SANBI
South African National
Bioinformatics Institute

ANNUAL REPORT 2022

UNIVERSITY of the
WESTERN CAPE

CONTENTS

1 CONTENTS

2 ABOUT SANBI

4 DIRECTOR'S MESSAGE

5 YEAR IN REVIEW

7 2022 IN SUMMARY

8 STAFF

14 CAPACITY DEVELOPMENT

16 SANBI STUDENTS 2022

17 WORKSHOPS AND COURSES

21 COMPUTATIONAL RESOURCES

24 RESEARCH OUTPUTS

32 RESEARCH LABORATORIES
  - 32 ANDERSON
  - 34 BENDOU
  - 36 CHRISTOFFELS
  - 38 CLOETE
  - 40 HARKINS
  - 45 HESSE
  - 46 TIFFIN

50 COLLABORATIONS AND ALUMNI

52 FINANCIALS

53 FUNDERS

ABOUT SANBI

Bioinformatics is a specialist discipline straddling the fields of biology, mathematics and computer sciences and it is integral to modern biological research.

WHO WE ARE

The South African National Bioinformatics Institute (SANBI) is situated at the University of the Western Cape (UWC) in Cape Town.

Our primary focus is the development and implementation of computational methodologies that allow biomedical researchers to accelerate their genomics data analyses. SANBI aims to heighten awareness of bioinformatics in South Africa and to assist the country in making optimal use of bioinformatics tools. As the leading bioinformatics entity in Africa, we continue to foster local and regional collaborations on health-related topics that cover both communicable and non- communicable diseases.

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

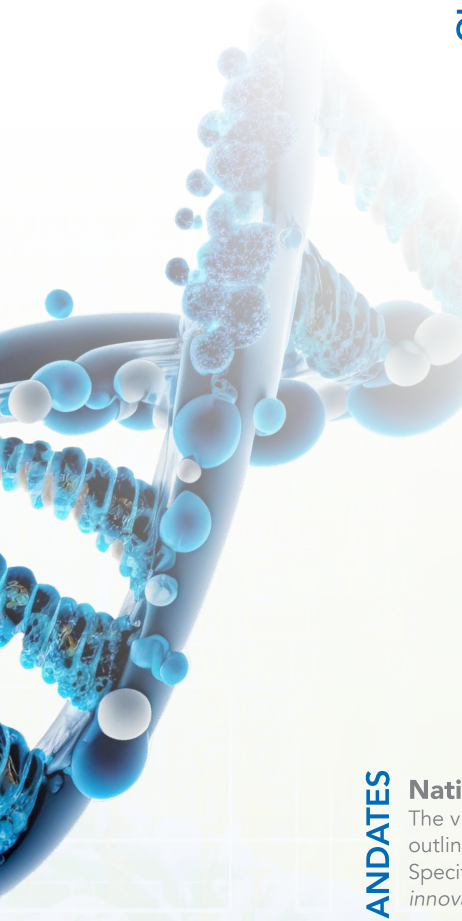
- the development of online specialised resources for genomics and genome informatics;
- capacity development in genomics and bioinformatics in Africa; and
- the development and implementation of genome analysis and interpretation methods.

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 (NRF) Strategy 2025

The vision and mission of SANBI aligns with the NRF's Strategy 2025. Specifically, *"promoting globally and competitive research and innovation"*.

The research and development at SANBI aligns with the Department of Science and Innovation White Paper on Science, Technology and Innovation 2019 by contributing to *"the development of human capabilities, knowledge expansion and innovation performance..."*.

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

As an extramural unit of the SAMRC, SANBI falls under the legislative mandates of the SAMRC. 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

2022 has seen a return to university for students and staff as we moved all our activities to in-person interaction. The shift away from virtual meetings has had a positive impact on student wellbeing and allowed us to continue our training activities in support of local and continental initiatives.

The annual report for 2022 demonstrates the cross border collaborations of our staff and the international engagements in different academic spaces. We are particularly excited by the 31 publications for 2022, many of these underpinning our efforts at translational research.

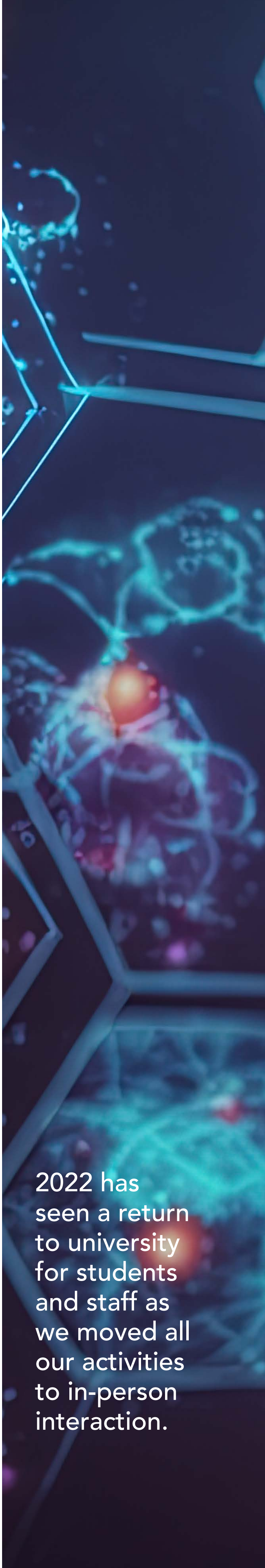
The past year has marked the end of a 15-year Department of Science and Innovation/South African National Research Foundation (DSI/NRF) Research Chair in Bioinformatics and Human Health. The investment by the NRF over the 15 years has provided critical investment into human capital development and catalysed numerous innovations. We have leveraged this investment to secure new funding streams for 2023 and beyond.

We look forward to a productive 2023 as we place community centeredness at the core of our mandate.

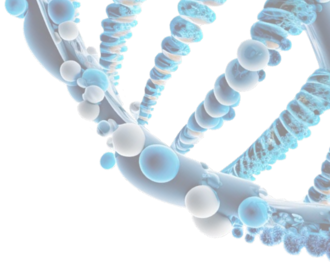
Professor Alan Christoffels

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

South African National Bioinformatics Institute
University of the Western Cape



2022 has seen a return to university for students and staff as we moved all our activities to in-person interaction.



YEAR IN REVIEW

2022 has been a demonstration of the international prestige that our staff and students enjoy. Scientific excellence has permeated every activity at SANBI as demonstrated by:

- Publications: 29 journal articles and 2 book chapters.
- Hub for international training workshops.
- Conference participation by staff and students including keynote presentations.
- The prestigious Fulbright Fellowship awarded to Ruben Cloete that enabled him to spend his sabbatical in the USA.

RESEARCH HIGHLIGHTS:

Africa CDC

The Africa CDC is a technical institution of the African Union strengthening the capacity of member states to respond quickly and effectively to disease threats. In September 2022, an Africa CDC team led by senior advisor Alan Christoffels and Sarah Mwangi (ex-SANBI student and current Africa CDC specialist scientist) met with Prof Maaroufi and Dr Barakat of Institut Pasteur du Maroc (IPM) in Casablanca, Morocco to establish terms of reference for IPM and its role to support regional genomic surveillance in Africa.

Asia Pathogen Genomics Initiative

A team from the Asia Pathogen Genomics Initiative based in Singapore visited SANBI in October 2022 to learn from our experience in setting up and leading the Public Health Alliance for Genomic Epidemiology (PHA4GE), global coalition whose secretariat is based at SANBI.

Beyond COVID-19: Pathogen Genomics and Bioinformatics for Health Security in Africa symposium

At least 140 public health officials, experts and researchers from Member States and international partners attended this event in November 2022 at the African Union headquarters in Addis Ababa. Alan Christoffels led the technical team who coordinated this event.

Alan Christoffels presenting at the Africa PGI Symposium in November 2022. *"Africa moved from being data consumers to data producers, thanks to exciting developments during Covid-19 in next generation sequencing capacity on the continent in past 2 years."*

COVID-19 Related Research

A range of projects related to COVID-19 culminated in high impact publications including:

- A pan-African study tracing the evolution of SARS-COV-2 epidemic in Africa.
- Another study with the focus on molecular modelling allowed Ruben Cloete and his group to demonstrate the flexible, extended coil of the PDZ-Binding Motif of three deadly human coronavirus E proteins and their role in pathogenicity.
- Gordon Harkins' work has focused on conserved recombination patterns across coronavirus subgenera and the unusual mutational changes in Omicron Lineage BA.1 that impacts Spike protein function.

PHA4GE Ethics and Data Sharing Working Group Workshop, Mauritius

Nicki Tiffin, Anja Bedeker and Jamie Southgate organised and attended the workshop from 14 – 17 November in Mauritius. This workshop provided training and support to recipients of PHA4GE Ethics and Data Sharing sub-awards which were awarded to four teams conducting research into ethics and data governance challenges in under-resourced or challenging research environments. Participants hailed from Botswana, South Africa, Ethiopia and Nepal.

Workshop participants – Team Ethiopia (Bizuneh Wakuma, Eba Abdisa, Gemechu Tiruneh), Team Botswana (Kagiso Ndlovu, Julia Gaorekwe, Tumalano Sekoto), Team South Africa (Jihaan Adonis), Team Nepal (Minu Singh) & PHA4GE Mentors (Victoria Nembaware, Nicki Tiffin, Anja Bedeker & Jamie Southgate).

Ethical Benefit Sharing in Health Research

SANBI staff led by Nicki Tiffin published a ground breaking framework for ethical benefiting sharing in health research. This work has provided the basis for further discussion on principles of data sharing (doi:10.1136/bmjgh-2021-008096).

Research Software and Systems Engineering Africa (RSSE Africa)

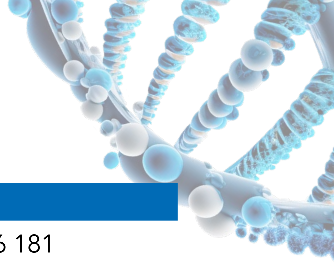
RSSE Africa was initiated by Peter van Heusden with support from Talarify (a startup founded by ex-SANBI student Anelda van der Walt that supports researchers who want to develop digital and computational skills). The events carried out in 2022, was aimed at stimulating the growth of an African community of practice around research software and system engineering practices.

Bioinformatics for Oncology

Hocine Bendou's team showed the role of FHL1, SPNS3, and MPZL2 as predictor of poor prognosis in pediatric acute myeloid leukemia patients with FLT3-ITD mutation.

Technology Development

- Alan Christoffels and Peter van Heusden concluded a 7 year project to build an analytics platform for studying M.tuberculosis genomics data.
- Peter van Heusden has been instrumental in setting up MinION sequencing infrastructure at UWC where the application has been demonstrated for the Rooibos *(Aspalathus linearis)* Genome.



2022 IN SUMMARY

INCOME
ZAR 37 586 181

EXPENDITURE
ZAR 24 849 024

DIVERSITY PROFILE

21 STAFF
11 MALE
10 FEMALE

19 NATIONALITY SOUTH AFRICAN

EDUCATION LEVEL

8	4	2	7
POSTDOCTORAL	MSC	BSC	DIP

38 STUDENTS
17 MALE
21 FEMALE

23 NATIONALITY SOUTH AFRICAN

EDUCATION LEVEL

4	19	15
POSTDOCTORAL	PHD	MSC

RESEARCH OUTPUTS

JOURNALS	BOOK CHAPTER	GRADUATIONS	AWARDS/HONOURS
29	2	3	2

KEYNOTES/ INVITED TALKS	WORKSHOPS & SEMINARS	CONFERENCE PRESENTATIONS	COURSE MODULE	EXPERT PANEL/ COMMITTEE MEMBERSHIP
12	5	7	1	19

46 RESEARCH PROJECTS WITHIN 25 RESEARCH THEMES

58 COLLABORATIONS ACROSS 13 LOCATIONS

ALUMNI IN 51 INSTITUTIONS AND CORPORATES

STAFF

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

As SANBI is a research institute, the Director reports through the faculty of Natural Sciences at UWC. SANBI academic staff are tasked with conducting research in their laboratories, securing funding, student graduations and capacity development.

Dominique Anderson, Dr
Senior Lecturer

Funder UWC

Anja Bedeker
Research Associate

Funder B&MGF

Hocine Bendou, Dr
Senior Researcher

Funder MRC

Alan Christoffels, Prof
Director,
SARChI Bioinformatics,
SA MRC Unit Director

Funder DSI/NRF Research
Chairs Programme

Ruben Cloete, Dr
Senior Lecturer

Funder UWC

Gordon Harkins, Dr
Associate Professor

Funder UWC

Judit Kumuthini, Dr
Senior Researcher

Funder H3A BioNetwork

Nicki Tiffin, Prof
Professor

Funder UWC

Technical staff develop and maintain the computing infrastructure at SANBI.

Quinton Coert
Software Developer

Funder Baobab LIMS

Zipho Mashologu
Software Developer

Funder SAMRC

Campbell Rae
Web Developer
(part time)

Funder DSI/NRF Research Chairs Programme

Peter van Heusden
Senior Systems Developer

Funder DSI/NRF Research Chairs Programme

Shadley Wentzel
Senior Systems Administrator

Funder SAMRC

SANBI administrators ensure the smooth running of daily operations.

Fungiwe Mpithi
PA/Administrator

Funder SAMRC

Ferial Mullins
Finance Administrator

Funder UWC

Junita Williams
Marketing Administrator
(part time)

Funder DSI/NRF Research Chairs Programme

The PHA4GE Secretariat hosted at SANBI provide administrative and project management support to the PHA4GE working groups.

Based at SANBI, Alan Christoffels leads the Secretariat of the Public Health Alliance for Genomic Epidemiology (PHA4GE), an organisation comprising more than 100 members based at academic institutions around the globe. This international network has worked hard to develop analytical tools and ethics frameworks to support public health interventions.

Michael Bridger
Business Development

Funder B&MGF

Rangarirai Matima
Communication Specialist

Funder B&MGF

Alecia Naidu, Dr
Technical Manager

Funder B&MGF

Jamie Southgate
Communications Officer

Funder B&MGF

Nawaal Nacerodien-Weitz
Administrator

Funder B&MGF

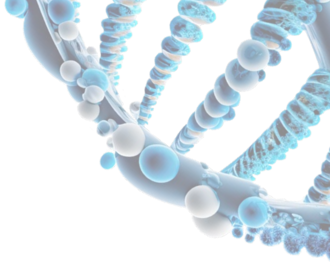
AWARDS AND HONOURS

Fulbright Fellow (2022-2023) Ruben Cloete

SANBI senior lecturer, Ruben Cloete, was awarded the prestigious Fulbright Fellowship to attend the University of California San Francisco (UCSF) from 12 September 2022 until 17 February 2023. During his 5-month sabbatical visit at the University of California San Francisco he established new collaborations and networks between SANBI, UWC and Prof Andrej Sali (Host), Prof Brain Shoichet and Prof John Irwin. His primary goal was to learn new methods in cell modelling namely integrative modelling and drug design in the laboratories of Prof Andrej Sali and Prof Brain Shoichet. The skill set and experience he acquired during his visit will be applied to existing projects such as Tuberculosis and SARS-Cov2 coronavirus drug discovery as well as to train students in the use of these methods towards completion of their degrees.

Royal Society of South Africa

In October, Alan Christoffels was elected a fellow of the Royal Society of SA, becoming one of only 3 RSSAf from UWC. Asked what this honour means to him, Alan responded: *"The Society gives you credibility and recognition. And I think that in turn there is a responsibility on the recipient's side to ensure that they continue to pursue scientific excellence in what they do."*



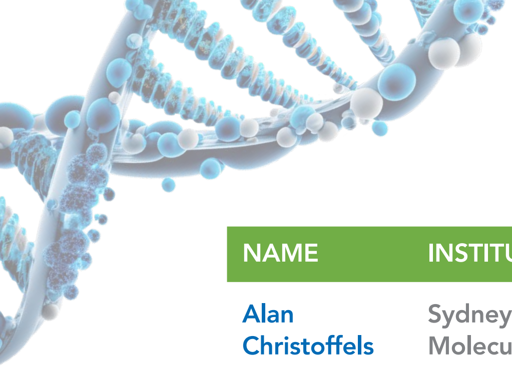
Grants/Funding Awarded

NAME	FUNDER DETAILS	AMOUNT	PROJECT SCOPE
Anja Bedeker	**BMGF (sub award by PHA4GE consortium)**	$5 000	Design and assessment of a research ethics short course for postgraduate students
Nicki Tiffin	**BMGF Calestous Juma Science Leadership Fellowship**	$968 546 (2021-2025)	The African Data and Biospecimen Exchange – facilitating equitable data and biospecimen sharing, and resolving barriers to data and biospecimen sharing in africa
	UKRI/MRC Seed funding grant	£201 523	The Khayelitsha Comorbidity Cohort: Establishing a multimorbidity cohort with integrated clinical, genomic and epidemiological data in South Africa
Peter van Heusden	**BMGF**	$97 795	Making pathogen sequencing accessible for meningitis response in the DRC

SANBI academic staff are actively involved in translating research into policy. In 2022, more than half of the academic staff were appointed on national and international committees and expert panels that develop research frameworks to accelerate data science and biomedical research.

Expert Panel, Committee Membership and Conference Review Panels

NAME	INSTITUTE	ROLE AND PURPOSE
Dominique Anderson	**Academy of Science of South Africa**	Standing Committee on POPIA – to guide the development of the POPIA Code of Conduct for Research
	National Scientific R Collections Platform	Medical Biobanks Working Group Co-Chair
Anja Bedeker	**PHA4GE Consortium**	Member of the Steering Committee and Co-Chair of the Ethics and Data Sharing Working Group
Alan Christoffels	**African Society for Bioinformatics and Computational Biology**	Member of the Governing Council
	Academy of Science of South Africa	Member of the Standing Committee on Biosafety and Biosecurity and Standing Committee on POPIA
	NHLS	Member of the Biospecimen and Data Access Ethics Committee

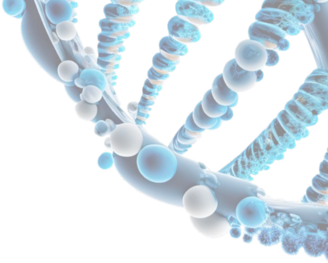


NAME	INSTITUTE	ROLE AND PURPOSE
Alan Christoffels	**Sydney Brenner Institute for Molecular Bioscience, Wits**	Research Advisory Committee
	Eastern Africa Network of Bioinformatics Training	Independent Scientific Advisory Board
	Global Emerging Pathogens Consortium	Member – host an annual conference on emerging pathogens as part of our mandate to create awareness of biosecurity threats across Africa, and to empower African scientists to respond to disease outbreaks
Ruben Cloete	**South African Society for Bioinformatics**	Secretary
	African Society for Bioinformatics and Computational Biology	Program Committee
Gordon Harkins	**Centre for the AIDS Programme of Research in South Africa (CAPRISA)**	Director of the Centre of Excellence in HIV Prevention at UWC
Nicki Tiffin	**University of Cape Town, Computational Biology Division**	Honorary Professor
	International Common Disease Alliance	Member – Executive Committee; Co-Lead – Global Equity Working Group
	International 100k+ Cohorts Consortium	Member – Executive Committee and Steer-ing Committee; Co-Chair – Policy and Data Sharing Working Group
	African Population Cohorts Consortium	Independent Advisory Group
	PHA4GE Consortium	Member – Steering Committee; Chair – Ethics and Data Sharing Working Group
	Global Health EDCTP3 Joint Undertaking	Member of the Scientific Committee
Peter van Heusden	**PHA4GE Consortium**	Infrastructure Working Group

SANBI IN THE MEDIA

Some coronaviruses kill, while others cause a common cold. We are getting closer to knowing why
In: The Conversation
Published: 3 October 2022
By: Dewald Schoeman, Burtram C. Fielding and Ruben Cloete.
https://theconversation.com/some-coronaviruses-kill-while-others-cause-a-common-cold-we-are-getting-closer-to-knowing-why-190690

The work by Anja Bedeker and Nicki Tiffin has been referenced in the media.
Why South Africa needs a national Data Transfer Agreement for health research
In: Daily Maverick
Published: 19 June 2022
By: Mahomed et al.
https://www.dailymaverick.co.za/article/2022-06-19-why-south-africa-needs-a-national-data-transferagreement- for-health-research/



Journal Editing and Reviews

NAME	JOURNAL
Hocine Bendou	Reviewer – Cancer Biomarkers
Alan Christoffels	Editor – Improper Scientist Editorial Board Member – Data Journal Reviewer – BMC Evolutionary Biology; BMC Genomics; BMC Bioinformatics; Bioinformatics; Journal of Parallel and Distributed Computing; PLoS ONE; Trends in Genetics; Genome Research.
Ruben Cloete	Reviewer – Heliyon
Nicki Tiffin	Journal Associate Editor – Nucleic Acids Research Genomics and Bioinformatics Reviewer – Health Policy and Technology Journal, International Journal of Population Genomics, BMJ Global Health
Peter van Heusden	Reviewer – Microbial Genomics

SANBI staff are actively involved in supporting other institutions of learning.

Thesis Examined for Students at Other Institutions

EXAMINER	INSTITUTION	DEGREE
Dominique Anderson	**University of Stellenbosch**	MSc
Alan Christoffels	**University of KwaZulu Natal**	MSc

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

Further Development of Staff

NAME	INSTITUTION	DATE	COURSE NAME + PURPOSE OF COURSE
Hocine Bendou, Ruben Cloete	**UWC**	15 May	Workshop on Professionalisation of Teaching and Learning
Nicki Tiffin	**ASSAf ELSI Gene Therapy Panel**	17 - 18 October	Writing workshop for: Ethical, Legal and Social Issues Related to Gene Therapy: A South African Perspective: Consensus Study

CAPACITY DEVELOPMENT

SANBI offers training programmes which are in keeping with its vision of becoming a centre 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 third-year Biotechnology students. The SANBI faculty, Ruben Cloete, Gordon Harkins, Hocine Bendou and Alan Christoffels taught 92 students various Bioinformatics topics such as Phylogenetics, Comparative Genomics, Biodatabases, Blast and Sequence Alignments, etc.

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.

Internship Programme

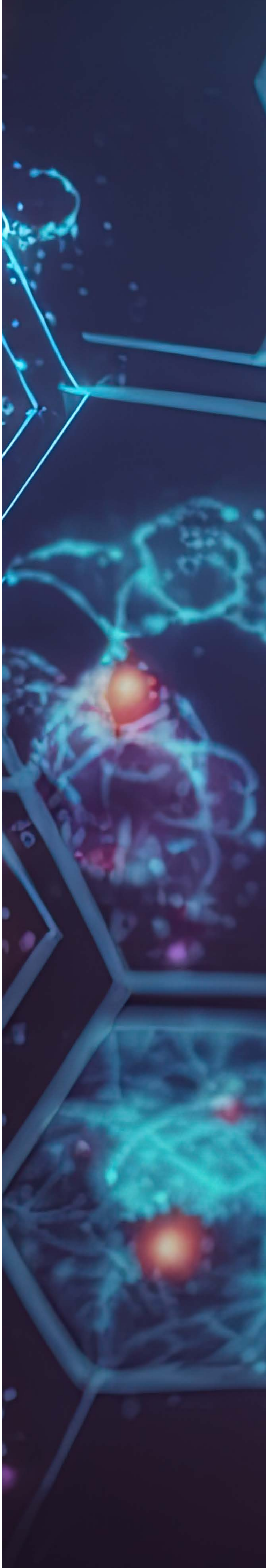
As part of the DST/NRF Research Chairs Programme, SANBI offers a 1-year internship programme to students who have graduated with a BSc degree.

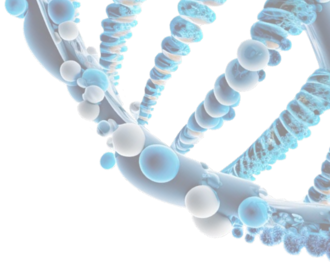
POSTGRADUATE TRAINING PROGRAMME

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

Bioinformatics Training Course (BITC)

For this annual postgraduate Bioinformatics training course there were 9 participants who attended both face to face meetings and virtual pre-recorded lecturers. The topics covered included UNIX Command Line Language, Python Programming, How to use a Cluster, Using the Galaxy Pipeline and Introduction to Statistics using R. Study Design was also incorporated into the course. Students were given assignments to assess whether they were able to understand the concepts and apply the knowledge to their research project.





Paper And Thesis Writing Workshops

Started in February 2022 and run by Nicki Tiffin, the aim of this series of weekly SANBI workshops has been to develop students' abilities and skills in paper writing and offer tips on how to approach thesis writing, while the informative and practical grant writing sessions aimed to increase the likelihood of success when applying for grants.

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.

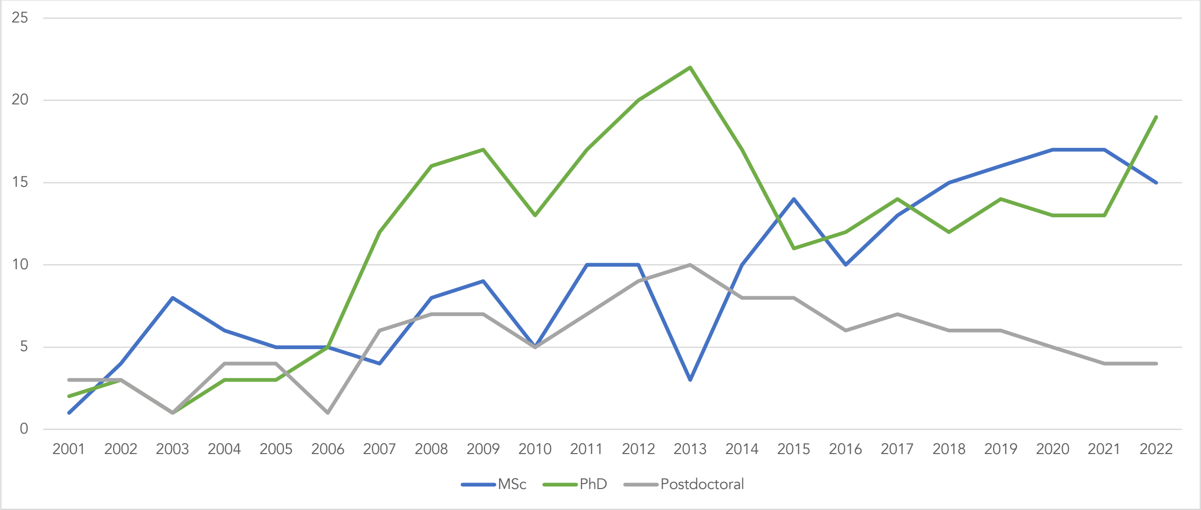
Doctoral Programme

Candidates should be in possession of a MSc 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 2001-2022



SANBI ALUMNI NEWS

SANBI is proud to celebrate the achievement of alumnus, Janet Kelso, who joined Prof Svante Paabo's group after her PhD in Bioinformatics in 2003 and subsequent Postdoctoral Fellowship at SANBI in 2004. She has worked on key pieces of research in his lab which led to Prof Paabo being awarded the 2022 Nobel Prize in Physiology or Medicine for his discoveries concerning the genomes of extinct hominins and human evolution.
(https://www.nobelprize.org/prizes/medicine/2022/press-release/)

SANBI STUDENTS 2022

The SANBI student complement comprises a cohort of diverse and multi-talented researchers. In total there are 4 Postdoctoral Fellows, 19 PhD and 15 MSc students as well as 1 Honours student supervised by the SANBI academic staff.

POSTDOCTORAL FELLOWS

Lulamba Eddie **Primary Supervisor** Tiffin
Olagbegi Oladapo **Primary Supervisor** Christoffels
Rossouw Catherine **Primary Supervisor** Bendou
Williams Wesley **Primary Supervisor** Christoffels

DOCTORAL STUDENTS

Abiodun Peter **Primary Supervisor** Anderson
Akinnuwesi Adetutu **Primary Supervisor** Egieyeh
Babajide Abiola **Primary Supervisor** Anderson
Calvert-JoshuaTracy **Primary Supervisor** Christoffels
DeRaedt Sarah **Primary Supervisor** Christoffels
Eshibona Nasr **Primary Supervisor** Bendou
Esterhuysen Faneschka **Primary Supervisor** Hesse
Giwa Abdulazeez **Primary Supervisor** Bendou
Hassan Mohammed **Primary Supervisor** Cloete
Juma John **Primary Supervisor** Christoffels
Langa Bridget **Primary Supervisor** Gamieldien
Livesey Michelle **Primary Supervisor** Bendou
Martin Darius **Primary Supervisor** Cloete
Mgwatyu Yamkela **Primary Supervisor** Hesse
Nkaule Anati **Primary Supervisor** Christoffels
Odugbemi Adeshina **Primary Supervisor** Egieyeh
Patten Victoria **Primary Supervisor** Bendou
Pillay Nikkita **Primary Supervisor** Christoffels
Tamuhla Tsaone **Primary Supervisor** Tiffin

MASTERS STUDENTS

Barendse Cleo-Rose **Primary Supervisor** Bendou
Brown Keagan **Primary Supervisor** Cloete
Diedricks Farzana **Primary Supervisor** Christoffels
Fernol Susan A **Primary Supervisor** Christoffels
Fredericks Ridaa **Primary Supervisor** Christoffels
Isaac Darren **Primary Supervisor** Cloete
Jassiem Wardah **Primary Supervisor** Bendou
Majuru Tatenda **Primary Supervisor** Anderson
Mazitshana Siphosetu **Primary Supervisor** Harkins
Mfuphi Nomlindelo **Primary Supervisor** Bendou
Ndlondlo Lunathi **Primary Supervisor** Bendou
Nyembezi Nkosazana **Primary Supervisor** Harkins
Omotoso Christianah **Primary Supervisor** Egieyeh
Serage Rudolf **Primary Supervisor** Anderson
van Heusden Peter **Primary Supervisor** Christoffels

HONOURS STUDENT

Adam E BSc Honours (MBS Department)
Co-supervisors Peter van Heusden, Megan Shaw

2022 SANBI GRADUATIONS

SANBI has great pleasure in presenting the 2022 graduants:

NOMXOLISI ANATI NKAULE, PhD Bioinformatics
Supervisor: PROF A CHRISTOFFELS
Thesis: Integrated Bioinformatics and Chemoinformatics Approach for identifying Bioactive compounds for *Mycobacterium Tuberculosis*

MFUPHI NOMLINDELO WITNESS, MSc Bioinformatics
Supervisor: DR H BENDOU, Co-supervisor: PROF A CHRISTOFFELS
Thesis: Reconstruction of gene regulatory networks of inflammation-associated genes in different clinical stages of diffuse large B-cell lymphoma

SERAGE RUDOLPH ABEL, MSc Bioinformatics *Cum Laude*
Supervisor: DR D ANDERSON, Co-supervisor: PROF A CHRISTOFFELS
Thesis: A deep learning approach to predicting potential virus species crossover using convolutional neural networks and viral protein sequence patterns

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. SANBI faculty are actively involved in teaching various topics.

WORKSHOPS ORGANISED/TAUGHT BY SANBI

The H3ABioNet Polygenic Risk Scores Workshop

Nicki Tiffin and Tsaone Tamuhla were organisers of, and attended the H3ABioNet Polygenic Risk Scores (PRS) workshop in Mauritius. We are members of the H3ABioNet Genetic Epidemiology working group, which I was chairing, and we led the application for funding to hold the PRS workshop on behalf of H3ABioNet. International experts in algorithms to calculate polygenic risk scores, Drs Paul O'Reilly and Conrad Iyegbe of Icahn School of Medicine at Mount Sinai, New York, joined the workshop as trainers to provide an excellent learning and hands-on experience for all participants.

Through our ongoing work with Dr Alice Matimba at Wellcome Trust Connecting Science, we were able to promote this course to be adopted and repeated with Wellcome Trust Connecting Science support, and Tsaone will act as a facilitator at the next PRS workshop to be held in Entebbe in 2023 under this new collaborative arrangement. This collaboration will ensure that the H3ABioNet workshop we held is just the first of many PRS workshops, with the sustainability offered through Wellcome Connecting Science support.

Participants in the H3ABioNet Polygenic Risk Scores workshop, University of Mauritius, May 2022.

Contribution to the Development of a Masters in Health Data Analytics

Nicki Tiffin participated in the development of a Masters in Health Data Analytics in an ongoing collaboration between UWC, Neu Ulm University of Applied Sciences in Germany, Kenya Methodist University, Muhimbili University of Health and Applied Sciences. She attended a workshop in Neu Ulm from 5 to 9 September 2022, with all partners, to complete the curriculum and course content design for this Masters. Nicki co-led the development of the Health Data module micro-curriculum with Dr Felix Sukums from Muhimbili University, and developed the content for the Omics Data unit,and the Epidemiology and Biostatistics unit. Other team members from UWC include Profs Anthea Rhoda, Jo-Celene de Jongh, James Kariuki, Simone Titus, Osden Jokonya and Verona Matthews.

MHDA Curriculum Development Workshop participants.

Nicki Tiffin's Open Access online teaching materials are available here under a Creative Commons license: they can be reused with attribution and without modification:

tinyurl.com/TIFFIN-Teaching-Material

Regional Pathogen Genomics and Data Management Training

Africa CDC aims to strengthen capacities and capabilities at public health institutions in Africa in order to detect and respond quickly and effectively to disease threats and outbreaks.

May 2022
Hosted by SANBI in collaboration with the Africa CDC and the Africa PGI a week-long course was held on SARS-CoV-2 sequence analysis. The attendees were from public health labs in 9 African countries – Morocco (Institut Pasteur du Maroc), Egypt (Central Public Health Laboratory), Ethiopia (Ethiopian Public Health Institute), Uganda (Center Public Health Institute), Kenya (National Public Health Laboratory), Senegal (Institut Pasteur de Dakar), Zambia (Zambian National Public Health Institute), Ghana (Noguchi Memorial Institute for Medical Research) and South Africa (National Institute for Communicable Diseases). Nigeria (Nigeria CDC) and DRC (Institut National de Recherche Biomédicale) participated via online streaming.

While the training focused on SARS-CoV-2, the discussions ranged across a number of other infectious diseases that these public health labs are responding to: HIV, TB, hepatitis, malaria, influenza and other pathogens.

Course participants from different African public health labs who attended the course.

October 2022
Following on from the training held earlier in the year, 16 delegates from 15 African countries attended the workshop entitled, 'Specialised Training Workshop on Bioinformatics Analysis for Pathogen Genomic Surveillance' directed at those working in African public health services.

The delegates were a host of lab technicians, data scientists and scientists from other disciplines, hailing from countries like Benin, Cameroon, the Ivory Coast, Kenya, Namibia, and Zambia.

One aim of the training is to help African labs become autonomous in both the generating and the analysis of genomic data.

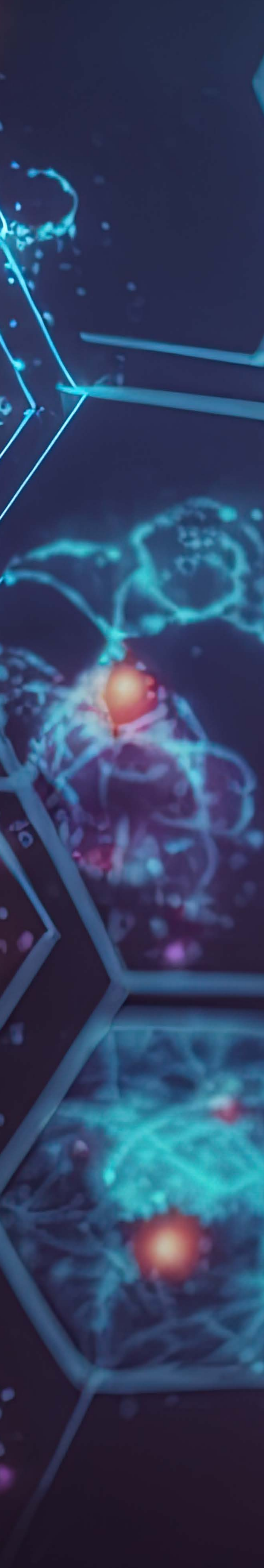
BAOBAB LIMS Training

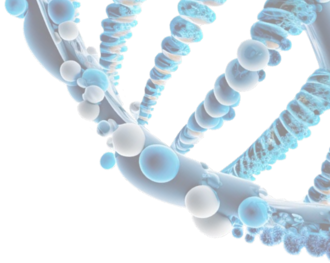
The Sickle In Africa Laboratory and Biorepository Workshop was hosted at UCT from 7-9 November. Participants from Ghana, Uganda, Tanzania, Nigeria, Zimbabwe and Zambia were in attendance. This workshop was aimed at research in Sickle cell disease and Dominique Anderson was requested to demonstrate Baobab LIMS at the workshop.

PHA4GE Ethics and Data Sharing Working Group Workshop

Nicki Tiffin, Anja Bedeker and Jamie Southgate organised and attended the workshop from 14 – 17 November in Mauritius. This workshop provided training and support to recipients of PHA4GE Ethics and Data Sharing sub-awards which were awarded to four teams conducting research into ethics and data governance challenges in under-resourced or challenging research environments. Participants hailed from Botswana, South Africa, Ethiopia and Nepal.

Workshop Participants – Team Ethiopia (Bizuneh Wakuma, Eba Abdisa, Gemechu Tiruneh), Team Botswana (Kagiso Ndlovu, Julia Gaorekwe, Tumalano Sekoto), Team South Africa (Jihaan Adonis), Team Nepal (Minu Singh) & PHA4GE Mentors (Victoria Nembaware, Nicky Tiffin, Anja Bedeker & Jamie Southgate).





COMPUTATIONAL RESOURCES

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

IT AND SYSTEMS SUPPORT

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.

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 serves.

SANBI HPC CLUSTER

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

SERVER COUNT
6
M710HD
64 GB - 128 GB
72 CPU CORES

232
CPU CORES

SERVER COUNT
3
M630
96 GB - 128 GB
48 CPU CORES

SERVER COUNT
1
M910HD
512 GB
16 CPU CORES

1952
GB OF RAM

SERVER COUNT
2
M610X
32 GB
24 CPU CORES

SERVER COUNT
6
M620
96 GB
72 CPU CORES

Research and Development driving SANBI's computational infrastructure deployment

PRINCIPAL INVESTIGATOR	THEME	PROJECT
Peter van Heusden	Research Computing Infrastructure	• Studying the deployment of research computing infrastructure and development of associated skills on the African continent through RSSE Africa.
	Pathogen surveillance Platforms	• Researching the use of pathogen genomics in a public health context for drug resistance prediction and genomic surveillance through the development of the COMBAT-TB and the COMBAT-SARS-CoV-2 Workbench projects.
	Poliovirus Surveillance	• Supporting poliovirus surveillance through automation of bioinformatics analysis pipelines, in collaboration with the National Institute for Communicable Diseases' Centre for Vaccines and Immunology.
	Meningococcal meningitis surveillance	• Developing field-deployment bioinformatics and sequencing solutions for meningococcal disease surveillance together with the Institut National de Récherche Biomedicale (INRB) in the Democratic Republic of Congo.

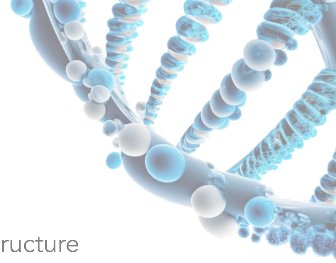
BAOBAB LIMS

Baobab LIMS is an open-source laboratory information management system developed by SANBI. Key to continuous development and improvement of Baobab LIMS following 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. 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. In addition, the development of a Viral Genome Analysis module which builds on the metadata standard, allows capture of basic workflows such as genomic material extraction, qPCR and sequencing, with the ability to crosslink sample metadata to sequencing QC in other systems, using a library ID.

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 occurs, 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.

Several workshops and demos for the use of Baobab LIMS have been provided throughout 2022, most being virtual. In an effort to upgrade the LIMS infrastructure parallel development on updated versions of Baobab LIMS (Lite and Pro) is currently underway.

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

COMBAT-TB AND COMBAT-SARS-COV-2 WORKBENCH

Building on SANBI's prior development of the COMBAT-TB Workbench, we unveiled the COMBAT-SARS-CoV-2 Workbench in 2022. This is a web based graphical user interface that makes it easier for laboratories sequencing the SARS-CoV-2 virus with either Illumina or Oxford Nanopore technology to assemble and analyse genomes.

The Workbench, like the previous COMBAT-TB Workbench, is based on IRIDA from the Public Health Agency of Canada. It uses Docker and Singularity container technology to ease deployment of the software. This ease of deployment allowed us to use the Workbench for training during 2022, where each lab participating in training was able to use their own instance of the Workbench (running on the Ilifu Cloud) to get hands-on experience in analysing SARS-CoV-2 sequences.

We also published a paper on the COMBAT-TB Workbench and demonstrated the platform during a visit to the Rwanda Biomedical Center in 2022.

GALAXY

SANBI remains an active part of the Galaxy Community. and continued to updated tools for microbial genomics and other bioinformatics use cases throughout 2022. Peter van Heusden participated in the second Galaxy Smörgåsbord global online training event in March 2022 and also helped develop the "Mycobacterial Tuberculosis NGS made easy" training resource. This introduction to M. tuberculosis sequence analysis, outbreak analysis and phylogenetics was offered as an online course twice in 2022, in a collaboration between SANBI, the Swiss Tropical and Public Health Institute (Swiss TPH) and IRCSS Ospedale San Raffaele (Italy) and the global diagnositics NGO FIND.

RESEARCH SOFTWARE AND SYSTEMS ENGINEERING AFRICA (RSSE AFRICA)

Initiated by Peter van Heusden with support from Talarify (a startup founded by ex-SANBI student Anelda van der Walt that supports researchers who want to develop digital and computational skills), a series of online presentations and community discussions were held from May - December 2022. The aim of these events is to stimulate the growth of an African community of practice around research software and system engineering practices. By facilitating and promoting collaboration between African RSSEs and other regions, there is benefit from sharing innovative solutions implemented in different contexts while at the same time highlighting the work and skills on the African continent. The series alternated between webinars and community meetups to provide the African community with an opportunity to learn and share their own experiences.

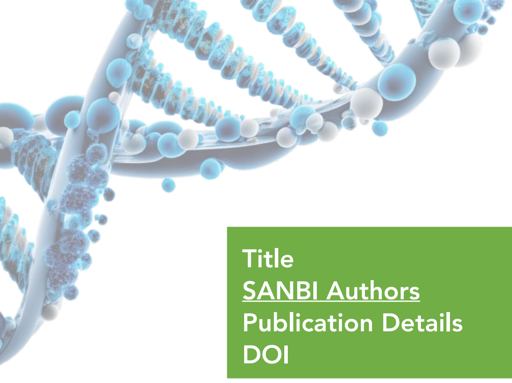
RESEARCH OUTPUTS

SANBI's profile of excellence is reflected in its ability to publish high impact scientific articles in international publications. Bioinformatics is an interdisciplinary field as evidenced by our collaborative publication outputs. Details of the institute's outputs are described in the tables that follow.

Journal Publications

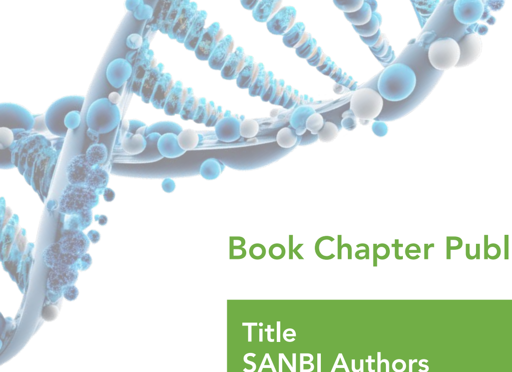
Title **SANBI Authors** Publication Details DOI	Published Date
Future-proofing and maximising the utility of metadata: The PHA4GE SARS-CoV-2 contextual data specification package. Griffiths EJ, Timme RE, Mendes CI, Page AJ, Alikhan N, Fornika D, Maguire F, Campos J, Park D, Olawoye IB, Oluniyi PE, **Anderson D, Christoffels A**, *et al.* and on behalf of the Public Health Alliance for Genomic Epidemiology (PHA4GE) consortium. GigaScience, 2022, Volume 11, Page1–13. DOI: 10.1093/gigascience/giac003	7 January
A framework for the promotion of ethical benefit sharing in health research. **Bedeker A**, Nichols M, Allie T, **Tamuhla T**, **van Heusden P**, Olorunsogbon O, **Tiffin N** for the PHA4GE Ethics and Data-Sharing Working Group. BMJ Global Health 2022;7:e008096. DOI: 10.1136/ bmjgh-2021-008096	15 January
One-step synthesis of picolinohydrazides from fusaric acid: DFT, structural characterisation and molecular inhibitory studies on metastatic tumor-derived exosomal and non-exosomal proteins. Anand K, Ramesh M, Singh T, Balakumar C, Chithravel V, Prasher P, Katari NK, Gupta G, Singh SK, Chellappan DK, Dua K, Chavda V, Laishevtcev A, **Shahbaaz M**, Abdellattif MH, Saravanan M, Chuturgoon AA. Journal of Molecular Structure 1255 (2022) 132442. DOI: 10.1016/j.molstruc.2022.132442	19 January
Repurposing Based Identification of Novel Inhibitors against MmpS5-MmpL5 Efflux Pump of *Mycobacterium smegmatis*: A Combined *In Silico* and *In Vitro* Study. **Shahbaaz M,** Maslov DA, Vatlin AA, Danilenko VN, Grishina M, **Christoffels A**. Biomedicines 2022; 10(2):333. https://doi.org/10.3390/biomedicines10020333	31 January
The COMBAT-TB Workbench: Making Powerful *Mycobacterium tuberculosis* Bioinformatics Accessible. **van Heusden P, Mashologu Z, Lose T,** Warren R, **Christoffels A.** ASM Journals, mSphere, Vol. 7, No. 1 https://doi.org/10.1128/msphere.00991-21	9 February
Current Status of Next-Generation Sequencing Approaches for Candidate Gene Discovery in Familial Parkinson´s Disease. **Pillay N S,** Ross O A, **Christoffels A**, Bardien S. Frontiers in Genetics, Volume 13 DOI: 10.3389/fgene.2022.781816	1 March

Title SANBI Authors Publication Details DOI	Published Date
High resolution HLA~A, ~B, ~C, ~DRB1, ~DQA1, and ~DQB1 diversity in South African populations. Tshabalala M, Mellet J, Vather K, Nelson D, Mohamed F, Christoffels A and Pepper MS. Frontiers Genetics 13:711944. DOI: 10.3389/fgene.2022.711944	4 March
Upregulation of FHL1, SPNS3, and MPZL2 predicts poor prognosis in pediatric acute myeloid leukemia patients with FLT3-ITD mutation. Eshibona N, Giwa A, Rossouw SC, Gamieldien J, Christoffels A, Bendou H. Leukemia & Lymphoma. DOI: 10.1080/10428194.2022.2045594	6 March
The 2021 WHO catalogue of *Mycobacterium tuberculosis* complex mutations associated with drug resistance: a genotypic analysis. Walker TM, Miotto P, Köser CU, Fowler PW, Knaggs J, Iqbal Z, Hunt M, Chindelevitch L, Farhat MR, Cirillo DM, Comas I, Posey J, Omar SV, Peto TEO, Suresh A, Uplekar S, Laurent S, Colman RE, Nathanson CM, Zignol M, Walker AS, Crook DW, Ismail N,Rodwell TC, Christoffels A as part of the CRyPTIC Consortium, the Seq&Treat Consortium. Lancet Microbe 2022;3: e265–73. https://doi.org/10.1016/S2666-5247(21)00301-3	8 March
microRNA profile of Hermetia illucens (black soldier fly) and its implications on mass rearing. DeRaedt S, Bierman A, van Heusden P, Richards C, Christoffels A. PLoS ONE 17(3): e0265492. https://doi.org/10.1371/journal.pone.0265492	17 March
Coronavirus Host Genetics South Africa (COHG-SA) database-a variant database for gene regions associated with SARS-CoV-2 outcomes. Barmania F, Mellet J, Ryder MA, Ford G, Herd CL, Tamuhla T, Hendricks C, Giles R, Kalua T, Joubert F, Tiffin N, Pepper MS. European Journal of Human Genetics (1018-4813). https://pubmed.ncbi.nlm.nih.gov/35351987/	29 March
Molecular Modeling of Subtype- Specific Tat Protein Signatures to Predict Tat-TAR Interactions That May Be Involved in HIV-Associated Neurocognitive Disorders. Williams ME and Cloete R. Frontiers in Microbiology. 13:866611. DOI: 10.3389/fmicb.2022.866611	7 April
Selection Analysis Identifies Clusters of Unusual Mutational Changes in Omicron Lineage BA.1 That Likely Impact Spike Function. Martin DP, Lytras S, Lucaci AG, Maier W, Grüning B, Shank SD, Weaver S, MacLean OA, Orton RJ, Lemey P, Boni MF, Tegally H, Harkins GW, Scheepers C, Bhiman JN, Everatt J, Amoako DG, San JE, Giandhari J, Sigal A; NGS-SA, Williamson C, Hsiao NY, von Gottberg A, De Klerk A, Shafer RW, Robertson DL, Wilkinson RJ, Sewell BT, Lessells R, Nekrutenko A, Greaney AJ, Starr TN, Bloom JD, Murrell B, Wilkinson E, Gupta RK, de Oliveira T, Kosakovsky Pond SL. Molecular Biology Evolution 39(4):msac061. DOI: 10.1093/molbev/msac061	10 April



Title **SANBI Authors** **Publication Details** **DOI**	**Published Date**
A View on Genomic Medicine Activities in Africa: Implications for Policy. Jongeneel CV, Kotze MJ, Bhaw-Luximon A, Fadlelmola FM, Fakim YJ, Hamdi Y, Kassim SK, **Kumuthini J**, Nembaware V, Radouani F, **Tiffin N** and Mulder N. Frontiers in Genetics 13:769919. DOI: 10.3389/fgene.2022.769919	27 April
Expanding the Galaxy's reference data. VijayKrishna N, Joshi J, Coraor N, Hillman-Jackson J, Bouvier D, van den Beek M, Eguinoa I, Coppens F, Davis J, Stolarczyk M, Sheffield NC, Gladman S, Cuccuru G, Grüning B, Soranzo N, Rasche H, Langhorst BW, Bernt M, Fornika D, de Lima Morais DA, Barrette M, **van Heusden P**, Petrillo M, Puertas-Gallardo A, Patak A, Hotz HR, Blankenberg D. Bioinformatics Advances 2(1):vbac030. DOI: 10.1093/bioadv/vbac030	29 April
Conserved recombination patterns across coronavirus subgenera. de Klerk A, Swanepoel P, Lourens R, Zondo M, Abodunran I, Lytras S, MacLean OA, Robertson D, Kosakovsky P, Sergei L, Zehr JD, Kumar, V, Stanhope MJ, **Harkins, G**, Murrell, B, Martin DP. Virus Evolution 8(2) 1-15. DOI: 10.1093/ve/veac054	9 July
Genomic Surveillance of Rift Valley fever Virus: from Sequencing to Lineage assignment. Juma J, Fonseca V, Konongoi SL, **van Heusden P**, Roesel K, Sang R, Bett B, **Christoffels A**, Oliveira T, Oyola SO. BMC Genomics (2022) 23:520. https://doi.org/10.1186/s12864-022-08764-6	18 July
The Flexible, Extended Coil of the PDZ-Binding Motif of the Three Deadly Human Coronavirus E Proteins Plays a Role in Pathogenicity. Schoeman D, **Cloete R**, Fielding BC. Viruses 2022, 14, 1707. https://doi.org/10.3390/v14081707	2 August
Cucumis melo var. momordica as a Potent Antidiabetic, Antioxidant and Possible Anticovid Alternative: Investigation through Experimental and Computational Methods. Yadav JP, Grishina M, **Shahbaaz M,** Mukerjee A, Singh SK, Pathak P. Chemistry and Biodiversity e202200200. DOI: 10.1002/cbdv.202200200	11 August
Establishing MinION Sequencing and Genome Assembly Procedures for the Analysis of the Rooibos (Aspalathus linearis) Genome. **Mgwatyu Y**, Cornelissen S, **van Heusden P**, Stander A, Ranketse M, **Hesse U**. Plants 2022, 11, 2156. https://doi.org/10.3390/ plants11162156	19 August
Record linkage for Routinely Collected Health Data in an African Health Information Exchange. Mutemaringa T, Heekes A, Boulle A and **Tiffin N.** International Journal of Population Data Science (2022) 7:3:245. https://doi.org/10.23889/ijpds.v7i3.2021	September

Title SANBI Authors Publication Details DOI	Published Date
The evolving SARS-CoV-2 epidemic in Africa: Insights from rapidly expanding genomic surveillance. Tegally H, San JE, Cotten M, Moir M, Tegomoh B, Mboowa G, Martin DP, Baxter C, Lambisia AW, Diallo A, Amoako DG, Diagne MM, Sisay A, Zekri AN, Gueye AS, Sangare AK, Ouedraogo AS, Sow A, Musa AO, Sesay AK, Abias AG, Elzagheid AI, Lagare A, Kemi AS, Abar AE, Johnson AA, Fowotade A, Oluwapelumi AO, Amuri AA, Juru A, Kandeil A, Mostafa A, Rebai A, Sayed A, Kazeem A, Balde A, **Christoffels A,** *et al*. Science. 2022 Sep 15:eabq5358. DOI: 10.1126/science.abq5358	15 September
The determination of the effect(s) of solute carrier family 22-member 2 (SLC22A2) haplotype variants on drug binding via molecular dynamic simulation systems. Abrahams-October Z, Johnson R, Benjeddou M, **Cloete R**. Scientific Reports (2022) 12:16936. https://doi.org/10.1038/s41598-022-21291-4	8 October
Neurexin 2 p.G849D variant, implicated in Parkinson's disease, increases reactive oxygen species, and reduces cell viability and mitochondrial membrane potential in SH-SY5Y cells. Cuttler K, de Swardt D, Engelbrecht L, Kriel J, **Cloete R**, Bardien S. Journal of Neural Transmission. DOI: 10.1007/s00702-022-02548-8	15 October
The Implementation of Laboratory Information Management System in Multi-Site Genetics Study in Africa: The Challenges and Up-Scaling Opportunities. Oluwole OG, Oosterwyk C, **Anderson D**, Adadey SM, Mnika K, Manyisa N, Yalcouye A, Wonkam ET, Aboagye, ET, Dia Y, *et al*. Journal of Molecular Pathology 2022, 3, 262–272. https://doi.org/10.3390/jmp3040022	22 October
Changes in subcutaneous adipose tissue microRNA expression in response to exercise training in obese African women. Pheiffer C, Dias S, Mendham AE, Jack B, Willmer T, **Eshibona N, Bendou H**, Pretorius A, Goedecke JH. Scientific Reports (2022) 12:18408. https://doi.org/10.1038/s41598-022-23290-x	1 November
Virion-Associated Nucleic Acid-Based Metagenomics: A Decade of Advances in Molecular Characterisation of Plant Viruses. Moubset O, François S, Maclot F, Palanga E, Julian C, Claude L, Fernandez E, Rott P, Daugrois J, Antoine-Lorquin A, Bernardo P, Blouin AG, Temple C, Kraberger S, Fontenele RS, **Harkins GW**, Ma Y, Marais A, Candresse T, *et al*. Phytopathology 2022 Nov;112(11):2253-2272. DOI: 10.1094/PHYTO-03-22-0096-RVW	November
An e-consent framework for tiered informed consent for human genomic research in the global south, implemented as a REDCap template **Tamuhla T**, **Tiffin N**, Allie T. BMC Medical Ethics, 23, 119 (2022). https://doi.org/10.1186/s12910-022-00860-2	24 November
Precision medicine for developmental and epileptic encephalopathies in Africa—strategies for a resource-limited setting, Esterhuizen AI, **Tiffin N**, Riordan G, Wessels M, Burman RJ, *et al*. Genetics in Medicine, ISSN 1098-3600. https://doi.org/10.1016/j.gim.2022.11.002.	8 December

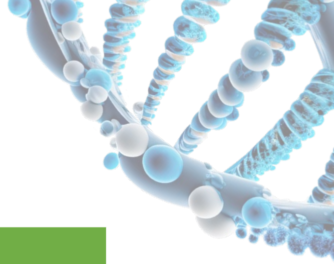


Book Chapter Publications

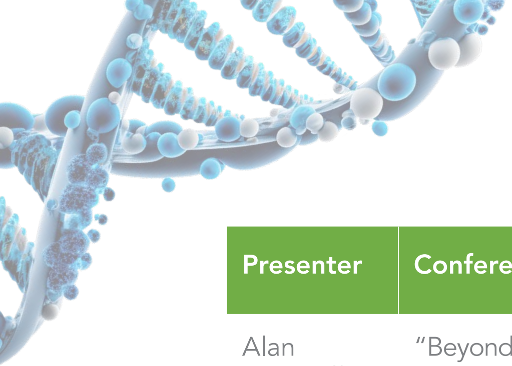
Title **SANBI Authors** **IN**	**Published Date**
Testing Times: COVID-19 Testing and Healthcare Workers in South Africa. Lewins K, **van Heusden P**, Baldwin-Ragaven L. **IN:** Caring on the Frontline during COVID-19. Contributions from Rapid Qualitative Research. By Cecilia Vindrola-Padros, Ginger A. Johnson.Publisher: Palgrave Macmillan, Singapore ISBN Hardcover: 978-981-16-6485-4 ISBN Softcover: 978-981-16-6488-5 https://doi.org/10.1007/978-981-16-6486-1	1 January
Software Tools for Biobanking in LMICs. **Anderson D, Bendou H**, Kipperer B, Zatloukal K, Muller H, **Christoffels A**. **IN:** Biobanks in Low- and Middle-Income Countries: Relevance, Setup and Management. 1st ed. 2022 pp 137 – 146 Editors: Sargsyan Karine, Huppertz Berthold, Gramatiuk Svetlana Publisher: Springer International Online ISBN: 978-3-030-87637-1 https://doi.org/10.1007/978-3-030-87637-1_20	16 March

Keynotes/Invited Talks Presented

Presenter	Conference Name and Venue	Title	Date
Nicki Tiffin	Festival of Genomics (online)	Promoting equitable sharing of samples biospecimens originating in Africa	26 January
Nicki Tiffin	Human Cell Atlas Meeting (online)	Ethical, legal and social responsibilities for health genomics research in Africa	21 March
Nicki Tiffin	ICDA scientific plenary (online)	Improving standards for global equity in research	8 March
Nicki Tiffin	Calestous Juma Fellowship Kick-off meeting (online)	Facilitating equitable data and biospecimen sharing in Africa	5 April
Alan Christoffels	Grand Challenges Conference 2022 – Public Policy Projects Webinar	Using technology to address science inequity in genomics	24 May
Alan Christoffels	32nd ECCMID (European Congress of Clinical Microbiology and Infectious Diseases), in Lisbon, Portugal	Invited talk: Strengthen Genomics and Bioinformatics in response to EIDs in Africa	22 - 26 April



Presenter	Conference Name and Venue	Title	Date
Nicki Tiffin	The 25th Human Genome Meeting, Human Genome Organisation. Tel Aviv, Israel	Advancing African health genomics research through innovative study design	22 May
Alan Christoffels	Regional workshop on the implementation of Public health pathogen genomics in the west Africa region. Accra Ghana.	Talk: Pathogen Genomics initiative update for 2022	31 May - 2 June
Alan Christoffels	Africa CDC PGI Consultative Workshop on Pathogen Genomics Data Architecture Roadmap. in Addis Ababa, Ethiopia	Invited Talk: Chair of the data technical working group for the Africa CDC. Strategy for a technology roadmap to implement a continental data platform.	14 - 16 June
Nick Tiffin, Tsaone Tamuhla	Multimorbidity Research in Sub-Saharan Africa: A Regional Retreat, Blantyre, Malawi	Exploring synergies between the health service and the research ecosystem: pragmatic multimorbidity research design	23 June
Alan Christoffels	5th Annual BioAfrica Convention. ICC Durban, South Africa	Invitation: Panel discussion: Innovative approaches for public health pathogen genomics and bioinformatics in Africa.	27 - 31 August
Nicki Tiffin	GA4GH 10th Plenary, Barcelona	Meeting Chair: Global Policy Forum Workshop	21 September
Nicki Tiffin	IHCC Educational Webinar Series (online)	Webinar moderator: COVID-19 genetics cohorts	3 October
Nick Tiffin	Bill & Melinda Gates Foundation Data, Grand Challenges Meeting, Brussels, Belgium	Science Global Call to Action – moderator. Session Chair: Genomic Surveillance Beyond SARS-CoV-2. J&J Janssen Beerse site Fireside Chat	20 - 27 October
Nicki Tiffin	UK Biobank Meeting: Population Health Studies and the African research community, Cape Town	Approach to Data Integration Platforms: Integrating and sharing African health data	2 - 3 November
Alan Christoffels	UK Biobank Meeting: Population Health Studies and the African research community, UCT, South Africa	Invited talk: Data analysis and skills gap	2 - 3 November
Nicki Tiffin	Data science in Africa, Global Health Network Meeting, Cape Town	Integrating and sharing African health data	21 November



Presenter	Conference Name and Venue	Title	Date
Alan Christoffels	"Beyond COVID-19: Pathogen Genomics and Bioinformatics for Health Security in Africa" symposium. African Union Commission in Addis Ababa, Ethiopia.	Chair of the conference: Talk: PHA4GE	29 November - 1 December
Nicki Tiffin	IHCC Scientific Meeting (online)	Policy and Data sharing working group presentations. Global Equity and Ethics Forum presentation.	29 - 31 November

Conference Participation

Presenter Name	Conference Name and Venue	Type and Title	Date
Ruben Cloete	BIO2022 at Stellenbosch Institute for Advance Study	Oral: Molecular docking of Zinc database natural compounds to SARSCOV-2 coronavirus proteins to identify novel inhibitors with antiviral activity.	24 - 27 April
Themba Mutemaringa	International Population Data Linkage Network Conference, Edinburgh, United Kingdom,	Poster and Conference Proceeding: Record linkage for Routinely Collected Health Data in an African Health Information Exchange.	7 - 9 September
Anja Bedeker, Nicki Tiffin	BMGF Grand Challenges Annual Meeting 2022. Brussels, Belgium	Poster: A framework for the promotion of ethical benefit sharing in health research.	22 October
Alan Christoffels	BMGF Grand Challenges Annual Meeting 2022. Brussels, Belgium	Poster: PHA4GE: Improving the structure of pathogen genomics data for effective public health action.	22 October
Jamie Southgate, Kevin Libuit	BMGF Grand Challenges Annual Meeting 2022. Brussels, Belgium	Poster: Bioinformatics pipelines and data visualisation working group: Integrating bioinformatics solutions to improve global Public Health.	22 October

Presenter Name	Conference Name and Venue	Type and Title	Date
Jamie Southgate, Rangarirai Matima, Alecia Naidu, Nawaal Weitz, Michael Bridger, Alan Christoffels	BMGF Grand Challenges Annual Meeting 2022. Brussels, Belgium	Poster: PHA4GE: Improving openness and interoperability in public health bioinformatics	22 October
Peter Van Heusden	BMGF Grand Challenges Annual Meeting 2022. Brussels, Belgium	Poster: PHA4GE Infrastructure Working Group: Framework for Compute Infrastructure for Pathogen Genomics Labs	22 October

The Africa CDC team at Grand Challenges Annual Meeting in Brussels.

RESEARCH LABORATORIES

DR DOMINIQUE ANDERSON

RESEARCH PROJECT THEMES

Informatics Tools

- Biobank informatics – multiple collaborations for biocollection sample quality management and e-infrastructure with a focus on Baobab LIMS.
- Pathogen genome databases.

Biomedical data governance

- Collaboration with biobank and law experts to evaluate data governance at tertiary institutions and biomedical research facilities.
- Cybersecurity and data management in Africa.
- Metadata standardisation and impacts of data management in public health.

Dental Genetics

- Dental genetics in collaboration with the Faculty of Dentistry investigating oral microbiomes.

One Health

- *In silico* research focussed on OneHealth.

HIGHLIGHTS OF 2022

- Supervisor of 2 PhD students and 2 MSc students
- Co-supervisor of 1 PhD student, 1 MSc student

The research group is investigating several areas with an interest in dental metagenomics, data security and privacy, biomedical data governance, informatics of biobanking and OneHealth. I have an ongoing role in a team focused on informatics solutions for biobanking with continued development, enhancement, and training in the open-source Baobab laboratory information management system. 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. The pandemic highlighted the need for the LIMS to become more metadata aware and the focus was to incorporate viral contextual metadata developed by the PHA4GE consortium into the system. A recent collaboration has been established with the NHLS, to evaluate data quality and metadata standards with a view to improve data linkage in data marts in the public health setting. In relation to additional interests, I am involved in community engagement and effective science communication and training.

With regards to data privacy and governance, my collaborators and I combine expertise in biobanking, bioinformatics and big data, and law to investigate how both regulations and best practices in ICT influence the biomedical data and data sharing landscape in Africa. The UWC Dentistry faculty and SANBI have an ongoing collaborative effort to increase research into dental genetics and this includes both wet-bench and dry-bench approaches, particularly to analysis of microbiomes. I am also currently working towards establishing networks within the agricultural, veterinary, medical, and environmental sphere in order to intensify research in the area of One Health and machine learning.

RESEARCH COLLABORATIONS

1. Data governance and the POPI Act

Collaborating Parties:
Dr Carmen Swanepoel – NHLS/Stellenbosch University Tygerberg
ASSAf
Nature and purpose:
To investigate the impact of local and international regulation on the sharing of biomedical research data in Africa. Drafting committee member for the POPIA code of conduct for research.

Output in the last 12 months:
Publication: Adams R, Adeleke F, Anderson D, Bawa A, Branson N, Christoffels A, *et al.* POPIA Code of Conduct for Research. S Afr J Sci. 2021;117(5/6), Art. #10933. (https://doi.org/10.17159/sajs.2021/10933)

2. Dental genetics

Collaborating Parties:
Drs Manogari Chetty and Anthea Jeftha – Faculty of Dentistry, UWC
Nature and purpose:
Incorporating molecular biology, genetics and bioinformatics into dentistry research.
Output in the last 12 months:
None
Future Direction:
Dental metagenomics. Impact of dental health in disease.

3. Biobank informatics and data quality management

Collaborating Parties:
Prof Alan Christoffels – SANBI, UWC
NHLS and University of the Witwatersrand
Africa CDC
PHA4GE Consortium
Nature and purpose:
Enhancement and customisation of Baobab LIMS for LMICs. Standardisation of e-infrastructure. Evaluating the impact of data standardisation in public health. Data linkage algorithms in data warehouses.
Output in the last 12 months:
Africa CDC funding.
One MSc graduate.
Book chapter: Anderson, D., Bendou, H., Kipperer, B., Zatloukal, K., Müller, H., Christoffels, A. (2022). Software Tools for Biobanking in LMICs. In: Sargsyan, K., Huppertz, B., Gramatiuk, S. (eds) Biobanks in Low- and Middle-Income Countries: Relevance, Setup and Management. Springer, Cham. https://doi.org/10.1007/978-3-030-87637-1_20

Publication: Griffiths EJ, Timme RE, Mendes CI, Page AJ, Alikhan NF, Fornika D, Maguire F, Campos J, Park D, Olawoye IB, Oluniyi PE, Anderson D, Christoffels A, da Silva AG, Cameron R, Dooley D, Katz LS, Black A, Karsch-Mizrachi I, Barrett T, Johnston A, Connor TR, Nicholls SM, Witney AA, Tyson GH, Tausch SH, Raphenya AR, Alcock B, Aanensen DM, Hodcroft E, Hsiao WWL, Vasconcelos ATR, MacCannell DR. Future-proofing and maximising the utility of metadata: The PHA4GE SARS-CoV-2 contextual data specification package. Gigascience. 2022 Feb 16;11:giac003. doi: 10.1093/gigascience/giac003. PMID: 35169842; PMCID: PMC8847733.

Publication: Oluwole, O.G.; Oosterwyk, C.; Anderson, D.; Adadey, S.M.; Mnika, K.; Manyisa, N.; Yalcouye, A.; Wonkam, E.T.; Aboagye, E.T.; Dia, Y.; Uwibambe, E.; Jonas, M.; Priestley, R.; Popel, K.; Manyashe, T.; de Cock, C.; Nembaware, V.; Wonkam, A. The Implementation of Laboratory Information Management System in Multi-Site Genetics Study in Africa: The Challenges and Up-Scaling Opportunities. J. Mol. Pathol. 2022, 3, 262-272. https://doi.org/10.3390/jmp3040022
Future Direction: New collaborations within Africa as well as training initiatives on the continent. Maintenance and enhancement of Baobab LIMS. Entrepreneurship focused on hybrid social and sustainability models for QM software.

4. OneHealth

Collaborating Parties:
Seeking collaborative partners.
Nature and purpose:
Multidisciplinary approaches to understanding infectious diseases. Development of metadata standards, novel AI based tools for surveillance and species cross-over and translational research in *OneHealth*.
Output in the last 12 months:
MSc graduate: Mr Rudolph Serage, *Cum Laude.*

RESEARCH PROJECT THEMES

Cancer Bioinformatics
- Novel genomic biomarkers for Pediatric and Adult Acute Myeloid Leukemia.
- Computational analysis of multi-omic data for the elucidation of molecular mechanisms of neuroblastoma.

Pipeline development
- Development of a Nextflow pipeline for the detection of enhancer-associated noncoding variants using ChIP-seq data.

Machine learning
- Prediction of synthetic lethal gene interaction in human cancer using supervised machine learning.

Gene regulatory networks
- Gene regulatory network study of diffuse large B-cell lymphoma.

Viral Study
- Identify impact of viral inserts in human esophageal cancer. A collaborative work with the University of Cape Town.

DR HOCINE BENDOU

HIGHLIGHTS OF 2022

Submission of an NIH grant application with my postdoctoral student, Dr Catherine Rossouw, aimed at the early-stage development of informatics technologies for cancer research and management.

My research group identified a methylation CpG signature capable of predicting the amplification of MYCN, a predictor of poor prognosis, in the early stages of neuroblastoma. This achievement has the benefit of helping clinicians design better treatment for patients with MYCN amplification. The work was published in the journal Future Oncology with an impact factor of 3.4 (2020).

My PhD student Abdulazeez Giwa, submitted his thesis entitled *"Computational analysis of multi-omic data for the elucidation of molecular mechanisms of neuroblastoma"* for examination.

My PhD student Nasr Eshibona, through principal component analysis and hierarchical clustering of RNA-seq gene expression counts from paediatric acute myeloid leukaemia samples, found that samples carrying the FLT3-ITD mutation (poor prog nosis) were grouped, in the same cluster, with samples carrying the NPM1 and the CEBPA mutations (favourable prognosis). This surprising result led to further bioinformatics analyses and the discovery of three genes potentially responsible for the poor prognosis in FLT3-ITD mutated patients (a research manuscript is under review).

Launch of a collaborative research project with Dr Carmen Pheiffer from the South African Medical Research Council to study the effect of physical training on the expression of miRNAs in gluteal adipose tissue in women with obesity.

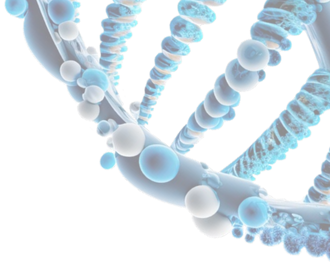
RESEARCH COLLABORATIONS

1. Molecular mechanisms of neuroblastoma

Collaborating Parties:
Dr Azeez Fatai – Department of Biochemistry, Lagos State University, Lagos, Nigeria
Abdulazeez Giwa – SANBI, UWC
Nature and purpose:
Elucidate the molecular mechanisms of neuroblastoma in high-risk patients using multi-omics data and machine learning approaches
Output in the last 12 months:
Publication article in Future Oncology (3.4 impact factor): Predicting amplification of *MYCN* using CpG methylation biomarkers in neuroblastoma.

2. Molecular Biology and Genetics of Oesophageal Cancer

Collaborating Parties:
Prof Iqbal Parker – University of Cape Town



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

Output in the last 12 months:
Whole genome sequencing analysis on paired tumour-normal samples from 60 participants. Prepare and publish a manuscript.

Future Direction:
Wet lab work for validation of the identified somatic mutations.

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

Collaborating Parties:
Dr Carmen Pheiffer – South African Medical Research Council
Nasr Eshibona – SANBI, UWC

Nature and purpose:
Provide bioinformatics expertise.

Output in the last 12 months:
Manuscript submitted to an international journal with a high impact factor.

RESEARCH PROJECT THEMES

Pathogen genomics

- In collaboration with the Tygerberg MRC Unit, we are developing methods to analyse high throughput sequencing data for microbial genomes.
- Identification of novel drug targets in pathways known to contain drug resistant genes.
- Predicting the interaction networks between human and mycobacteria
- Identification of operons & non-coding RNAs to understand gene regulation in pathogen genomes.

Pathogen surveillance

- Developing systems to support reproducible analyses
- Development of global standards for data sharing and integration
- Public Health Alliance for Genomic Epidemiology (PHA4GE)

Drug discovery

- Identifying drug molecules against multi-drug resistant M.tuberculosis
- In partnership with the School of Pharmacy at UWC, delineate the pharmacokinetics of drugs in response to host genetic factors
- In partnership with the School of Pharmacy at UWC, use of machine learning techniques to prioritise molecules in natural compounds.

PROF ALAN CHRISTOFFELS

HIGHLIGHTS OF 2022

During 2022, we maintained an active research programme. Our work on COMBAT-TB has allowed us to leverage funding from Africa CDC to develop a data analytics platform to analyse SARS-CoV-2 sequencing data. The lab continues to work with the Africa CDC to conceptualise a data sharing platform. In October 2021, we formed a partnership with WHO-AFRICA regional office (WHO-AFRO) to establish a regional center for bioinformatics and genomic surveillance that would support and strengthen the SADC region in response to disease outbreaks, including COVID-19.

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

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.

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

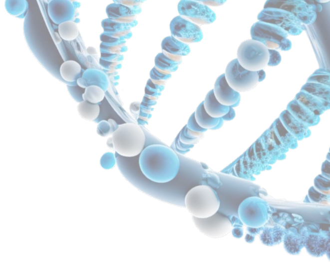
Recently we used quadruplet amino acids as a feature in training model for protein interactions.
We are extending this work to improve on the accuracy of our models with small datasets.

RESEARCH COLLABORATIONS

1. Develop a biobank informatics management system

Collaborating Parties:

Dr Dominique Anderson – SANBI, UWC and Africa CDC



Nature and Purpose:
Development of a laboratory management system for infectious disease labs that will integrate with other biobanking tools.
Output in the last 12 months:
Added a freezer monitoring schedule. Submitted a book chapter on biobank software for low and middle income countries.

2. Analysing genetic networks in 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 is currently benchmarked against two international methods.
Using a sequence conversation approach and a RNA folding approach, we are identifying putative small RNAs.

3. Identification of novel drug targets for drug resistant tuberculosis

Collaborating Parties:
Profs Samantha Sampson and Rob Warren – University of Stellenbosch
Dr Ruben Cloete – SANBI, UWC
Prof Sarel Malan – School of Pharmacy, UWC
Prof Valery Danilenko and Dr Dmitry Maslov – Vavilov Institute of General Genetics, Russia
Nature and Purpose:
Current 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. And submitted a funding application via the BRICS-country funding mechanism.

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

Collaborating Parties:
Profs Samantha Sampson and Rob Warren – University of Stellenbosch
Peter van Heusden – SANBI, UWC
Nature and Purpose:
Develop a computational platform to store Tuberculosis omic data and to provide a visualisation tool.
Output in the last 12 months:
Added further analysis tools and a user manual. Our manuscript was published in *mSphere* Journal.

5. Biosecurity Preparedness in Africa

Collaborating Parties:
Dr Dominique Anderson – SANBI, UWC
Global Consortium on emerging infectious diseases https://www.getafrica.org/.
Nature and Purpose:
We aim to establish infrastructure and research capacity to respond to highly infectious emerging Pathogens.
Output in the last 12 months:
We have reviewed the information security policies at SA Universities with a view to propose guidelines for data management that impacts biomedical researchers.

6. Public health alliance for genomic epidemiology (PHA4GE)

Collaborating Parties:
Africa CDC, Oxford University, Washington University, University of Melbourne, USA CDC, University of Birmingham UK, University of British Columbia and BC Centre for Disease Control, Broad Institute in Boston, and H3Africa.
Nature and Purpose:
The PHA4GE consortium was established to bring Bioinformatics closer to Public Health – to build data standards.
Output in the last 12 months:
PHA4GE data standards working group developed a metadata standard for COVID-19 biospecimen collection.

7. African Genomics Archive

Collaborating Parties:
Africa CDC
Nature and Purpose:
To establish a data platform in Africa to manage disease surveillance data with a view to strengthen public health institutes.
Output in the last 12 months:
A series of workshops to gather information on various use cases.

DR RUBEN CLOETE

RESEARCH PROJECT THEMES

Pathogen resistance

- 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

- Previous pathway mapping and gene prioritisation methods allowed us to identify known and novel *Mycobacterium tuberculosis* drug targets. Here we exploit the drug targets using virtual screening methods to identify novel drugs to treat drug resistant tuberculosis. The drugs identified are tested for activity against *Mycobacterium tuberculosis* using a bioluminescent reporter assay system. *In silico* discovery of compounds with activity against the novel SARS-coronavirus and *in vitro* testing.

Human disease genomics

- 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

- Using machine learning methods to identify putative compounds with activity against *Mycobacterium tuberculosis* and HIV.

Diagnostic tool development

- Develop structural methods to aid in the identification of biomarkers useful for Tuberculosis and Ebola virus diagnosis.

HIGHLIGHTS OF 2022

- Currently, I supervise two PhD students, Mr Mohammed Hassan and Mrs Adetutu Akinnuwesi and I co-supervise two students, Miss Katelyn Kuttler from Stellenbosch University and Mr Darius Martin from the Biotechnology Department, UWC.
- My UWC Masters students, Mr Darren Isaacs and Miss Maryam Hassan, graduated in 2021.
- Two new MSc students joined my group in 2021: Mr Keaghan Brown (University of Stellenbosch) and Miss Lunathi Mihlali Ndlondlo (UWC).
- **Grants/Funding Awarded:** Co-Principal investigator on a SAMRC RCDI grant with Prof Megan Shaw for 2021-2022.

RESEARCH COLLABORATIONS

1. Novel drug discovery against *Mycobacterium tuberculosis*

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

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

Output in the last 12 months:
One published article: Cloete R, Shahbaaz M, Grobbelaar M, Sampson SL, Christoffels A. *In silico* repurposing of a Novobiocin derivative for activity against latency associated *Mycobacterium tuberculosis* drug target nicotinate-nucleotide adenylyl transferase (Rv2421c). PLoS One. 2021 Nov 2;16(11):e0259348. doi: 10.1371/journal.pone.0259348. PMID: 34727137.

Future Direction:
Draft two manuscripts one targeting *Mycobacterium tuberculosis* drug target Rv2196 and a follow up article on Rv2421c to identify novel drug molecules to treat Tuberculosis infections.

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

Collaborating Parties:
The late Dr Graeme Jacobs – Stellenbosch University (deceased)

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

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

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

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

Collaborating Parties:
Prof Soraya Bardien – Stellenbosch University
Nature and purpose:
To identify novel genes associated with Parkinson's disease development using Whole Exome sequencing and using structural computational methods to understand the impact of mutations on protein structure and function.
Output in the last 12 months:
Publications: 1) Sebate B, Cuttler K, Cloete R, Britz M, Christoffels A, Williams M, *et al*. Prioritisation of candidate genes for a South African family with Parkinson's disease using in-silico tools. PloS one. 2021;16(3):e0249324.
2) Cuttler K, Hassan M, Carr J, Cloete R, Bardien S. Emerging evidence implicating a role for neurexins in neurodegenerative and neuropsychiatric disorders. Open Biol. 2021 Oct;11(10):210091. doi: 10.1098/rsob.210091. Epub 2021 Oct 6. PMID: 34610269; PMCID: PMC8492176.
Future Direction:
Two manuscripts under preparation for publication following up on neurexin and neuroligin interaction associated with PD and another on a novel gene candidate identified in a South African family with PD using exome sequencing.

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:
Manuscript submitted for publication.

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

Collaborating Parties:
Prof Burtram Fielding – Department of Medical Biosciences, UWC
Nature and purpose:
Perform topology predictions, structural modelling as well as simulation studies to understand the structural differences between four coronavirus envelope proteins and their interaction with the human host PALS1 protein.
Output in the last 12 months:
None
Future Direction:
Manuscript in preparation.

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

Collaborating Parties:
Prof Mervin Meyer – Biotechnology Department, UWC
Nature and purpose:
The aim of the research is to develop multiplex lateral flow devices (LFDs) for the detection of serum human biomarker proteins for TB and Ebola diagnosis using DNA aptamers.
Output in the last 12 months:
Publications: Martin DR, Sibuyi NR, Dube P, *et al*. Aptamer-Based Diagnostic Systems for the Rapid Screening of TB at the Point-of-Care. Diagnostics (Basel, Switzerland). 2021 Jul;11(8). DOI: 10.3390/diagnostics11081352. PMID: 34441287; PMCID: PMC8391981.
https://doi.org/10.3390/diagnostics11081352
Future Direction:
To draft a research article on the identification of potential human biomarkers that can be targeted by DNA aptamers for Tuberculosis diagnosis. Complete an *in silico* research paper on the development of novel DNA aptamers against Ebola virus nucleoproteins as well as drafting a review article on Ebola virus diagnostic tools.

PROF GORDON HARKINS

RESEARCH PROJECTS

SARS-CoV-2

- The characterisation and clinical manifestation of the SARS-CoV-2 outbreak in Uruguay
- The evolutionary dynamics of SARS-CoV-2 lineages during the first epidemic wave in New York City
- The role of natural selection in the emergence and ongoing evolution of the N501Y lineages (V1-V3) of SARS-CoV2
- Selection Analysis Identifies Unusual Clustered Mutational Changes in Omicron Lineage BA.1 that Likely Impact Spike Function
- Conserved recombination in coronavirus subgenera
- T cell immune escape mutations at HLA-binding anchor motifs found in SARS-CoV-2 epitopes from genomes isolated in South Africa and subSaharan Africa

Human immunodeficiency virus (HIV-1)

- HIV-1 latent viral reservoir dynamics

Phytovirus Diversity

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.

HIGHLIGHTS OF 2022

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

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

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

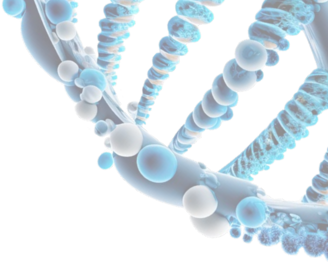
RESEARCH THEMES AND PROJECTS

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

RESEARCH PROJECTS AND 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 Zappiled 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 epicentre of the COVID-19 epidemic in the United States.

Output in the last 12 months:
One paper was published in *PLoS Pathogens* in 2021.

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

3. Investigating the role of natural selection in the emergence and ongoing evolution of the N501Y lineages (V1-V3) of SARS-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 published in the journal *Cell* in 2021.

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

4. Selection Analysis Identifies Unusual Clustered Mutational Changes in Omicron Lineage BA.1 that Likely Impact Spike Function.

Collaborating Parties:
Darren P Martin, Arne De Klerk – Institute of Infectious Diseases and Molecular Medicine, Division Of Computational Biology, Department of Integrative Biomedical Sciences, University of Cape Town, South Africa.
Philippe Lemey – Department of Microbiology, Immunology and Transplantation, Rega Institute, KU Leuven, Leuven, Belgium.
Steven Weaver, Stephen D Shank, Sergei L Kosakovsky Pond – Institute for Genomics and Evolutionary Medicine, Department of Biology, Temple University, Pennsylvania, USA
Houryiah Tegally, Emmanuel James San, Eduan Wilkinson, Jennifer Giandhari, Richard J Lessells, Anton Nekrutenko, Tulio De Oliveira – KwaZulu-Natal Research Innovation and Sequencing Platform (KRISP), School of Laboratory Medicine & Medical Sciences, University of KwaZulu-Natal, Durban, South Africa NGS-SA.
Ben Murrell – Department of Microbiology, Tumor and Cell Biology, Karolinska Institutet, Stockholm, Sweden.
Carolyn Williamson – Division of Medical Virology, Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Cape Town, South Africa.
Spyro Lytras, Oscar S MacLean, David L Robertson, Richard J Orton – MRC-University of Glasgow Centre for Virus Research, University of Glasgow, Glasgow G61 1QH, UK
Alexander G Lucaci – Institute for Genomics and Evolutionary Medicine, Department of Biology, Temple University, Philadelphia, PA 19122, USA.
Wolfgang Maier, Bjorn Gruning – Bioinformatics Group, Department of Computer Science, University of Freiburg, Freiburg, Germany.
Maciej F Boni – Center for Infectious Disease Dynamics, Department of Biology, Pennsylvania State University, University Park, PA, USA.
Cathrine Scheepers, Jinal N Bhiman – National Institute for Communicable Diseases (NICD) of the National Health Laboratory Service, (NHLS), Johannesburg, South Africa, SA MRC Antibody Immunity Research Unit, School of Pathology, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa.
Josie Everatt, Daniel G Amoako – National Institute for Communicable Diseases (NICD) of the National Health Laboratory Service, (NHLS), Johannesburg, South Africa.
James Emanuel San, Jennifer Giandhari – KwaZulu-Natal Research Innovation and Sequencing Platform (KRISP), School of Laboratory Medicine & Medical Sciences, University of KwaZulu- Natal, Durban, South Africa.
Alex Sigal, Ravindra K Gupta – Africa Health Research Institute, Durban, South Africa.
Nei-yuan Hsiao – Division of Medical Virology, University of Cape Town and National Health Laboratory Service, Cape Town South Africa.
Anne von Gottberg – National Institute for Communicable Diseases (NICD) of the National Health Laboratory Service, (NHLS), Johannesburg, South Africa
Robert W Shafer – Division of Infectious Diseases, Department of medicine, Stanford university, Stanford, CA, USA
Robert J Wilkinson – Wellcome Center for Infectious Diseases Research in Africa, Institute of Infectious Disease and Molecular Medicine and Department of Medicine, University of Cape Town, South Africa
Brian Trevor Sewell – Structural Biology Research Unit, Department of Integrative Biomedical Sciences, Institute for Infectious Diseases and Molecular Medicine, University of Cape Town, South Africa
Allison J Greaney, Tyler N Starr, Jesse D Bloom – Basic Sciences Division and Computational Biology Program, Fred Hutchinson Cancer Research Center, Seattle; Department of Genome Sciences & Medical Scientist Training Program, University of Washington, Seattle, Howard Hughes Medical Institute, Seattle, WA

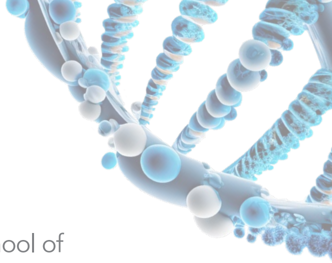
Nature and purpose:
To determine what drives the mutational changes in the Omicron variant of concern (VOC).

Output in the last 12 months:
One paper published in 2022 in the Journal Molecular Biology and Evolution.

Future Direction:
This is an ongoing collaboration.

5. Conserved recombination in coronavirus subgenera

Collaborating Parties:
Darren P Martin, Arné de Klerk, Phillip Swanepoel, Mpumelelo Zondo, Isaac Abodunran – Institute of Infectious Diseases and Molecular Medicine, Division Of Computational Biology, Department of Integrative Biomedical Sciences, University of Cape Town



Rentia Lourens – Division of Neurosurgery, Neuroscience Institute, Department of Surgery, University of Cape Town
Spyros Lytras, Oscar A MacLean, David Robertson – MRC-University of Glasgow Centre for Virus Research, University of Glasgow, Glasgow
Sergei L Kosakovsky Pond, Jordan D Zehr – Institute for Genomics and Evolutionary Medicine, Department of Biology, Temple University, Philadelphia
Venkatesh Kumar, Ben Murrell – Department of Microbiology, Tumor and Cell Biology, Karolinska Institutet, Stockholm, Sweden
Michael J. Stanhope – Department of Population and Ecosystem Health, College of Veterinary Medicine, Cornell University, Ithaca, NY

Nature and purpose:

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

Output in the last 12 months:

One article has been accepted in the journal Virus Evolution.

Future Direction:

This is an ongoing collaboration.

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

Collaborating Parties:

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

Nature and purpose:

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

Output in the last 12 months:

None.

Future Direction:

This is an ongoing collaboration funded by SAMRC for 2022.

7. HIV-1 Latent Viral Reservoir Dynamics

Collaborating Parties:

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

Nature and purpose:

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

Output in the last 12 months:

None.

Future Direction:

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

8. Phytovirus Diversity

Collaborating Parties:

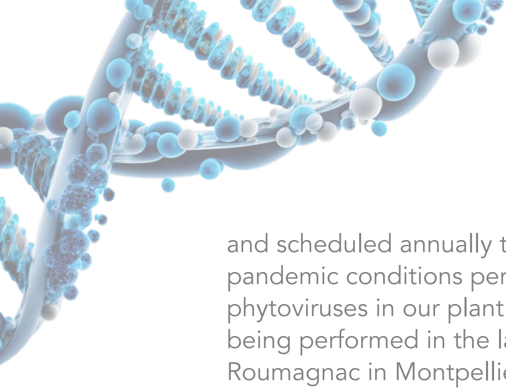
Philippe Roumagnac, Denis Filloux, Charlotte Julian and Emmanuel Fernandez – CIRAD, Montpelier, France.
Pierre Lefeuvre, Frederic Chiroleux and Jean-Michel Lett – CIRAD, UMR PVBMT, Réunion Island, France.
Thierry Candresse, Armelle Marais, Marie Lefebvre and Chantal Faure – INRA, UMR BFP, Villenave d'Ornon France.
Darren Martin – UCT, South Africa.
Arvind Varsani – Arizona State University, USA.
Pascal Gentit and Benoir Remenant – ANSES France.
Francois Mesleard and Hugo Fontes – La Tour du Valat, France.
Christophe Levergne – CBN CPIE Mascarin.
Anne Duputie and Francois Massol – UMR Evo-Eco-Paleo, France.

Nature and purpose:

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

Output in the last 12 months:

Plant samples were first collected in South Africa in 2019 but the field sampling trips planned for 2020 and 2021 was postponed until October 2022



and scheduled annually thereafter until 2025, pandemic conditions permitting. Efforts to detect phytoviruses in our plant samples are currently being performed in the laboratory of Philippe Roumagnac in Montpellier France.

Future Direction:

This is an ongoing project that is funded until 2023.

DR ULJANA HESSE

(Senior Lecturer, UWC Department of Biotechnology)

RESEARCH THEME

Rooibos Genomics

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 and 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 Program aims to

- locally establish all methodologies essential for medicinal plant genome analysis,
- generate a high-quality assembly of the rooibos genome with extensive functional annotation of the rooibos genes,
- identify rooibos genes involved in medicinal compound production and plant stress tolerance, and
- develop machine learning tools that can be trained to mine plant genomes for diverse genomic features (specific protein families, gene clusters, non-coding RNA, etc.).

Collaboration with SANBI: I have a longstanding partnership with SANBI. Since the initiation of the Rooibos Genomics Program, all postgraduate students of my research team have received training in bioinformatics through the 1-month course offered by SANBI. Establishment of biocomputational data analysis procedures is conducted or initiated at SANBI.

Progress 2022:

1. Published method for MinION sequencing and data analysis used to generate a high quality assembly of the 1.25Gbp rooibos genome.
2. Evaluated methods for repeat masking and completed repeat masking of the rooibos genome.
3. Optimised 3rd generation sequencing procedures for the analysis of plant transcriptomes using MinION from Oxford Nanopore.
4. Sequenced 12 transcriptomes from commercial and wild rooibos plants using MinION.
5. Established methods for gene prediction using hints from short and long read rooibos transcriptome data.
6. Established JBrowse for visualisation of the rooibos genome.
7. Established a collection of leaf samples from 105 rooibos ecotypes for population studies (9 rooibos growth types; 1-3 populations per growth type; 5 plants per population; 3 harvests: Feb 2022, Oct 2022, Feb 2023) which will permit comparative gene expression analyses between different growth types, years and before/after summer to identify genes associated with different morphological characteristics, biochemical profiles and with drought/heat stress tolerance.
8. Analysed all plant samples from the Feb 2022 harvest using HPLC to identify concentrations of 12 different rooibos phenolic compounds.
9. Established a greenhouse experiment with rooibos seedlings for comparative transcriptomics and proteomics associated with drought stress.
10. Established and evaluated a Convolutional Neural Networks foundation model for identification and classification of novel plant proteins to predict their functions.

Publication:

Mgwatyu, Yamkela, Stephanie Cornelissen, Peter van Heusden, Allison Stander, Mary Ranketse, and Uljana Hesse. "Establishing MinION Sequencing and Genome Assembly Procedures for the Analysis of the Rooibos (Aspalathus linearis) Genome."; Plants 11, no. 16 (2022): 2156.

RESEARCH PROJECT THEMES

Health data integration

- V-CAMM: Establishing a virtual cohort for African multi-morbidity, through the integration of longitudinal routine health data with genotype data for consenting participants.
- Epidemiological analysis of health and treatment profiles of public healthcare clients in the Western Cape, using routine health data

Facilitating sharing of African data and biospecimens

- ADBEx: The African Data and Biospecimen Exchange: an online platform to facilitate the ethical and equitable sharing of African datasets and biospecimen collections

Ethics and governance for research in Africa

Working group activities for international consortia and networks:

- International Common Disease Alliance (ICDA) Global Equity Working Group
- International 100k+ Cohorts Consortium (IHCC) Policy and Systems Working Group
- Public Health Alliance for Genetic Epidemiology (PHA4GE) Ethics and Data Sharing Working Group

PROF NICKI TIFFIN

HIGHLIGHTS OF 2022

Students

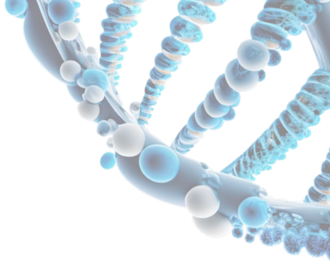
2022 was a busy year for the postgraduate students in the health data integration group: Ph D candidates Tsaone Tamuhla and Richard Osei-Yeboah, and MPH student Pierre Dane finished their research and completed their theses for submission at UCT at the beginning of 2023. Themba Mutemaringa and Florence Phelanyane continue part-time Ph D degrees under my supervision at the University of Cape Town, with Themba having his first research chapter accepted for publication in a peer-reviewed journal (in press in 2023). I continue to co-supervise a Ph D student at the University of Pretoria (Fatima Barmania), and at the University of Cape Town (Ryan Aylward), and co-supervised the MMed thesis of Graeme Hofmeyr which has also been completed and submitted to UCT.These students have a variety of projects working with large routine health datasets as well as conducting pilot studies to link genomic and routine health data. In addition to wrapping up her Ph D research work, Tsaone completed the first year of her MPH degree in 2022, passing her modules in the Epidemiology and Biostatistics track with flying colours.

We were also delighted to welcome Dr Eddie Lulamba to the team in 2022, to work on the ADBEx project, particularly in the first instance to undertake extensive work to understand the data standards, ontologies and controlled vocabularies that can be used for defining meta-data for sample collections and datasets. Eddie has an engineering and biotechnology background, and an interest in microbial genomics that complements the human genomics focus in the team. Ms Abiola Babjide has also joined the ADBEx project as a part-time administrator, in parallel with her ongoing Ph D programme at SANBI.

Funding

In 2022 we completed our first year of funding with the Bill & Melinda Gates Foundation, provided through the Calestous Juma Science Leadership Fellowship award to myself. This fellowship provides five years of funding for building the African Data and Biospecimen Exchange (ADBEx) as well as opportunities for networking and training in non-scientific expertise for the fellows.

We were also delighted to have our seed funding from UKRI/MRC reinstated after changes and disruptions to funding and awards from this programme over the last few years. This fund will support the development of processes, pipelines and a pilot for building a virtual genotyped multi-morbidity population cohort through the linkage of routine health data to genotype data for consenting participants in the Western Cape, South Africa. The research programme will kick off in June 2023, and Tsaone has started to pilot these processes with a proof-of-concept study together with



our service partners at the Central Analytical Facility at Stellenbosch.

Funding from the H3Africa Bioinformatics Network, H3ABioNet (www.h3abionet.org) came to a close in 2022. Luckily there are many excellent follow-on initiatives that will allow us to continue working with our valued African collaborators from this ten-year programme. We would like to take this opportunity to specifically recognise the funding from H3ABioNet that made it possible for Tsaone to undertake her Ph D in Bioinformatics at UCT, and which also provided me with funding for the original SANBI node in the first round of funding, and the subsequent public health and informatics-focused node that I have been leading at UCT and then UWC. The support, opportunities, learnings and strong collaborative partnerships we have had through H3ABioNet will help us to continue the work we are doing, going forward.

RESEARCH PROJECTS AND COLLABORATIONS

1. Multimorbidity in the Western Cape

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

1.1 Establishing a virtual population cohort using routine health data for epidemiological analyses

Nature and purpose: We are analysing anonymised and perturbed routine health data from a virtual cohort of all health care clients who attended any government health facility in the Khayelitsha sub-district in Cape Town during 2016 and 2017. Analysis of these data provides insights into multimorbidity in this population group that suffers from a high burden of both infectious and non- communicable diseases. In addition, we have continued to work in collaboration with the Provincial Health Data Centre (PHDC) at the Western Cape Department of Health in epidemiological and service-delivery projects working with routine health data from the Western Cape.

Researchers:

Richard Osei-Yeboah, Tsaone Tamuhla

Outputs in the past 12 months:

COVID-19 among adults living with HIV: Correlates of mortality in a general population in a resource-limited setting. medRxiv Preprint November 4, 2022. Kassanjee R, Davies MA, Ngwenya O, Osei-Yeboah R, Jacobs T, Morden E, Timmerman V, Britz S, Mendelson M, Taljaard J, Riou J, Boulle A, Tiffin N, Zinyakatira N

Future Directions:

Ongoing research in preparation for publication includes the use of longitudinal dispensing data to define complex phenotypes relating to persistence and adherence to diabetes medication (Tsaone Tamuhla).

1.2 Piloting a virtual genotyped population cohort linking genotype and routine health data:

Nature and purpose: With informed consent from health care clients, it is possible to link individuals' genotype data with their routine health data in order to establish a genotyped virtual cohort with complex clinical phenotype data that can be updated into the future using routine health records. We are piloting this approach by establishing the informed consent process, sample and data workflows to create, in the first instance, a virtual genotyped cohort for patients with diabetes in Cape Town. Funding from UKRI/MRC will fund a pilot study for proof-of-principle establishment of a virtual genotyped population cohort using this approach, as well as establishing the infrastructure to return actionable findings from the analysis to clinicians at the Western Cape Department of Health.

Researchers:

Tsaone Tamuhla, Nicki Tiffin.

Outputs in the past 12 months:

Tamuhla T, Tiffin N, Allie T. An e-consent framework for tiered informed consent for human genomic research in the global south, implemented as a REDCap template. BMC Med Ethics. 2022 Nov 24;23(1):119. doi: 10.1186/s12910-022-00860-2.

Multimorbidity Research in Sub-Saharan Africa: A Regional Retreat. Malawi-Liverpool Wellcome Trust Clinical Research Programme, Blantyre, Malawi. 22nd – 24th June 2022.

Nicki Tiffin, Talk: Exploring synergies between the health service and the research ecosystem: pragmatic multimorbidity research design

Tsaone Tamuhla, Talk: Piloting the creation of a genotyped virtual cohort.

Future Directions:

A proof-of-concept pilot project is ongoing to establish the analysis pipeline with an appropriate informed consent process, collection of buccal swabs and preparation of high quality DNA sample, generation and analysis of genotype data, and linkage to routine health data for phenotype definition (Tsaone Tamuhla). In addition we will be building infrastructure to ensure that actionable genetic findings from these studies can be appropriately returned to the Western Cape Department of Health to inform evidence-based care.

1.3 Optimising data linkage of African health client data

Nature and purpose:
When integrating individualised data from a variety of sources, different linkage algorithms can be used to ensure the most accurate linkage and deduplication of records that originate from the same individual. Most of these algorithms, however, have been developed using Euro-centric datasets and do not necessarily perform as well in African contexts, and we are working on a systematic approach to improve data linkage for routine health data. In 2022 we completed a descriptive analysis of the current iteration of the linkage algorithm at the PHCD at the Western Cape Department of Health.

Researcher:
Themba Mutemaringa

Collaborators:
Andrew Boulle, Alexa Heekes, Western Cape Department of Health

Outputs in the past 12 months:
International Population Data Linkage Network Conference, Edinburgh, United Kingdom, 7-9 September 2022. Poster and Conference Proceedings: Mutemaringa, Themba & Heekes, Alexa & Boulle, Andrew & Tiffin, Nicki. (2022). Record linkage for Routinely Collected Health Data in an African Health Information Exchange. International Journal of Population Data Science. 7. 10.23889/ijpds.v7i3.2022.

In Press: Record linkage for routinely collected health data in an African health information exchange. Mutemaringa *et al.*, International Journal of Population Data Science, 2022.

Future Directions:
In this project**,** Themba will continue to lead the validation of existing linkage and de-duplication algorithms, and will develop and test new methodologies and algorithms to best achieve de-duplication and data linkage. In addition, he is working on a thorough review of the governance protocols and ethical requirements for this ongoing work using routine health data at the PHDC.

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

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

Researchers: Nicki Tiffin, Eddie Lulamba, Tsaone Tamuhla.

Collaborators: Hominum Global, www.hominum.global – User Interface and Software Specification and Development

Outputs in the past 12 months: Technical specification for development of the ADBEx platform. In addition, an initiative by the Calestous Juma Fellows programme made it possible for Tsaone to attend the B&MGF Grand Challenges meeting with me in Brussels in September. This was a great opportunity to see other similar projects, and to present the ADBEx project.

Posters:
GA4GH 10th Plenary Meeting, Barcelona, September 2022: Poster presentation: Tamuhla T, Lulamba E, Babajide A, Tiffin N. African Data and Biospecimen Exchange (ADBEx) – a platform to facilitate equitable and ethical sharing of African data and biospecimens.

B&MGF Grand Challenges Meeting, Brussels, October 2022: Poster presentation: Tamuhla T, Lulamba E, Babajide A, Tiffin N. African Data and Biospecimen Exchange (ADBEx) – a platform to facilitate equitable and ethical sharing of African data and biospecimens.

Future Directions:
In year 2 of this programme, the development of the platform will commence, guided by the technical and front end. Two stakeholder workshops will be hold, engaging with researchers and other stakeholders working with human health genomics resources (workshop 1) and pathogen genomics resources (workshop 2).


(Photo: Davies Mbela, AAS Communications)

Calestous Juma fellows at a leadership workshop, with B&MGF program officers.

The Calestous Juma Fellowship

https://gcgh.grandchallenges.org/article/nicki-tiffin

In November 2021 the Bill & Melinda Gates Foundation launched the Calestous Juma Fellowship program for scientists permanently located at African research organisations. Nicki Tiffin was one of fourteen recipients of these fellowships, and is participating in a series of leadership trainings, networking opportunities and skills development as part of this program.
In 2022, she participated in leadership and communications training as well as participating in the B&MGF Grand Challenges meeting in Brussels. In addition she launched her own research project funded by this fellowship.

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

Nature and purpose:
In order to undertake ethical research in Africa and other regions of the global South, there are multiple considerations around risks, benefits, informed consent and equitable research. We explore different ways to better do research and to ensure that health research in Africa is ethical and equitable through the development of tools and frameworks.
I also participate in multiple ethics and governance working groups of international consortia including PHA4GE (Chair: Ethics and Data Sharing Working Group), ICDA (co-Lead: Global Equity Working Group) and IHCC (co-Chair: Systems and Policy Working Group). I have continued to lead the Ethics Working Group Forum, which brings together ELSI working groups from the many different large health genomics research networks and consortia in order to meet and share information about their activities, challenges and resources. This Forum met in person at the GA4GH meeting in Barcelona in

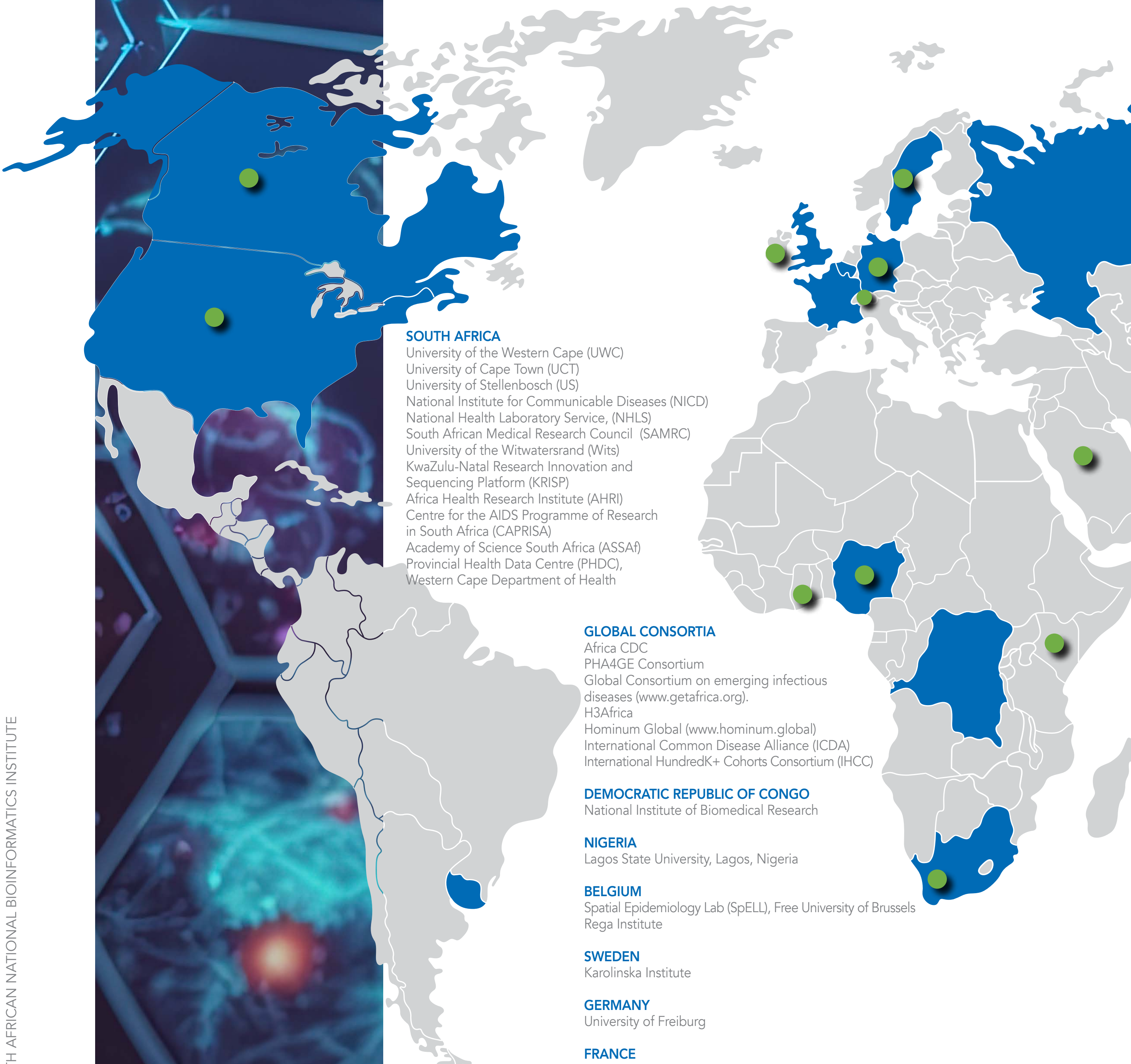
Outputs in the past 12 months:
Peer reviewed publication: Bedeker A, Nichols M, Allie T for the PHA4GE Ethics and Data-Sharing Working Group, *et al.* A framework for the promotion of ethical benefit sharing in health research BMJ Global Health 2022;7:e008096.
This work has also been referenced in the South African national press.
Why South Africa needs a national Data Transfer Agreement for health research
IN: Daily Maverick
Published: 19 June 2022
By: Mahomed *et al.*
https://www.dailymaverick.co.za/article/2022-06-19-why-south-africa-needs-a-national-data-transfer-agreement-for-health-research/

Future Directions:
Ongoing projects, with manuscripts in preparation include:

- A risk analysis framework for genomic health research
- A systematic review on global equity practices in human genomics research in Africa
- Development of a Global Equity standard (ICDA) for monitoring and evaluation of global equity practices in international networks and consortia

COLLABORATIONS

SANBI researchers have established a vast network of partnerships and collaborations with peers all over the world (indicated in blue) in the map below:



SOUTH AFRICA
University of the Western Cape (UWC)
University of Cape Town (UCT)
University of Stellenbosch (US)
National Institute for Communicable Diseases (NICD)
National Health Laboratory Service, (NHLS)
South African Medical Research Council (SAMRC)
University of the Witwatersrand (Wits)
KwaZulu-Natal Research Innovation and Sequencing Platform (KRISP)
Africa Health Research Institute (AHRI)
Centre for the AIDS Programme of Research in South Africa (CAPRISA)
Academy of Science South Africa (ASSAf)
Provincial Health Data Centre (PHDC),
Western Cape Department of Health

GLOBAL CONSORTIA
Africa CDC
PHA4GE Consortium
Global Consortium on emerging infectious diseases (www.getafrica.org).
H3Africa
Hominum Global (www.hominum.global)
International Common Disease Alliance (ICDA)
International HundredK+ Cohorts Consortium (IHCC)

DEMOCRATIC REPUBLIC OF CONGO
National Institute of Biomedical Research

NIGERIA
Lagos State University, Lagos, Nigeria

BELGIUM
Spatial Epidemiology Lab (SpELL), Free University of Brussels
Rega Institute

SWEDEN
Karolinska Institute

GERMANY
University of Freiburg

FRANCE
La Tour du Valat
ANSES (French Agency for Food, Environmental and Occupational Health & Safety)
CIRAD (Centre for International Cooperation in Development-Oriented Agronomical Research), Montpelier
CIRAD, UMR PVBMT (Plant Populations and Bio-aggressors in Tropical Ecosystems Joint Research Unit), Réunion Island
INRA (French National Agronomical Research Institute),
UMR BFP (Fruit Biology and Pathology), University of Bordeaux
CBN CPIE (National Botanical Conservatory Permanent Centre for Environmental Studies), Mascarin.
UMR Evo-Eco-Paleo, University of Lille

ALUMNI

SANBI has produced many alumni over the years who have taken their research to institutions and corporates all over the world (indicated in green) in the map below.

RUSSIA
Vavilov Institute of General Genetics, Russia

UNITED KINGDOM
Oxford University
University of Cambridge
University of Birmingham
MRC-University of Glasgow
Centre for Virus Research

UNITED STATES OF AMERICA
Washington University
USA CDC
Broad Institute
Johns Hopkins School of Medicine
NYU Langone Health
NYU Grossman School of Medicine
University of California San Diego
Temple University
Pennsylvania State University
Cornell University
Stanford University
Fred Hutchinson Cancer Research Center
Howard Hughes Medical Institute
Arizona State University

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

URUGUAY
Department of Biotechnological Development, Institute of Hygiene (IH), Faculty of Medicine (Fmed) of the University of the Oriental Republic of Uruguay (UdelaR)
Molecular Biology Laboratory, First Spanish Health Association, Montevideo
Department of Biodiversity and Genetics, Biological Research Institute, Clemente Stable, Montevideo

AUSTRALIA
University of Melbourne

SOUTH AFRICA
University of the Western Cape
University of KwaZulu-Natal
University of Cape Town
University of Stellenbosch
University of Pretoria
University of South Africa
University of the Free State
University of the Witwatersrand
SA Medical Research Council
National Institute of Communicable Diseases
Centre for Proteomic and Genomic Research
Centre for High Performance Computing
H3AbioNet

INDUSTRY
- Hyrax Biosciences
- IBM
- Roche
- B&M Scientific
- GENEdiagnostics
- City of Cape Town
- World Bank
- SA Government Administration
- KappaBiosystems
- Takealot
- Konga.com
- Monash
- Entersekt

KENYA
Egerton University, Njoro
International Livestock Research Institute

GHANA
University of Ghana

NIGERIA
Lagos State University
University of Nairobi

SAUDI ARABIA
King Abdullah University of Science and Technology

GERMANY
Institute for Inorganic Chemistry, RWTH Aachen University

UNITED STATES OF AMERICA
Harvard School of Public Health
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
University of British Columbia

IRELAND
National University of Ireland, Galway

SWEDEN
Karolinska Institute

SWITZERLAND
Syngenta Crop Protein AG, Basel

RUSSIA
Institute of Cytology and Genetics, Novosibirsk

SINGAPORE
Nanyang Technology University
University of Singapore

MALAYSIA
Novocraft Technologies Resear

FINANCIALS

SANBI's income and expenditure trends for 2022 are shown in this section.

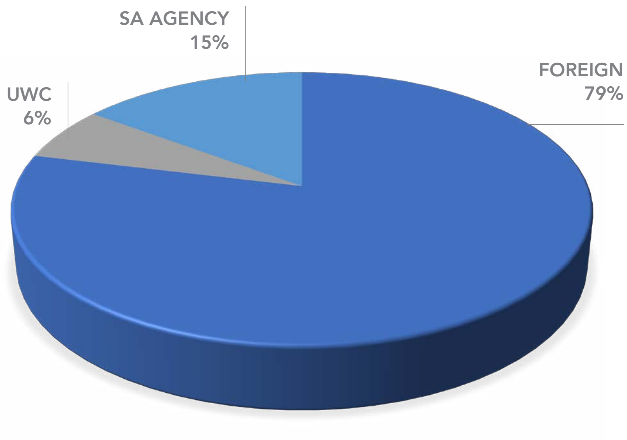


FIGURE 1. Distribution of income received from all sources 2022.

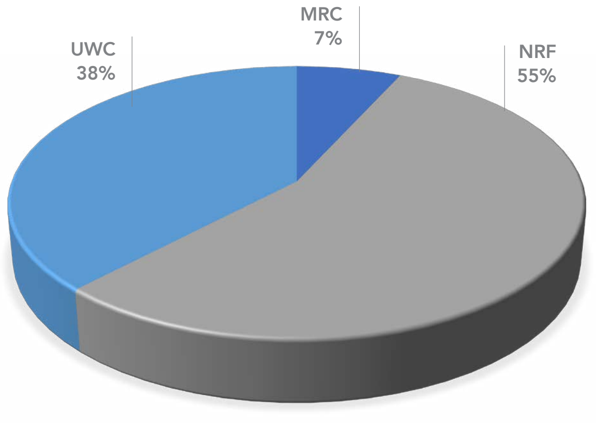


FIGURE 2. Distribution of income received from SA sources 2022.

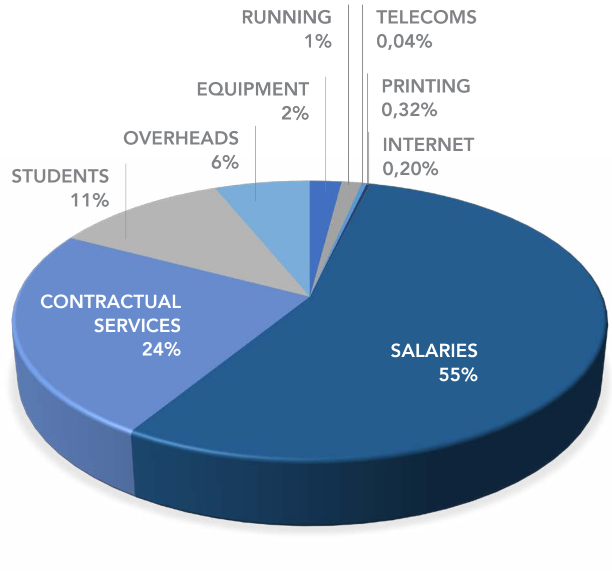


FIGURE 3. Expenditure in 2022.

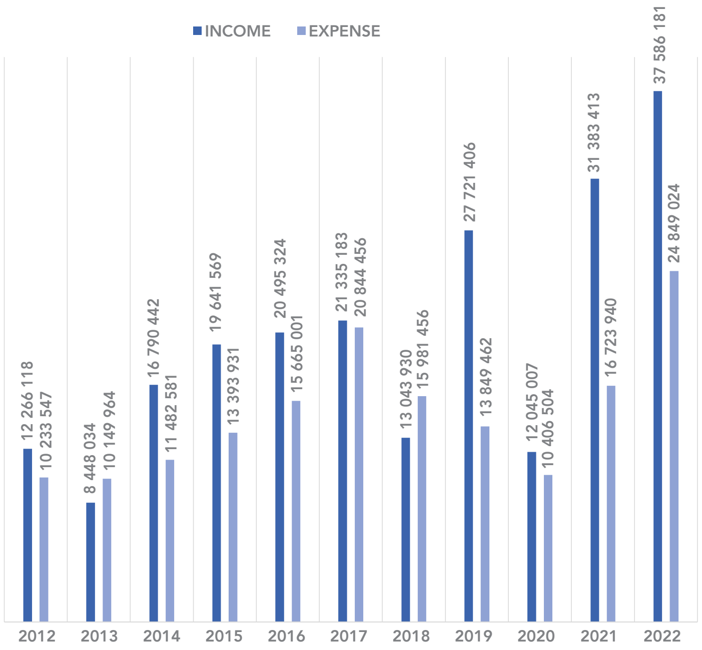


FIGURE 4. INCOME VS EXPENDITURE 2012 – 2022.

FUNDERS

SANBI
South African National Bioinformatics Institute

UNIVERSITY of the WESTERN CAPE

Research Institute at UWC since 1997

NRF National Research Foundation

National Research Foundation funding since 1998

saMRC advancinglife

South African Medical Research Council Bioinformatics Unit since 2002

science & innovation
Department:
Science and Innovation
REPUBLIC OF SOUTH AFRICA

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

H3ABioNet
Pan African Bioinformatics Network for H3Africa

H3Africa BioNetwork (H3ABioNet) Node 2020-2022

BILL & MELINDA GATES foundation

Bill & Melinda Gates Foundation funding since 2014

technology innovation AGENCY

Technology Innovation Agency funding since 2016

PRF
Poliomyelitis Research Foundation

Poliomyelitis Research Foundation Funding since 2019

AgriProtein

Agriprotein funding since 2019

Funding since 2021

World Health Organization

World Health Organization
REGIONAL OFFICE FOR Africa

African Union

AFRICA CDC
Centres for Disease Control and Prevention
Safeguarding Africa's Health

SANBI
South African National
Bioinformatics Institute

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

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

TELEPHONE: +27 (0)21 959 3645
EMAIL: info@sanbi.ac.za
WEBSITE: www.sanbi.ac.za
TWITTER: @SANBI_SA