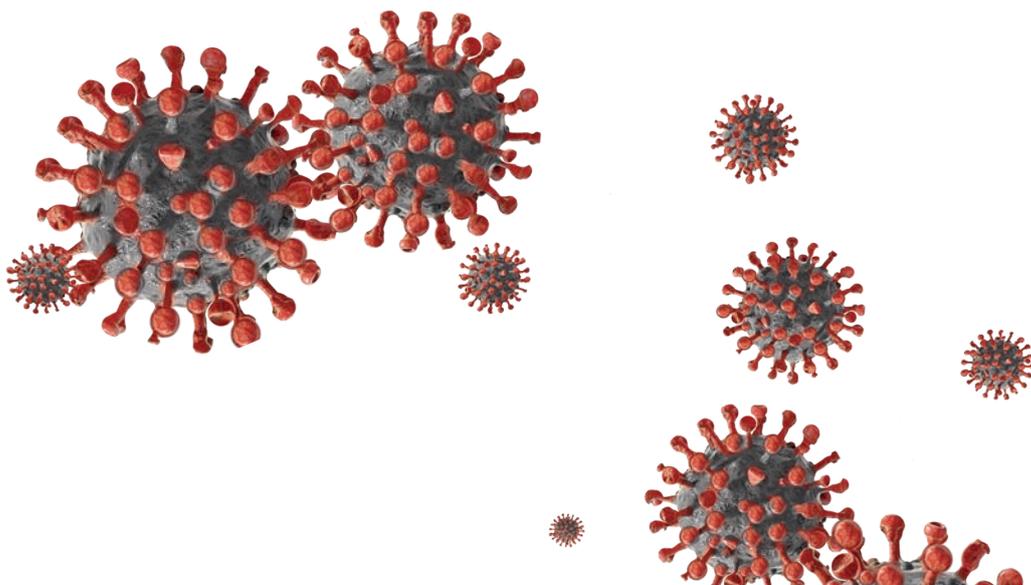
PI Name	Journal
Alan Christoffels	Editor: The Improper Scientist: Quarterly magazine for the African Society for Bioinformatics and Computational Biology
	Editorial Board Member: Data Journal
	Reviewer: BMC Genomics, BMC Bioinformatics, Bioinformatics, PLoS ONE
Peter van Heusden	Reviewer: Journal of Open Source Education, Bioinformatics, Bioinformatics and Biology Insights, PLoS Computational Biology

THESIS EXAMINED FOR STUDENTS AT OTHER INSTITUTIONS

Examiner	Degree	Institute
Hocine Bendou	MSc	University of the Witwatersrand
Alan Christoffels	PhD	Nelson Mandela African Institution for Science and Technology, Tanzania
	MSc	Rhodes University
	MSc	University of Pretoria

EXTERNAL MODERATION

Name	Institution	Course
Alan Christoffels	Stellenbosch University	Biochemistry Department: 3 rd year and Honours modules



RESEARCH PROJECT THEMES

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

PI	Theme	Projects
Dominique Anderson	Informatics Tools	<ul style="list-style-type: none"> • Biobank informatics - multiple collaborations for biocollection sample quality management and e-infrastructure with a focus on Baobab LIMS. • NHLS collaboration for data marts in public health settings.
	Biomedical data governance	<ul style="list-style-type: none"> • Collaboration with biobank and law experts to evaluate data governance at tertiary institutions and biomedical research facilities.
	Dental Genetics	<ul style="list-style-type: none"> • Dental genetics in collaboration with the Faculty of Dentistry investigating oral microbiomes.
	One Health	<ul style="list-style-type: none"> • <i>In silico</i> research focussed on OneHealth.
Hocine Bendou	Cancer Bioinformatics	<ul style="list-style-type: none"> • Novel genomic biomarkers for Pediatric and Adult Acute Myeloid Leukemia. • Computational analysis of multi-omic data for the elucidation of molecular mechanisms of neuroblastoma.
	Pipeline development	<ul style="list-style-type: none"> • Development of a Nextflow pipeline for the detection of enhancer-associated noncoding variants using ChIP-seq data.
	Machine learning	<ul style="list-style-type: none"> • Prediction of synthetic lethal gene interaction in human cancer using supervised machine learning.
	Gene regulatory networks	<ul style="list-style-type: none"> • Gene regulatory network study of diffuse large B-cell lymphoma.
	Viral Study	<ul style="list-style-type: none"> • Identify impact of viral inserts in human esophageal cancer. A collaborative work with the University of Cape Town.
Alan Christoffels	Pathogen genomics	<ul style="list-style-type: none"> • In collaboration with the Tygerberg MRC Unit, we are developing methods to analyse high throughput sequencing data for microbial genomes. • Identification of novel drug targets in pathways known to contain drug resistant genes. • Predicting the interaction networks between human and mycobacteria • Identification of operons & non-coding RNAs to understand gene regulation in pathogen genomes.
	Pathogen surveillance	<ul style="list-style-type: none"> • Developing systems to support reproducible analyses • Development of global standards for data sharing and integration • Public Health Alliance for Genomic Epidemiology (PHA4GE)
	Drug discovery	<ul style="list-style-type: none"> • Identifying drug molecules against multi-drug resistant <i>M.tuberculosis</i> • In partnership with School of Pharmacy at UWC, delineate the pharmacokinetics of drugs in response to host genetic factors • In partnership with School of Pharmacy at UWC, use of Machine learning techniques to prioritise molecules in natural compounds.

Ruben Cloete	Pathogen resistance	<ul style="list-style-type: none"> NGS sequencing of Tuberculosis isolates and HIV plasma samples allowed the identification of novel mutations. To identify mutations associated with drug resistance we develop structural computational pipelines to further our understanding of HIV-1 and tuberculosis drug resistance. To identify drug resistance pathways in HIV-1 using Phylogenetics.
	Drug discovery	<ul style="list-style-type: none"> Previous pathway mapping and gene prioritisation methods allowed us to identify known and novel <i>Mycobacterium tuberculosis</i> drug targets. Here we exploit the drug targets using virtual screening methods to identify novel drugs to treat drug resistant tuberculosis. The drugs identified are tested for activity against <i>Mycobacterium tuberculosis</i> using a bioluminescent reporter assay system. In silico discovery of compounds with activity against the novel SARS-coronavirus and in vitro testing.
	Human disease genomics	<ul style="list-style-type: none"> Whole genome sequencing identified variants that co-segregate within South African families with Parkinson's disease. Here we use various software tools to prioritise these variants to identify causal variants that might be associated with Parkinson's disease.
	Machine learning approaches	<ul style="list-style-type: none"> Using machine learning methods to identify putative compounds with activity against <i>Mycobacterium tuberculosis</i> and HIV.
	Diagnostic tool development	<ul style="list-style-type: none"> Develop structural methods to aid in the identification of biomarkers useful for Tuberculosis and Ebola virus diagnosis.
Ravnit Grewal	Precision Medicine	<ul style="list-style-type: none"> The role of precision medicine in the diagnosis of HIV related lymphomas.
Gordon Harkins	SARS-CoV-2 Phylogeography	<ul style="list-style-type: none"> The characterisation and clinical manifestation of the SARS-CoV-2 outbreak in Uruguay.
	SARS-CoV-2 Dynamics	<ul style="list-style-type: none"> The dispersal history and dynamics of SARS-CoV-2 lineages during the first epidemic wave in New York City.
	SARS-CoV-2 Evolution	<ul style="list-style-type: none"> The role of natural selection in the emergence and ongoing evolution of the N501Y lineages (V1-V3) of SARS-CoV2.
	HIV-1 Latent Viral Reservoir Dynamics	<ul style="list-style-type: none"> Reservoir size determination and timing of entry of HIV-1 variants into the latent reservoir. The viral dynamics involved in HIV-1 subtype C latent reservoir formation, maintenance and evolution. The role of HIV-1 Nef and LTR functions in determining latent reservoir size.
	HIV-1 Superinfection	<ul style="list-style-type: none"> Dating the origin of independently acquired viral lineages in HIV-1 superinfection.
	HIV -1 Immunology	<ul style="list-style-type: none"> Determining whether correlates of protection are associated with clinical outcome in natural infection.
	Virulence Evolution	<ul style="list-style-type: none"> An investigation of how virulence and pathogenicity are evolving in viral crop pathogens.
	Viral Emergence	<ul style="list-style-type: none"> Investigation of the spatial and temporal diversity, demographics and evolutionary dynamics of viral communities inhabiting natural ecosystems that are linked to the emergence of socially relevant crop-infecting geminiviral diseases.
	Endogenous Viral Elements (EVEs)	<ul style="list-style-type: none"> Investigation of the role of EVEs in virus evolution and their functions in plants to address fundamental biological issues that are critical to the understanding of plant virus evolution.
Phytovirus Diversity	<ul style="list-style-type: none"> Determining phytoviral species richness of plant communities within several unmanaged and managed ecosystems. 	
Peter van Heusden	Research Computing Infrastructure	<ul style="list-style-type: none"> Studying the deployment of research computing infrastructure and development of associated skills on the African continent.
	Pathogen surveillance Platforms	<ul style="list-style-type: none"> Researching the use of pathogen genomics in a public health context for drug resistance prediction and genomic surveillance.
	Scientific workflows for pathogen genomics	<ul style="list-style-type: none"> Developing workflow systems for pathogen genomics with a focus on Galaxy.

RESEARCH LABORATORIES

RESEARCH LABORATORY OF DOMINIQUE ANDERSON



Front row Left to Right: Abiola Babajide (PhD candidate), Rudolph Serage (MSc candidate), Ziphozakhe Mashologu (Software developer).
Back row Left to Right: Quinton Coert (Software developer), Oluwafemi Peter Abiodun (PhD candidate), Dominique Anderson (PI). Absent: Lunga Baliwe (Software developer).

Highlights of the lab for 2020

- Supervisor of 1 PhD student and 1 MSc student
- Co-supervisor of 1 PhD student, 1 MSc student
- Grants: Africa CDC awarded for LIMS customisation.

RESEARCH COLLABORATIONS

1. Data governance for biobanks

Collaborating Parties:

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

Nature and purpose:

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

Output in the last 12 months:

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

2. Dominique Anderson, Oluwafemi Peter Abiodun, Alan Christoffels, Information security at South African universities – implications for biomedical research, International Data Privacy Law, Volume 10, Issue 2, May 2020, Pages 180–186,
<https://doi.org/10.1093/idpl/ipaa007>

Future Direction:

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

2. Dental genetics

Collaborating Parties:

Dr Manogari Chetty - Faculty of Dentistry, UWC
Dr Anthea Jeftha - Faculty of Dentistry, UWC

RESEARCH PROJECTS

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

With regards to data privacy and governance, my collaborators and I combine expertise in biobanking, bioinformatics and big data, and law to investigate how both regulations and best practices in ICT influence the biomedical data and data sharing landscape in Africa. The UWC Dentistry faculty and SANBI have also started a collaborative effort to increase research into dental genetics. In the future, I hope to establish networks within the agricultural, veterinary, medical, and environmental sphere in order to intensify research in the area of *One Health*.

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.

3. Biobank informatics and community engagement

Collaborating Parties:

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

Nature and purpose:

Enhancement and customisation of Baobab LIMS for LMICs. Standardisation of e-infrastructure. Science communication and community engagement.

Output in the last 12 months:

TIA seed fund grant award.
West Africa Health Organisation/ Institute Pasteur Côte d'Ivoire LIMS customisation.

One publication:

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

Future Direction:

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

4. OneHealth

Collaborating Parties:

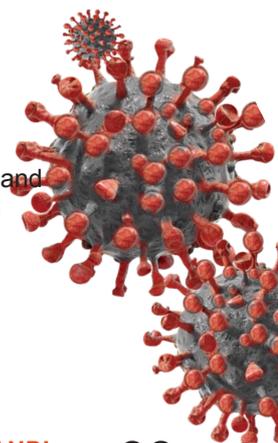
Seeking collaborative partners

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.

Output in the last 12 months: None

Future Direction: None



RESEARCH LABORATORY OF HOCINE BENDOU



Members of the Bendou lab.

RESEARCH COLLABORATIONS

1. Molecular mechanisms of neuroblastoma

Collaborating Parties:

Dr Azeez Fatai - Department of Biochemistry, Lagos State University, Lagos, Nigeria

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:

Giwa A, Fatai A, Gamielien J, Christoffels A, Bendou H. Identification of novel prognostic markers of survival time in high-risk neuroblastoma using gene expression profiles. *Oncotarget*. 2020;11(46):4293-4305. Published 2020 Nov 17. doi:10.18632/oncotarget.278082. Molecular Biology and Genetics of Oesophageal Cancer

2. Molecular Biology and Genetics of Oesophageal Cancer

Collaborating Parties:

Prof Iqbal Parker - UCT
Dr Hocine Bendou - SANBI, UWC

Nature and purpose:

Provide bioinformatics expertise and PhD student co-supervision.

Output in the last 12 months:

Ongoing work on viral DNA integrated in human oesophageal cancer.

Future Direction:

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

Highlights of the lab for 2020

Student supervision:

Nomlindelo Mfuphi - MSc, SANBI
Wardah Jassiem - MSc, SANBI
Cleo-Rose Barendse - MSc, SANBI
Abdulazeez Giwa - PhD, SANBI
Nasr Eshibona - PhD, SANBI
Michelle Livesey - PhD, SANBI
Victoria Patten - PhD, UCT

RESEARCH PROJECTS

Our interest is to study cancer using bioinformatics approaches. Cancer is a dynamic disease that evolves to become more aggressive and heterogeneous. This heterogeneity has been shown to cause resistance to cancer therapies and lead to shortened patient survival. Precise stratification of tumor heterogeneity is essential to administer and develop more effective therapies. Computational methods are needed to assess tumor heterogeneity. Artificial Intelligence by Machine Learning (ML) can be used to demonstrate the existence of heterogeneity within tumors from same cancer. It can also be used in combination with other methods such as differential gene/methylation analyses to find prognostic signatures of a cancer condition. We have machine learning algorithms to different cancer research projects involving postgraduate students to investigate heterogeneity in Neuroblastoma, Pediatric Acute Myeloid leukemia, Oesophageal cancer, Colorectal cancer and DLBCL.

RESEARCH LABORATORY OF ALAN CHRISTOFFELS



Members of the Christoffels lab.

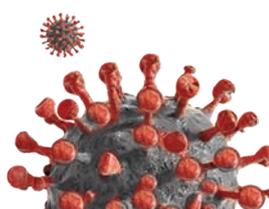
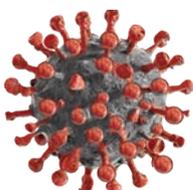
Highlights of the lab for 2020

During lockdown in 2020, we maintained an active research programme without our group. 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. In August 2020, the work done by members of the team were recognised when SANBI was identified as a reference lab for the World Health Organisation Africa region to respond to disease outbreaks, including COVID-19.

Data security continues to be high on the national agenda in South Africa. Our work on assessing data security at universities was published in a prestigious international law journal. Subsequently I (Alan Christoffels) was appointed to the ASSAF-led steering committee to develop the code of conduct for research that aligns to the Protection of Personal Information Act.

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. This work was presented at the GrandChallenges 2020 conference.

A total of 2 PhDs and 1 MSc student completed their degrees. Their work was embedded in a total of 18 publications from the group.



RESEARCH PROJECTS

The projects below underpin our translational work:

Tuberculosis

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

Medicinal plant target prioritisation

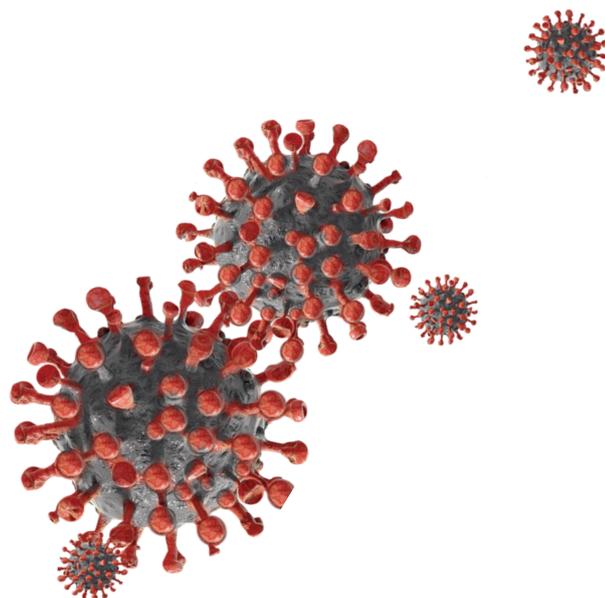
The medicinal benefits of plant extracts generates a plethora of data. In collaboration with the UWC School of Pharmacy, we are extending our predictive analytics capability to prioritize candidate molecules for drug resistant bacteria. This work builds on our work on anti-malarial targets.

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.

Applications of Machine Learning methods to protein-protein interactions

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



RESEARCH COLLABORATIONS

1. Prediction of human-tuberculosis interaction networks

Collaborating Parties:

Prof Peter Witbooi - Mathematics Department, UWC.

Nature and Purpose:

Use machine-learning techniques to improve the prediction of host-pathogen interactions between human and pathogens, and model disease outbreaks.

Output in the last 12 months:

Developed a population model for the 2017 listeriosis outbreak in SA.

2. Develop a biobank informatics management system

Collaborating Parties:

Dr Dominique Anderson - SANBI, UWC; Africa Centers for Disease Control.

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 module. Submitted a book chapter on biobank software for low and middle income countries.

3. Analysing genetic networks in *M.tuberculosis*

Collaborating Parties:

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

Nature and Purpose:

Identification of Operons: We rewrote an algorithm previously published by a team at Johns Hopkins University.

Identification of non-coding RNAs.

Output in the last 12 months:

Our operon detection algorithm has been tested with different metrics. And currently benchmarked against two international methods.

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

4. Identification of novel drug targets for drug resistant tuberculosis

Collaborating Parties:

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

Nature and Purpose:

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

Output in the last 12 months:

We continue to refine our computational screening system.

5. 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 and show anti-plasmodial activity. In past we have successfully used a range of computational tools to compare plant extracts that show anti-plasmodial 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 Pharmacy has completed his thesis on "Hit Prioritisation of Natural Products with Activities Against Methicillin-Resistant *Staphylococcus aureus* (MRSA)".

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

Collaborating Parties:

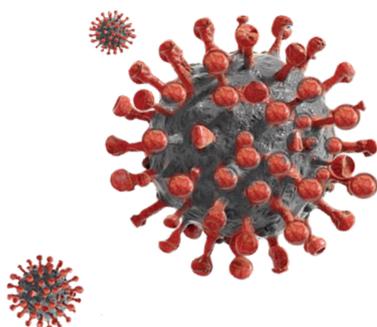
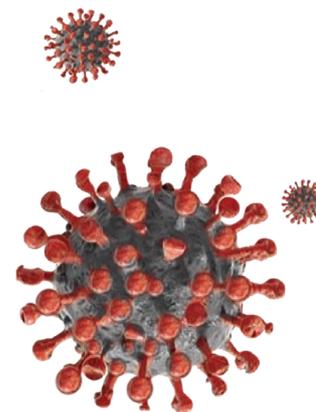
Prof Samantha Sampson - University of Stellenbosch
Prof Rob Warren - University of Stellenbosch

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.



7. Identification of SNPs implicated in rare diseases

Collaborating Parties:

Professor Manogari Chetty - Dentistry Faculty, UWC.

Nature and Purpose:

We aim to utilise bioinformatics tools to support the Dentistry faculty to analyse the genetic basis of some of the syndromes observed in the 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.

8. Biosecurity Preparedness in Africa

Collaborating Parties:

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

Nature and Purpose:

We aim to establishing infrastructure and research capacity to respond to highly infectious emerging Pathogens.

Output in the last 12 months:

We have reviewed the information security policies at SA Universities with a view to propose guidelines for data management that impacts biomedical researchers.

9. 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 meta data standard for COVID-19 biospecimen collection.

10. African Genomics Archive

Collaborating Parties:

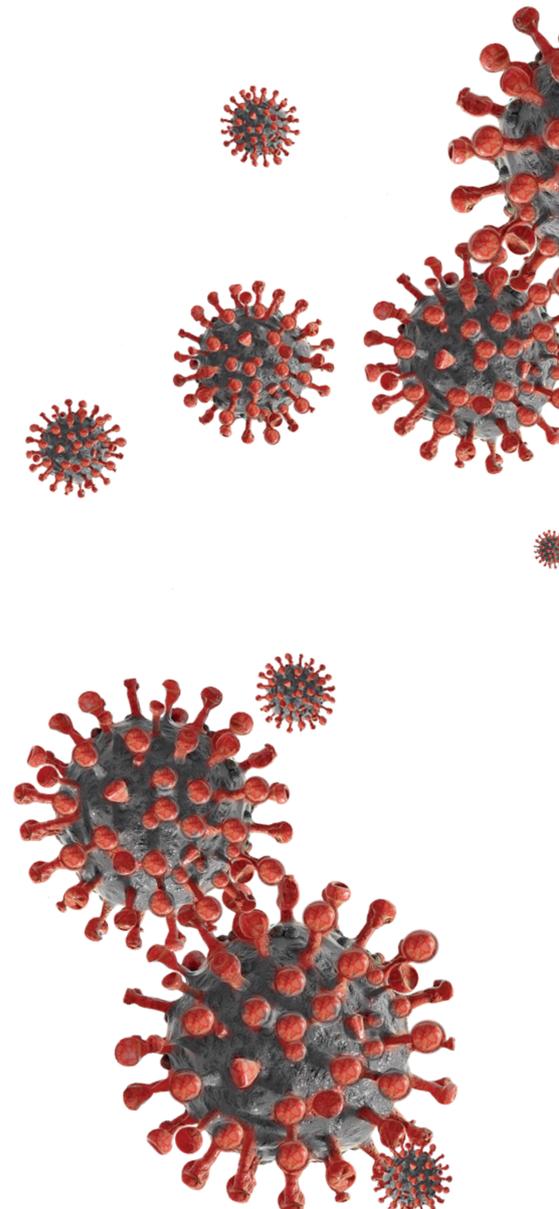
Africa CDC

Nature and Purpose:

To develop 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.



RESEARCH LABORATORY OF RUBEN CLOETE



Members of the Cloete lab.

Highlights of the lab for 2020

Students:

Sello Given Mikasi, the PhD at Stellenbosch University who I co-supervised, graduated. Additionally, I supervised two Honours students (Kimberly Coetzer and Lunathi Mhlangeni Ndlondlo) from the Department of Biotechnology, University of the Western Cape. Another PhD student that I co-supervise at Stellenbosch University, Katelyn Cuttler, won third prize at the Centre for High Performance Computing (CHPC) conference 2020.

Our research collaboration with Stellenbosch University yielded 4 journal publications on the project: Structural impact of resistance associated mutations in the South African HIV-1C integrase protein.

RESEARCH PROJECTS

My primary research interests focus on understanding drug resistance in *Mycobacterium tuberculosis* and Human Immune deficiency virus (HIV-1) integrase protein and the identification of causal variants in Parkinson's disease development. Tuberculosis and HIV-1 drug resistance is a huge problem in South Africa and calls for the identification of newer drugs to curb the spread of these diseases. The use of computational methods to speed up the process of drug discovery can reduce the cost and time spent pursuing drugs that later fail in clinical trials. We have developed computational pipelines within my laboratory to interrogate drug targets to screen for drugs that inhibit enzyme targets that can be purchased and tested in vitro for activity against *Mycobacterium tuberculosis*.

We also use computational methods to validate mutations and their effects on protein structure and function. This is also very useful to determine if a drug remains within an enzyme active pocket and if the mutation results in reduced binding or no binding.

The result of this research may provide improved treatment regimens for South African populations to combat infectious disease and non-communicable diseases like Parkinson's.

RESEARCH COLLABORATIONS

1. Novel drug discovery against *Mycobacterium tuberculosis*

Collaborating Parties:

Prof Alan Christoffels - SANBI, UWC
Dr Melanie Grobbelaar - Stellenbosch University
Prof Samantha Samson - Stellenbosch University

Nature and purpose:

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

Output in the last 12 months:

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

Future Direction:

Rv2421c manuscript submitted to PLoS ONE under review.



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

Collaborating Parties:

Dr Graeme Jacobs - Stellenbosch University

Nature and purpose:

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

Output in the last 12 months:

1) Mikasi SG, Gichana JO, Van der Walt C, Brado D, Obasa AE, Njenda D, Messembe M, Lyonga E, Assoumou O, Cloete R, Ikomey GM, Jacobs GB. HIV-1 Integrase Diversity and Resistance-Associated Mutations and Polymorphisms Among Integrase Strand Transfer Inhibitor-Naive HIV-1 Patients from Cameroon. *AIDS Res Hum Retroviruses*. 2020 May;36(5):450-455. doi: 10.1089/AID.2019.0264. Epub 2020 Jan 13. PMID: 31830799.

2) Drug Resistance Mutations Against Protease, Reverse Transcriptase and Integrase Inhibitors in People Living with HIV-1 Receiving Boosted Protease Inhibitors in South Africa. Obasa, Adetayo Emmanuel; Mikasi, Sello Given; Brado, Dominik; Cloete, Ruben; Singh, Kamlendra; Neogi, Ujjwal; Jacobs, Graeme Brendon ISSN: 1664-302X, 1664-302X; DOI: 10.3389/fmicb.2020.00438. *Frontiers in microbiology*, 2020, Vol.11, p.438

3) Chitongo R, Obasa AE, Mikasi SG, Jacobs GB, Cloete R (2020) Molecular dynamic simulations to investigate the structural impact of known drug resistance mutations on HIV-1C Integrase-Dolutegravir binding. *PLoS ONE* 15(5): e0223464. <https://doi.org/10.1371/journal.pone.0223464>

4) Isaacs D, Mikasi SG, Obasa AE, Ikomey GM, Shityakov S, Cloete R, Jacobs GB. Structural Comparison of Diverse HIV-1 Subtypes using Molecular Modelling and Docking Analyses of Integrase Inhibitors. *Viruses*. 2020 Aug 26;12(9):936. doi: 10.3390/v12090936. PMID: 32858802; PMCID: PMC7552036.

Future Direction:

Two more manuscripts in preparation.

3. Prioritising mutations identified in South African Parkinson's disease patients using structural methods

Collaborating Parties:

Prof Soraya Bardien - Stellenbosch University

Nature and purpose:

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

Output in the last 12 months: None

Future Direction:

Revised Manuscript is under review.

4. Haplotype variation within 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:

Manuscripts in preparation.

5. Comparative studies of more and less virulent coronavirus envelope proteins to understand pathogenicity

Collaborating Parties:

Prof Burtram Fielding - Molecular Biology and Virology Research Laboratory, Department of Medical Biosciences, University of the Western Cape.

Nature and purpose:

Perform topology predictions, structural modeling as well as simulation studies to understand the pathogenic differences between the seven coronavirus envelope proteins

Output in the last 12 months: None

Future Direction:

Manuscript under review.

RESEARCH LABORATORY OF RAVNIT GREWAL

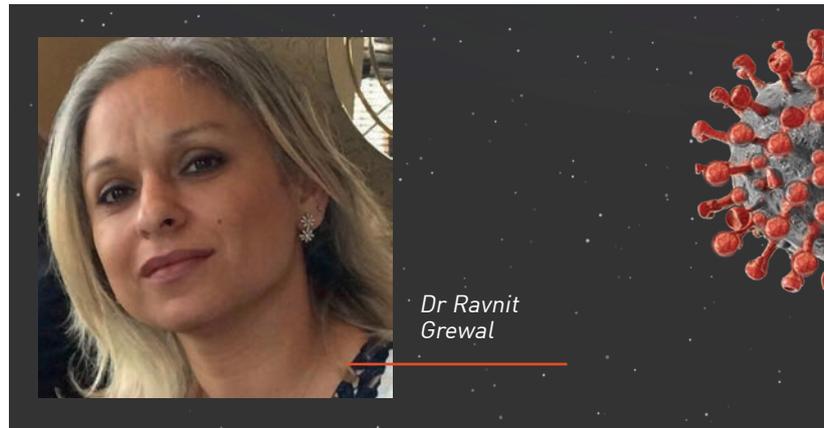
As a Haemato-Pathologist my primary research interest is in HIV related lymphomas. As a consequence of this enquiry, I spearheaded a research group that studied lymphomas as it became evident that we were experiencing a surge in HIV associated complex lymphoma pathology in our laboratory. As people with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

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

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

RESEARCH COLLABORATIONS:

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



Dr Ravnit Grewal

RESEARCH OUTPUTS:

1. WHO Strategic Advisory Group of Experts (SAGE) on In vitro Diagnostics (2020). Essential Diagnostic List

Reviewed the Haematology related diagnostic tests to be implemented in middle to low income countries. <https://www.who.int/publications/item/9789240019102>

2. Identification of a novel WAS mutation in a South African patient presenting with atypical Wiskott-Aldrich syndrome: a case report.

Brigitte Glanzmann, Marlo Möller, Mardelle Schoeman, Michael Urban, Paul D. van Helden, Lisa Frigati, **Ravnit Grewal**, Hermanus Pieters, Ben Loos, Eileen G. Hoal, Richard H. Glashoff, Helena Cornelissen, Helena Rabie, Monika M. Esser & Craig J. Kinnear
BMC Medical Genetics 21, 124 (2020).
<https://doi.org/10.1186/s12881-020-01054-6>

3. Differential Diagnosis of Malignant Lymphadenopathy Using Flow Cytometry on Fine Needle Aspirate: Report on 269 Cases

Carla Griesel, Minodora Desmirean, Tonya Esterhuizen, Sergiu Pasca, Bobe Petrushev, Cristina Selicean, Andrei Roman, Bogdan Fetica, Patric Teodorescu, Carmen Swanepoel, Ciprian Tomuleasa and **Ravnit Grewal**.
Journal of Clinical Medicine 2020, 9, 283
doi:10.3390/jcm9010283

RESEARCH LAB OF ULJANA HESSE



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 includes over 3000 plant species actively used in traditional medicine. These plants produce a wide range of medicinally active compounds, many of which are rare or even species specific. One of the more famous endemic South African medicinal plant species is rooibos, best known as a herbal tea. It only grows in the Cederberg Mountain Region and represents a key cash crop for the local farming communities. Rooibos is increasingly recognised as a potential phytopharmaceutical: the species produces a wide range of phenolic compounds, which are associated with diverse medicinal properties of the plant (including anti-diabetic, cardioprotective, antispasmodic, anti-aging effects). It therefore represents a suitable pilot plant species to initiate medicinal plant genomics research that focuses on the endemic flora of South Africa. The Rooibos Genomics Programme aims to 1) locally establish all methodologies essential for medicinal plant genome analysis, 2) generate a high-quality assembly of the rooibos genome with extensive functional annotation of the rooibos genes, 3) identify rooibos genes involved in medicinal compound production and plant stress tolerance.

Dr Uljana Hesse has had a longstanding partnership with SANBI that dates back to her postdoc at SANBI. Currently she supervises students that are registered at SANBI and in her current academic department (Biotechnology).

Collaboration with SANBI:

The rooibos genome and eight rooibos transcriptomes have been sequenced using Illumina sequencing technologies. Establishment of biocomputational data analyses procedures is conducted at SANBI. In 2020, my research team has:

- 1) Completed phylogenetic analyses of rooibos genes involved in biosynthesis of phenolic compounds;
- 2) Established 3rd generation sequencing (Oxford Nanopore; MinION) at UWC;
- 3) Generated long reads (2-40kb) for the rooibos genome (1x genome coverage);
- 4) Investigated Convolutional Neural Networks towards establishing machine learning algorithms for plant protein classification.

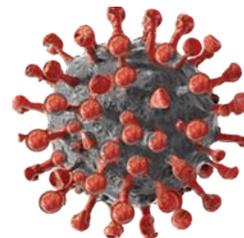
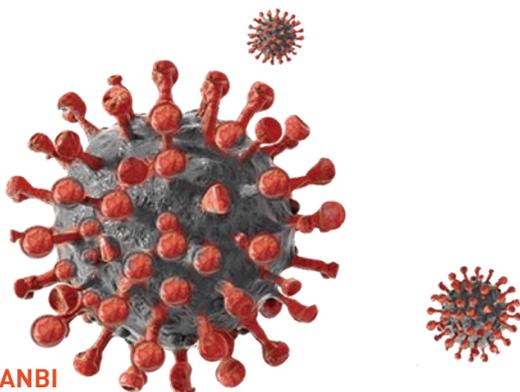
Student graduated in 2020:

Allison Stander, MSc Bioinformatics. Thesis Title: "De novo assembly of the rooibos genome".

Publications:

Mgwayu, Y., Stander, A.A., Ferreira, S., Williams, W. and Hesse, U., 2020. Rooibos (*Aspalathus linearis*) Genome size estimation using flow cytometry and k-mer analyses. *Plants*, 9(2), p.270.

Stander, E.A., Williams, W., Mgwayu, Y., Heusden, P.V., Rautenbach, F., Marnewick, J., Roes-Hill, M.L. and Hesse, U., 2020. Transcriptomics of the Rooibos (*Aspalathus linearis*) Species Complex. *BioTech*, 9(4), p.19.



RESEARCH LABORATORY OF GORDON HARKINS



*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 tackle the unprecedented volume of SARS-CoV-2 genomic and spatial data that was being generated and shared with the scientific community to gain real-time insights into the virus transmission during a viral pandemic.

RESEARCH PROJECTS:

A summary of some of the research projects that my laboratory has been involved in during 2020 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 where the virus 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.

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 epicenter of the pandemic. In order to gain insights on 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

Highlights of the lab:

In collaboration with researchers from the Institute of Infectious Diseases and Molecular Medicine at UCT and Johns Hopkins University, School of Medicine in Baltimore, USA, we secured funding from the National Institutes of Health, USA (NIH) and the South African Medical Research Council to the amount of R13.7 million (UWC sub-award R4.5 million) over the next five years to characterise the dynamics of the HIV-1 subtype C latent viral reservoir and determinants thereof in African populations.

Four manuscripts and one book chapter were submitted for publication in 2020 three of which were published within the calendar year and the fourth, in 2021. Included as co-authors on the 2021 article were my honours student from 2020 Miss Sabine Smidt and a previous masters student of mine Dr Batsirai Mabvakure.

dispersal history and dynamics of viral lineages at the state and city level. Our finding 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-CoV2.

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) Immunological responses and their impact on clinical outcomes in epidemiologically linked HIV-1 transmission

In collaboration with researchers from Department of Pathology, NYU Grossman School of Medicine, New York, NY, USA, the Johns Hopkins School of Medicine, Baltimore, MD, USA,

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

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.

Virulence Evolution

This is a long-term collaborative effort to investigate how the traits of virulence and transmission rates are evolving in viral pathogens important to human health. Using a combination of laboratory experimentation and state-of-the-art computational analyses, we are investigating the spatiotemporal dynamics and evolution of a broad range of viral pathogens focusing primarily on viruses that cause important crop diseases. Collectively, these studies are aimed at improving our understanding of the different life history strategies of pathogens, including reproduction, migration between and within hosts and virulence. By explicitly connecting our pathogenicity and virulence measurements to the natural history of the pathogen, we are confident this will provide a broader perspective on the evolution of pathogens for which current findings are mainly limited to theoretical or experimental settings. By producing infectious clones of ancient genomes from herbarium specimens and re-synthesised computationally

inferred ancestral genomes, it will be possible to pinpoint when and where over the past few centuries, major changes in virus virulence have occurred. This project will complement the results obtained from our previous work on virulence evolution in Maize streak virus and increase our understanding of the evolutionary and epidemiological processes that shape the diversity in an important group of crop pathogens.

Viral Emergence

In collaboration with researchers from CIRAD France, the State University of Arizona and UCT, we have adopted a “geo-metagenomics” approach to investigate whether ecological disturbances such as intensive agriculture, cause measurable changes in the spatial and temporal diversity, demographics and evolutionary dynamics of viral communities inhabiting natural ecosystems that are linked to the emergence of socially- relevant crop-infecting geminiviral diseases. Our study population is the fynbos ecosystem situated in the Western Cape region of South Africa – a biodiversity hot-spot which is extremely threatened by increasing human population densities, urbanisation and agriculture. To complement the data on contemporary viral diversity

derived from field studies, we are sequencing “ancient” viral genomes present in fynbos plant specimens stored within French and South African herbaria to more accurately estimate when and where, key ancestral plant viruses existed.

Phytovirus Diversity

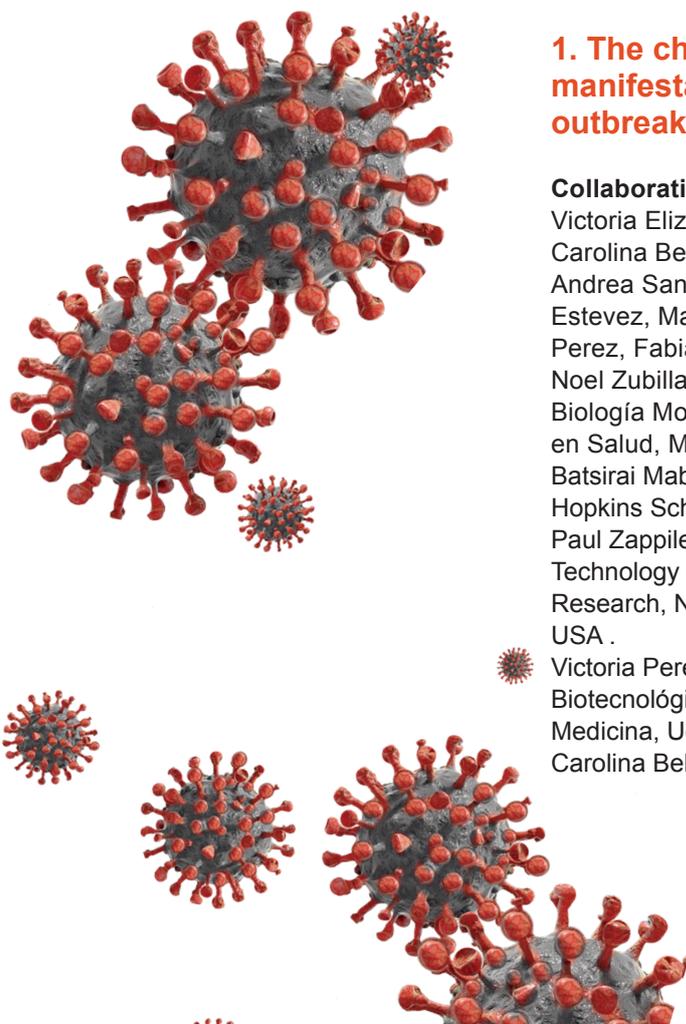
Emergent diseases of plants, a high proportion of which are caused by phytoviruses, are a significant burden on the food security and economic stability of society. However, a comprehensive view of the geographical distribution of phytovirus diversity does not exist, including both the numbers or richness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

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, MD, USA.
Paul Zappile Christian Marier- Genome
Technology Center, Office for Science and
Research, NYU Langone Health, New York, NY,
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, NY, USA.
Simon Dellicour- Spatial Epidemiology Lab. (SpELL), Université Libre de Bruxelles, Bruxelles, 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, NY, 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, Bruxelles, 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, NY, USA.
Antoine Chaillon - Division of Infectious Diseases and Global Public Health, University of California San Diego, CA, USA.
Sitharam Ramaswami, Paul Zappile, Christian Marier - Genome Technology Center, Office for Science and Research, NYU Langone Health, New York, NY, USA
Adriana Heguy, Ralf Duerr - Department of Pathology, NYU Grossman School of Medicine, New York, NY, USA.

Nature and purpose:

Phylogeographic investigation to elucidate the circulation of viral lineages during the first months of the New York outbreak – the epicenter of the COVID-19 epidemic in the United States.
Output in the last 12 months:
One paper was submitted to PloS Pathogens in 2020 and is currently under review

Future Direction:

This is an ongoing collaboration that will shift it's 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-CoV2.

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 submitted to the journal Cell in 2020 and is currently under review

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. Immune Correlates of Disease Progression in Linked HIV-1 Infection

Collaborating Parties:

Ralf Duerr Michael Tuen, Jude S. Bimela, Andrew N. Banin; Sonal Soni; Luzia Mayr; Aubin J. Nanfack; Mirosław K. Gorny and Phillippe N. Nyambi – New York University School of Medicine, New York, New York, USA. Jude S. Bimela; Judith N. Torimiro – University of Yaoundé, Yaoundé, Cameroon. Shilei Ding – Centre de Recherche du CHUM, and Université de Montréal, Montréal, QC, Canada. Vincenza Itri – Icahn School of Medicine at Mount Sinai, New York, USA.

Allison R. Durham – National Institutes of Health-National Institute of Allergy and Infectious Diseases, Bethesda, Maryland, USA.

Stephen F. Porcella – Rocky Mountain Laboratories, Division of Intramural Research, NIAID, NIH, Hamilton, Montana, USA. Josephine Meli – Yaoundé General Hospital, Yaoundé, Cameroon.

Marcel Tongo – Institute of Medical Research and Study of Medicinal Plants (IMPM), Yaoundé, Cameroon and Nelson R Mandela School of Medicine, University of KwaZulu-Natal, South Africa.

Xiaohong Wang – Veterans Affairs New York Harbor Healthcare Systems, New York, USA. Xiang-Peng Kong – Arthur Nadas New York University School of Medicine, New York, USA. Daniel E. Kaufmann – The Scripps Research Institute, La Jolla, California, USA.

Zabrina L. Brumme – Simon Fraser University, and British Columbia Centre for Excellence in HIV/AIDS, St. Paul's Hospital, Vancouver, Canada.

Thomas C. Quinn; Andrew D. Redd -Department of Medicine, Johns Hopkins University, Baltimore, USA.

Andrés Finzi – McGill University, Montréal, Canada.

Nature and purpose:

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

Output in the last 12 months:

One online book chapter was published in 2020

Future Direction:

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

5. 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, Cape Town, South Africa. Salim Abdool Karim – Centre for the AIDS Programme of Research in South Africa, University of KwaZulu-Natal, Durban, South Africa. Andrew Redd – Johns Hopkins University, School of Medicine in Baltimore, USA.

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.

6. Virulence Evolution

Collaborating Parties:

Philippe Roumagnac - CIRAD, Montpellier, France. Darren P. Martin, Adérito Luis Monjane, Dionne Natalie Shepherd – UCT, South Africa.

Philippe. Lemey and Simon Dellicour – Katholieke Universiteit, Leuven, Belgium.

Pierre Lefevre and Jean-Michel Lett – CIRAD, Reunion Island.

Arvind Varsani – Arizona State University, USA.

Nature and purpose:

We received funding from the National Research Foundation (NRF) to investigate how the traits of virulence and transmission rates are evolving in viral pathogens important to human health. Using a combination of laboratory experimentation and state-of-the-art computational analyses, we investigated the spatiotemporal dynamics and evolution of a broad range of viral pathogens focusing primarily on viruses that cause important crop diseases such as Tomato yellow leaf curl disease (TYLCD), Maize streak disease (MSD) and Cassava mosaic disease (CMD).

Output in the last 12 months:

One article published in the scientific journal *E-Life* in 2020

Future Direction:

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

7. Molecular Mechanisms of Viral Emergence

Collaborating Parties:

Pauline Bernardo, Sarah François, Maëlle Deshoux, Denis Filloux, Emmanuel Fernandez, Serge Galzi, Romain Ferdinand, Martine Granier, Cécile Richet, Michel Peterschmitt and Philippe Roumagnac – CIRAD-INRA-SupAgro, Montpellier, France.
Brejnev Muhire, Darren P. Martin – UCT, South Africa.

Simona Kraberger, Arvind Varsani – Arizona State University, USA.

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

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

Pablo Monge, Fernando Escru – Unidad de Sanidad Vegetal Zaragoza, Spain.

Nature and purpose:

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

Output in the last 12 months:

A manuscript was published in the *Archives of virology* in 2020.

Future Direction:

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

is related to other newly discovered species that cause severe infections in French beans in India and alfalfa in France. Three other geminiviruses viruses that we discovered infecting the uncultivated plant species *Limeum africanum*, *Exomis microphylla* and *Polygala garcinii* DC that group in the with begomoviruses, becurtoviruses and capulaviruses respectively were investigated by a MSc student at SANBI.

8. Endogenous Viral Elements (EVEs)

Collaborating Parties:

Philippe Roumagnac, Denis Filloux and Emmanuel Fernandez - CIRAD, Montpellier, France.

Darren Martin – UCT.

Véronique Jamilloux, Florian Maumus – INRA, URGI, Versailles, France.

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

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

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

Andrew Geering – the University of Queensland, Australia.

Nature and purpose:

To study endogenous viral elements (EVEs) and their role in virus evolution and their functions in plants to address fundamental biological issues that are critical to the understanding of plant virus evolution and the contribution of plant viruses to both the host genome evolution and biology. This involves approaches in virology, bio-informatics, genomics, proteomics, evolution and synthetic biology, and aims at disentangling key mechanisms of adaptive plasticity of plant and viral genomes such as gene domestication, genetic and epigenetic

regulation of plant gene expression and deciphering mechanisms that are central to molecular plant-virus interactions such as recombination, transcriptional or post transcriptional gene silencing.

Output in the last 12 months:

None

Future Direction:

We received funding to support this research from the French National Research Foundation (ANR) in December 2017 and subsequently begun generating data. Two annual meetings of all of the major stakeholders were held at CIRAD Réunion Island in September 2017 and Montpellier in 2018 to assess the progress to date and final meeting was scheduled for 2020 but has been rescheduled for 2021.

9. 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:

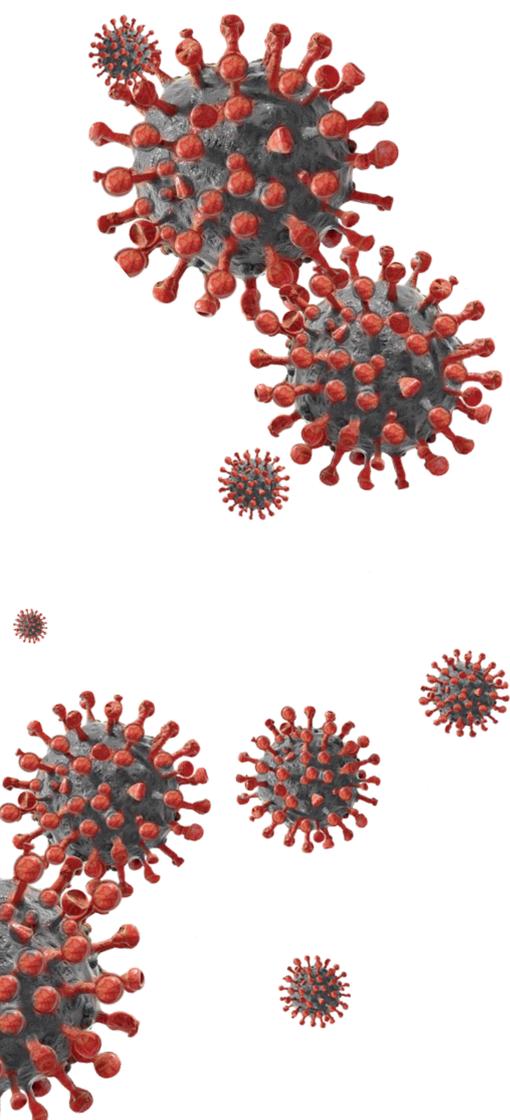
While it is well established that emerging viruses generally originate in uncultivated hosts with which they have well-established interactions, we currently know almost nothing about the diversity and the distribution of phytoviruses circulating in either natural or managed ecosystems. This project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

Output in the last 12 months:

The first year of funding for this project was 2019 and to date we have completed the plant sampling efforts in South Africa in and Arizona for 2019. Further sampling trips are planned for both countries in 2021 while efforts to detect phytoviruses in our plant samples are currently being performed in the laboratory of Philippe Roumagnac in Montpellier France.

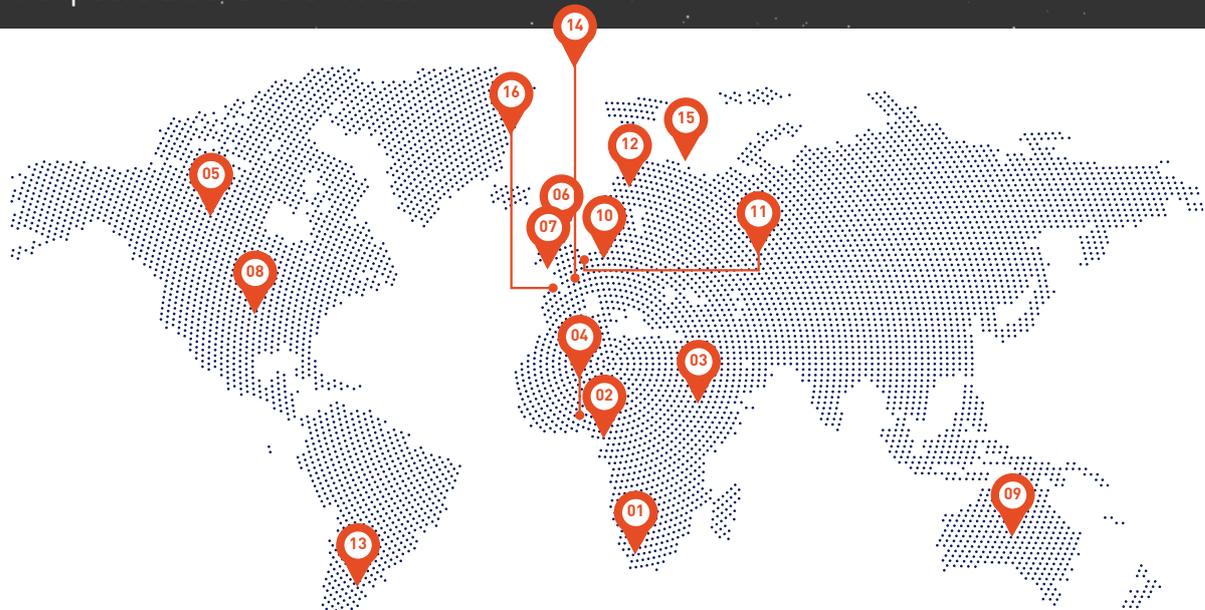
Future Direction:

This is an ongoing project that is funded until 2023.



COLLABORATIONS

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



01 SOUTH AFRICA

University of the Western Cape
University of Cape Town
University of Stellenbosch
NHLS/Stellenbosch University Biorepository,
Tygerberg Hospital
National Institute for Communicable Diseases
(NICD)
Centre for the AIDS Programmed of Research
in South Africa (CAPRISA)
H3Africa
KwaZulu-Natal Research Innovation and
Sequencing Platform (KRISP), School of
Laboratory Medicine & Medical Sciences,
University of KwaZulu- Natal
Africa Health Research Institute, KwaZulu-Natal
Nelson R Mandela School of Medicine, University
of KwaZulu-Natal

02 CAMEROON

University of Yaoundé, Yaoundé
Institute of Medical Research and Study
of Medicinal Plants (IMPM), Yaoundé

03 ETHIOPIA

Africa CDC

04 NIGERIA

Global Consortium on Emerging Infectious
Diseases
Department of Biochemistry, Lagos State
University, Lagos
Lagos Biobank

05 CANADA

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

06 ENGLAND

School of Law, Middlesex University
Oxford University
University of Birmingham
COVID-19 Genomics UK (COG-UK)

Clinical Microbiology, University of Cambridge,
Cambridge

07 SCOTLAND

MRC-University of Glasgow Centre for Virus
Research, Scotland

08 UNITED STATES

USA CDC
Broad Institute
Pacific North West Research Institute, Seattle
University of Washington, Seattle
Cleveland Clinic Lerner College of Medicine
Genome Technology Center, Office for Science
and Research, NYU Langone Health, New York
Department of Pathology, NYU Grossman School
of Medicine, New York
Institute for Systems Genetics, NYU Grossman
School of Medicine, New York
Genome Technology Center, Office for Science
and Research, NYU Langone Health, New York
Veterans Affairs New York Harbor Healthcare
Systems, New York
Arthur Nadas New York University School
of Medicine, New York
Icahn School of Medicine at Mount Sinai, New York
Institute for Genomics and Evolutionary Medicine,
Department of Biology, Temple University,
Pennsylvania
Department of Biochemistry and Molecular Biology,
The Pennsylvania State University, Pennsylvania
Department of Medicine, Johns Hopkins School
of Medicine, Baltimore, Maryland
National Institutes of Health-National Institute
of Allergy and Infectious Diseases, Bethesda,
Maryland
Rocky Mountain Laboratories, Division of
Intramural Research, NIAID, NIH, Hamilton,
Montana
Division of Infectious Diseases and Global Public
Health, University of California San Diego
The Scripps Research Institute, La Jolla, California
Department of Medicine, University of California
San Diego, La Jolla
Arizona State University

09 AUSTRALIA

University of Melbourne
University of Queensland

10 SWITZERLAND

World Health Organisation

11 BELGIUM

Spatial Epidemiology Lab. (SpELL), Université
Libre de Bruxelles, Bruxelles, Belgium;
Department of Microbiology, Immunology and
Transplantation, Rega Institute, Leuven
Katholieke Universiteit

12 SWEDEN

Department of Microbiology, Tumor and Cell
Biology, Karolinska Institutet, Stockholm.

13 URUGUAY

Departamento de Desarrollo Biotecnológico,
Instituto de Higiene, Facultad de Medicina,
Udelar, Montevideo
Departamento de Biodiversidad y Genética.
Instituto de Investigaciones Biológicas Clemente
Estable, Montevideo
Laboratorio de Biología Molecular, Asociación
Española Primera en Salud, Montevideo

14 FRANCE

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

15 FINLAND

University of Helsinki, Finland.

16 SPAIN

Unidad de Sanidad Vegetal Zaragoza

FINANCIALS

SANBI's income and expenditure trends for 2020 are shown in this section.

Figure 1. Distribution of income received from all sources 2020.

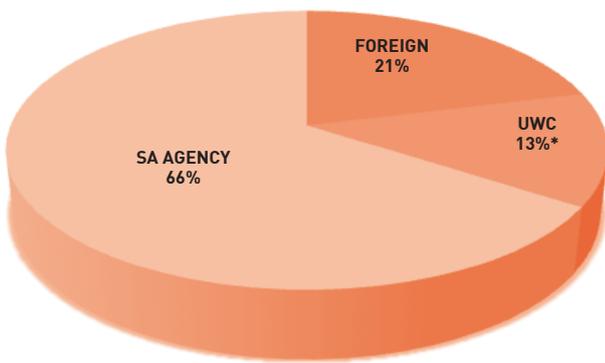
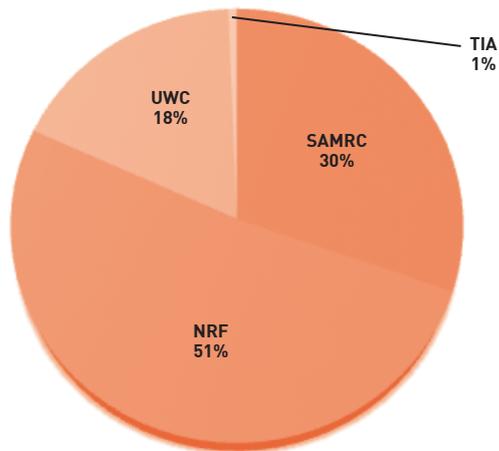
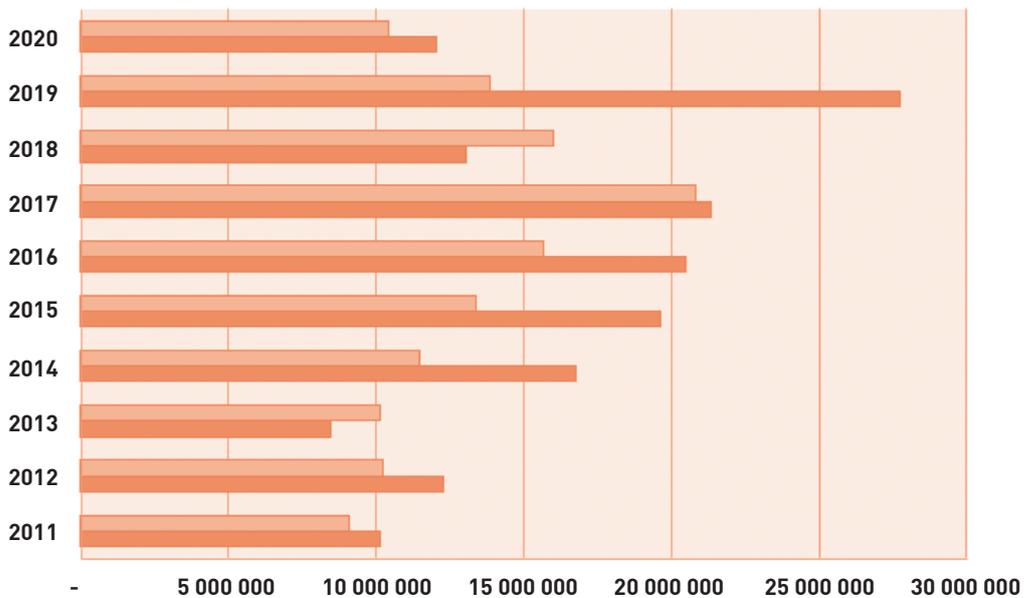


Figure 2. Income received from SA sources 2020.



* excluding SARChI Chairholder

Figure 3. Income vs. Expenditure 2010 – 2020.



ZAR	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
EXPENSES	9 101 523	10 233 547	10 149 964	11 482 581	13 393 931	15 665 001	20 844 456	15 981 456	13 849 462	10 406 504
INCOME	10 138 202	12 266 118	8 448 034	16 790 442	19 641 569	20 495 324	21 335 183	13 043 930	27 721 406	12 045 007

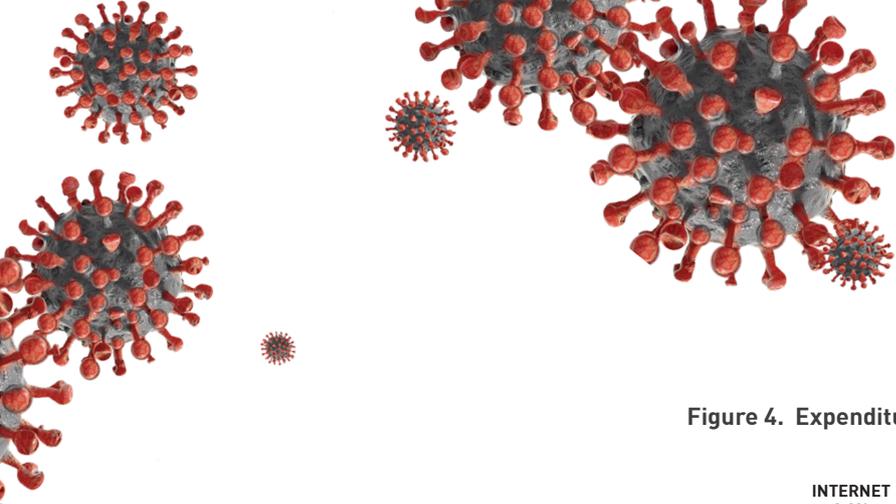
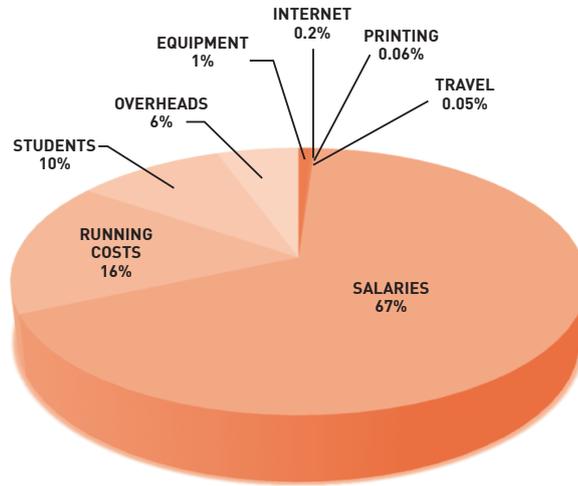


Figure 4. Expenditure in 2020.



ALUMNI

SANBI has produced many alumni 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
- Univeristy of South Africa
- 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
- Hyrax Biosciences
- IBM
- ROCHE
- B&M Scientific
- GENE diagnostics
- City of Cape Town
- World Bank
- SA Government Administration
- KappaBiosystems
- Takealot

KENYA

- Egerton University, Njoro, Kenya
- International Livestock Research Institute, Kenya

GHANA

- University of Ghana

NIGERIA

- Lagos State University
- University of Nairobi

UNITED STATES

- Harvard Stem Cell Institute
- Princeton University
- Washington University
- Jackson State University
- University of California, San Diego
- Wayne State University, Detroit, Michigan
- Pacific Northwest Diabetes Research Institute, Washington
- Beth Israel Deaconess Medical Centre, Harvard Medical School
- Johns Hopkins University

CANADA

- University of Western Ontario, Canada
- University of British Columbia

SAUDI ARABIA

- King Abdullah University of Science and Technology

IRELAND

- National University of Ireland, Galway

SWITZERLAND

- Syngenta Crop Protein AG, Basel, Switzerland

RUSSIA

- Institute of Cytology and Genetics, Novosibirsk

SINGAPORE

- Nanyang Technology University
- University of Singapore

MALAYSIA

- Novocraft Technologies

FUNDERS



Research Institute at UWC
since 1997



National Research Foundation
funding since 1998



South African Medical Research Council
Bioinformatics Unit since 2002



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



H3ABioNet Node 2020-2022



BILL & MELINDA
GATES *foundation*

Bill & Melinda Gates Foundation



Technology Innovation Agency funding since 2016



Poliomyelitis Research Foundation
funding since 2019



Agriprotein funding
since 2019



POSTAL ADDRESS

South African National Bioinformatics Institute
University of the Western Cape
Private Bag X17
Bellville
7535

PHYSICAL ADDRESS:

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

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

