



SANBI

South African National
Bioinformatics Institute

ANNUAL REPORT 2016



UNIVERSITY of the
WESTERN CAPE

FROM HOPE TO ACTION THROUGH KNOWLEDGE

Contents

About SANBI	2
Director's Message	4
Year in Review	5
Staff	9
Capacity Development	12
SANBI Graduations	17
Workshops and Courses	19
Computational Resources	23
Awards and Honours	25
SANBI in the Media	26
Research Outputs	29
Research Projects	36
Research Laboratories:	38
Prof Alan Christoffels	38
Prof Junaid Gamieldien	41
Dr Gordon Harkins	44
Dr Uljana Hesse	49
Prof Simon Travers	51
Research Collaborations	55
Financials	56
Alumni	58



About SANBI

Who we are

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

The South African National Bioinformatics Institute (SANBI) is situated at the University of the Western Cape (UWC) in Cape Town. Our primary focus is the development and implementation of computational methodologies which allow biomedical researchers to accelerate their genomics data analyses. SANBI aims to heighten awareness of bioinformatics in South Africa and to assist the country in making optimal use of this technology. As the leading bioinformatics entity on the African continent, 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 an online specialised resource for genomics and genome informatics;
- capacity development in genomics and bioinformatics in South Africa; and
- the development and implementation of genome annotation methods.

Our Vision

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

Our Mission

To conduct cutting edge bioinformatics and computational biology research relevant to South African, African and global populations.

To develop human resources in bioinformatics and computational biology by educating and mentoring scientists.

To increase awareness of and access to bioinformatics and computational biology resources.

Our Goals

To generate and publish high quality, relevant biomedical research.

To train and graduate competent and productive researchers.

To add value to the academic program of the University of the Western Cape.

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 (2012 – 2016)

The vision and mission of SANBI align with Draft Zero of the National Strategic Plan (NSP) which specifies Research and Innovation as a key enabler of the NSP, and argues that “relevant research provides information and the impetus for innovation within the implementation of the NSP” and that strategic priorities should include “concrete plans to improve capacity for research” and “a budget for research”.

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

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

The MRC Act (Act 58 of 1991)

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

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



Director's Message

The past year has seen SANBI substantially capitalise on human and technological capacity built during the past 15 years as a South African Medical Research Council (SAMRC) Bioinformatics Capacity Development Unit.

Notable achievements for 2016 include 20 peer-reviewed publications, multiple new funding streams generated, nine postgraduate degrees and several national and international awards recognising the institute's research excellence, including a TechWomen Award for the second year running. With 28 registered postgraduate students and postdoctoral fellows, co-supervision of students in other UWC departments and universities, and hosting of specialised training courses, our always-growing commitment to bioinformatics capacity development on the African continent has again been a hallmark of the year.

Our strategic and intentional investment into in-house next generation sequencing (NGS) skills and technology to meet our prediction of an approaching global bioinformatics bottleneck has again borne fruit in multiple application areas in 2016, with SANBI being able to rapidly remove stumbling blocks for multiple high profile projects, building new collaborations and strengthening existing ones in the process. Moreover, the quality of our research in this burgeoning niche was highlighted by the successful launch of an NGS focused spinoff company and validated by the rapid uptake of its flagship project by several HIV research and diagnostic laboratories worldwide.

2016 ended on a particularly high note, with the SAMRC committing to fund SANBI as an external Bioinformatics Unit on a 5-year cycle starting in 2017. We are excited about the year ahead and look forward to increasing our substantial momentum towards the advancement of national and international bioinformatics-driven basic research and applied science.

Prof Junaid Gamielien, PhD

*Acting Director: South African National Bioinformatics Institute
University of the Western Cape, South Africa*

Year in Review

Thanks to the commitment of academics, staff and students, SANBI has continued in the vein of producing excellent results in all areas of its mission, objectives and goals.

Research Highlights

Exatype

The official launch of exatype in February 2016 has led to many significant spin-off effects:

- research and diagnostic laboratories all over the world are using this computational tool routinely for their analysis.
- an application for a worldwide patent which is the second patent to arise from the exatype project.
- the exatype team, represented by Imogen Wright, was awarded second prize in the Innovation Prize for Africa, resulting in wide coverage in online and print media.
- Simon Travers and Imogen Wright met with Bill Gates and a global health team where they presented exatype and described plans of expanding into new disease lines and moving towards a point of care TB diagnostic and drug susceptibility test.
- Hyrax Biosciences (the exatype spinoff company) was invited to be a signatory on the Declaration by the Pharmaceutical, Biotechnology and Diagnostics Industries on Combating Antimicrobial Resistance that was launched at the World Economic Forum in Davos, Switzerland.
- Further, Simon Travers was a contributor to the final report of the UK Review on Antimicrobial Resistance that was launched in May 2016.



Imogen Wright and Simon Travers (far right back) with Bill Gates and members of the global health team.

Asian Seabass Genome

The SANBI team was successful in finding the solution that was finally adopted by the international consortium, when the genome annotation component of the project ran into trouble. This project is one of the few genomes with an assembly of such accuracy. A resultant journal article was published in *PLoS Genetics* and a website was developed at <http://seabass.sanbi.ac.za/>.

Awards

2016 has been another prolific year for SANBI in terms of recognition received for significant strides in the field of science and technology.

Alan Christoffels was awarded a **Fulbright Visiting Scholar Award** and spent 6 months at the Broad Institute-MIT, Boston, USA.

The **TechWomen Mentorship Programme** as well as the **Innovation Prize for Africa** were awarded to Imogen Wright.

Alan Christoffels was a finalist for the **NSTF-South32 Award**.

Funding

Institutional Funding

The SA Medical Research Council has renewed its 5-year cycle of funding to support SANBI as an external Bioinformatics Unit starting in 2017.

Research Project Funding

SAMRC/DST Strategic Health Innovation Partnership grants were awarded for:

- continuing development of the **exatype** platform to facilitate performing drug resistance testing locally on a point of care device;
- the development of a **TB** drug susceptibility testing pipeline and a real-time disease incidence and drug resistance surveillance platform; and
- development of a **clinical exome sequencing solution** and its application to diabetes mellitus and related disorders.

SAMRC Flagship

- **COMBAT-TB**: a multi- institutional project aimed at building an analytical platform that can be distributed across multiple sites to analyse *Mycobacterium tuberculosis* data. This platform will be piloted between SA and Nigeria in 2017.

NRF Research and Technology Fund and NRF Competitive Funding for Unrated Researchers

- Rooibos **transcriptome analyses**
- Rooibos **genome sequencing**

Internationalisation

In March 2016, in recognition of their productivity in terms of scientific publications (34 in five years), Gordon Harkins along with his collaborators from CIRAD, Montpellier (Dr Philippe Roumagnac) UCT (Associate Prof Darren Martin) and Stanford University (Dr Breynev Muhire) were invited to present their research findings at SANBI on the evolution and molecular epidemiology of important crop infecting viral pathogens to the **French ambassador to South Africa, Mrs Elisabeth Barbier**.

In May 2016, Junaid Gamielien, Gordon Harkins and Uljana Hesse hosted **Russian Federation Deputy Minister of Education and Science, Lyudmila Ogorodova**, at SANBI to identify and discuss possible areas of mutual interest.



Russian delegation meeting with SANBI.

The Brian O'Connell Fellowship for visiting African Scholars facilitates engagement between UWC staff and prestigious African scholars with a view of initiating closer interaction. SANBI hosted Prof Alash'le Abimiku from the University of Maryland School of Medicine and Director at the Institute of Human Virology Nigeria, during October - December 2016 as a recipient of this fellowship. Prof Abimiku has had interaction with Alan Christoffels during the past three years as they both served on an NIH-funded biobank consortium. During her visit, Prof Abimiku facilitated discussions between SANBI staff and the Center for Disease and Control, USA, and explored bioinformatics training initiatives for Nigerian students to gain bioinformatics experience at SANBI.

Research Outputs

2016 was a productive year in terms of research outputs. **20 Publications** appeared in **16 journals** and a total of **12 publications** had **first or last authors**. One **conference proceeding** was also published.

Capacity Development

The student cohort comprised **6 Postdoctoral** fellows, **12 Doctoral** and **10 Masters** students in 2016.

Of the total 28 SANBI students, **50% are female** and **93% are South African**.

SANBI graduated **3 PhD**, **4 MSc** and **2 Honors** degree students in 2016.

9 Training courses or workshops were organised or hosted by SANBI.

Alumni

SANBI's reputation of providing excellent graduate education and international exposure to its students is reflected in the positions which recent graduates and alumni have taken up in corporate and academic spheres:

Michael Berry joined Kapa Biosystems (acquired by Roche) as a Bioinformatics Scientist.

Rosaline Macharia joined the University of Nairobi as a Lecturer in 2016.

Darlington Mapiye is a Research Scientist in the Data Driven Healthcare department of IBM Research Africa in Johannesburg.

Mahjoubeh Jalali joined the Pacific Northwest Diabetes Research Institute in Seattle, Washington, USA as a Postdoctoral Fellow.

Zahra Jalali is currently a Postdoctoral Fellow at the Centre for Molecular Medicine and Therapeutics at the University of British Columbia, Canada.

Imogen Wright is a SANBI Postdoctoral Fellow and co-founder of Hyrax Biosciences.

Colleen Saunders is a Junior Research Fellow in the Division of Emergency Medicine at the University of Cape Town.

Staff

As a research institute, the Director of SANBI reports through the faculty of Natural Sciences at the University of the Western Cape. SANBI staff are made up of a diverse group of scientists and administrative staff all contributing towards a dynamic productive working environment.

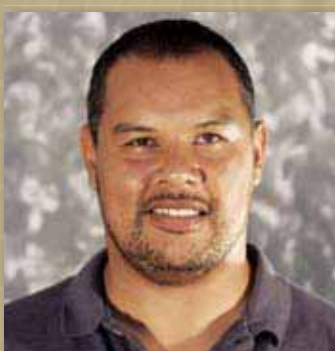
Academic staff are tasked with research, student graduations and capacity development. The computing infrastructure is maintained by a team of technical staff while administrators ensure the smooth running of daily operations.

Academic



Alan Christoffels, Prof

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



Junaid Gamiieldien, Prof

Associate Professor
UWC



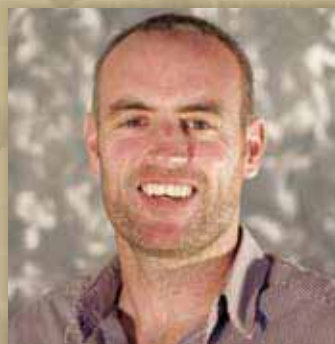
Gordon Harkins, Dr

Senior Lecturer
UWC



Uljana Hesse, Dr

Senior Lecturer
UWC



Simon Travers, Prof

Associate Professor
UWC

Technical



Hocine Bendou
Software Developer
NIH (H3Africa)



Phillip Labuschagne
Senior Software Developer
HVTN funding from the Bill and Melinda Gates Foundation



Thoba Lose
Software Developer
SA Medical Research Council



Baruch Lubinsky
Software Developer
DST-HIV funding



Zipho Mashologu
Software Developer
SA Medical Research Council



Campbell Rae
Web Developer
DST/NRF Research Chairs Programme



Peter Van Heusden
Senior Systems Developer
UWC



Long Yi
Software Developer
SA Medical Research Council

Administration



Fungiwe Mpithi

Reception
SA Medical Research Council



Ferial Mullins

Finance Administrator
UWC Dean's budget



Maryam Salie

Student Administrator
B3Africa



Natasha Schoeman

PA and HR Administrator
*DST/NRF Research Chairs
Programme*



Junita Williams

Administrator part-time
*DST/NRF Research Chairs
Programme*

Capacity Development

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

Students are expected to publish their work in peer-reviewed journals and to present at laboratory meetings and to wider audiences. All students participate in journal club meetings where relevant publications are prepared, presented and discussed.

The support provided by SANBI exposes students to internationally competitive research environments.

Undergraduate Training Programme

Bioinformatics Module (BTN 315)

The UWC undergraduate Bioinformatics Module is taught to approximately 110 third year students by Drs Gordon Harkins, Uljana Hesse, Mahjoubeh Jalali and Simon Travers from SANBI. Students attend lectures, practicals and tutorials on the following topics:

- Databases and their Applications in Disease Genomics;
- Introduction to Molecular Evolution and Phylogenetics;
- Comparative Genomics; and
- Alignments and Database Searching.

Honours 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. Two students, Lynley Abdoll and Warren Jacobus, graduated with an Honours Degree after working with SANBI PIs on their thesis projects and they have subsequently registered with SANBI for a Masters degree in 2016.

Internship Programme

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

In 2016, SANBI hosted Yamkele Mgwatya as an intern in the lab of Uljana Hesse.

Postgraduate Training Programme

Masters Programme

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



Lynley Abdoll
Alan Christoffels



Toluwaleke Ademuyiwa
Alan Christoffels



Olabode Ajayi
Alan Christoffels



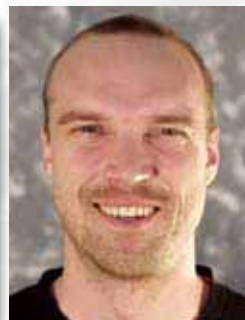
Eugene De Beste
Alan Christoffels



Ereshia Gabier
Simon Travers



Warren Jacobus
Alan Christoffels



Phillip Labuschagne
Simon Travers



Clint Mercur
Simon Travers



Lunga Sizani
Alan Christoffels



Peter Van Heusden
Alan Christoffels

**Supervisor in italics*

Doctoral Programme

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



Ibrahim Ahmed
Alan Christoffels



Hocine Bendou
Alan Christoffels



Tracey Calvert-Joshua
Alan Christoffels



**Mmakamohelo
Direko**
Alan Christoffels



Roux-Cil Ferreira
Simon Travers



Kay Fortuin
Alan Christoffels



Bridget Langa
Junaid Gamieldeen



Anati Nkaule
Alan Christoffels



Catherine Rossouw
Alan Christoffels



Emily Stander
Uljana Hesse



Emil Tanov
Gordon Harkins



Larry Van Vuuren
Nicki Tiffin

Postdoctoral Programme

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



Ruben Cloete
Alan Christoffels



Samuel Egieyeh
Alan Christoffels



Mahjoubeh Jalali
Junaid Gamielien



Zahra Jalali
Alan Christoffels



Colleen Saunders
Junaid Gamielien



Imogen Wright
Simon Travers

International Training Opportunities

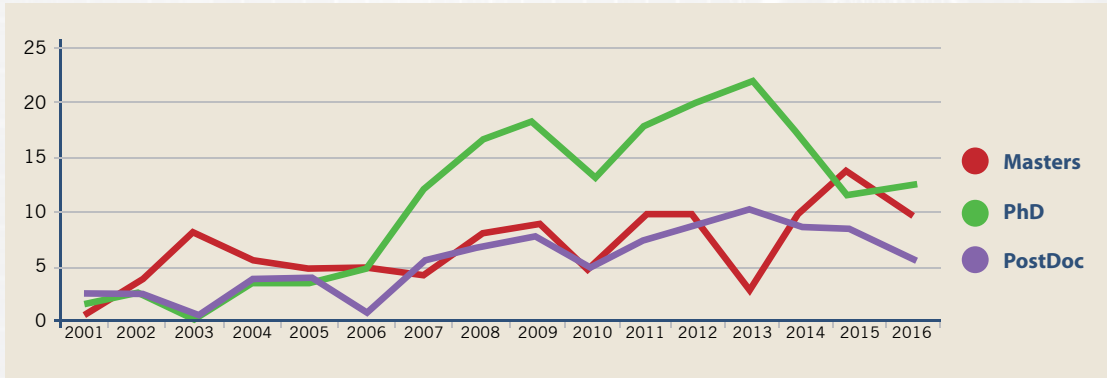
SANBI students enjoy the benefit of internships at international research labs. These internships usually occur as a result of a collaborative research project.

In April 2016, PhD Student Phillip Labuschagne visited the Fred Hutchinson Cancer Research in Seattle, Washington for 2 weeks to work with a collaborator on the development of methods for the analysis of sequence data from HIV vaccine trials.

2016 SANBI Postgraduate Registration





In 2016 the student cohort comprised of 6 Postdoctoral fellows, 12 Doctoral and 10 Masters students.





SANBI Postgraduate Registration Trend 2001 – 2016:



SANBI Graduations

SANBI has great pleasure in announcing the following graduates for 2016:

Student Name	Thesis Title
 <p>Samuel Egieyeh PhD <i>Alan Christoffels</i></p>	<p>Computational strategies to identify, prioritize and design potential antimalarial drugs from natural products.</p>
 <p>Rosaline Macharia PhD <i>Alan Christoffels</i></p>	<p>Characterisation of olfactory responsive genes in selected Tsetse Species: Annotation and comparative analyses of chemosensory proteins in the Genus Glossina.</p>
 <p>Darlington Mapiye PhD <i>Junaid Gamiieldien</i></p>	<p>Computational genomics approaches for kidney diseases in Africa.</p>
 <p>Tracey Calvert-Joshua MSc <i>Nicki Tiffin</i></p>	<p>Integrating regulatory and epigenetic data for clear cell renal cell carcinoma variant discovery.</p>
 <p>Eugene Madzokere MSc cum laude <i>Gordon Harkins</i></p>	<p>The phylogeography, epidemiology and determinants of maize streak virus dispersal across Africa and adjacent Indian Ocean islands.</p>

Student Name	Thesis Title
 <p>Stephanie Pitts MSc <i>Nicki Tiffin</i></p>	<p>An Investigation into the MicroRNA-gene interactions involved in the pathogenesis of Systemic Lupus Erythematosus.</p>
 <p>Werner Veldsman MSc cum laude <i>Alan Christoffels</i></p>	<p>SNP based literature and data retrieval.</p>
 <p>Lynley Abdoll BSc Medical Biosciences (Honours) <i>Alan Christoffels</i></p>	<p>The effects of single nucleotide polymorphisms (SNPs) on the functionality of human arylamine N-acetyltransferase1 (NAT1).</p>
 <p>Warren Jacobus BSc Computer Science (Honours) <i>Alan Christoffels</i></p>	

Workshops and Courses

In keeping with its mission to develop human resources in bioinformatics and computational biology, SANBI faculty are involved in numerous capacity development efforts in the form of workshops and courses throughout the year.

Organised/Hosted by SANBI

Introduction to Linux + SANBI queues + Python course

This 5-day workshop was presented by Peter van Heusden from 07 – 11 March to graduate students in bioinformatics and other researchers with the aim to teach them basic lab skills for scientific computing. This hands-on workshop covered basic concepts and tools, including program design, version control, data management and task automation.

ENSEMBL

On 6 June SANBI hosted a one-day workshop on the Ensembl Genome Browser. Presented by Dr Ben Moore from the Vertebrate Genomics Team at the European Biotechnology Institute, this course was partially sponsored by the Department of Science and Technology through the Bioinformatics Support Platform.

Basic probability theory/statistics and the R language

SANBI hosted the first of a two-part course on statistics and the R statistical programming language (<https://www.r-project.org/>) from 18 - 22 July. The first part of the course was aimed at providing a basis for participants to understand and start performing statistical analyses in R. The course was presented by Jean-Baka Domelevo-Entfellner and was attended by 25 participants from UWC, CPUT, University of Stellenbosch and UCT.

Software Carpentry Seminar and Workshop

On 29 August, Michael Crusoe, one of the founders of the Common Workflow Language project presented a seminar at SANBI on CWL and facilitated discussions on building portable, reproducible scientific workflows.

The seminar was followed by a hands-on workshop held 30 – 31 August, presented by Peter Van Heusden and Warren Jacobus. The workshop was aimed at post-graduate students, postdoctoral research fellows, and researchers affiliated with UWC and surrounding institutions. Beginner-level material suitable for programming novices was taught.



Participants at the Gene Annotation Workshop.

2-Day Gene Annotation Workshop

From 05 – 06 December, Robert Waterhouse from the Swiss Bioinformatics Institute brought together 20 participants from SANBI, University of Stellenbosch, UWC and UCT to learn about arthropod genomics and biocuration of gene and genome data. The programme focused on practical training in biocuration skills that helped the participants to make the most of the publically-available genomic resources such as at VectorBase (www.vectorbase.org) to further their own research projects.

H3AfricaBionet Activities

SANBI continues to be actively involved in providing training through the H3ABioNet workshops around Africa. SANBI node members were particularly active as lecturers and trainers at the following training interventions.

H3ABioNet Introduction to Bioinformatics online course (IBT_2016)

From 06 July – 09 October Colleen Saunders, Ruben Cloete and Jean-Baka Domellevo-Entfellner of SANBI taught different aspects of the online course. The course was aimed at individuals with a basic understanding of genetics and/or biochemistry who would like to become basic bioinformatics users.

Over the 3 month period, approximately 350 participants in 19 classrooms across 11 African countries focussed on bioinformatics tools, algorithms and resources in the twice weekly theoretical and practical sessions.

B3Africa Consortium

Funded by the EU Horizon 2020, the purpose of the Bridging Biobanking and Medical Research Across Europe and Africa (B3Africa) initiative is to implement a cooperation platform and technical informatics framework for biobank integration between Africa and Europe.

B3Africa 1st Technical Jamboree

From 03 - 06 May, 16 participants from 6 institutions in South Africa and Europe attended the first B3Africa Technical Jamboree at SANBI to propose and test solutions for the integration and implementation of the technical platform, the eB3Kit.



Participants of the B3Africa 1st Technical Jamboree at SANBI offices.

B3Africa Open Forum

SANBI co-convened this forum with the National Health Laboratory Services (NHLS) and Stellenbosch University Biobank (NSB) on 09 December at the University of Stellenbosch Tygerberg Medical School.

The goal of this meeting was to bring together experts in the fields of biobanking, laboratory information management systems (LIMS) and Ethical, Legal and Social Issues (ELSI) to discuss concepts such as data/information security, ethical and regulatory issues, harmonization of data standards, and biosecurity in resource limited settings. Alan Christoffels presented a talk entitled "Data standards for Biobanking".

B3Africa Site Visit

This workshop was held from 12 – 14 December at the University of Stellenbosch Tygerberg Medical School with the aim of training users on biobanking tools. Countries represented by the 21 participants include Switzerland, Sweden, Austria, Spain, Ghana, The Gambia, Nigeria

and South Africa. Hocine Bendou presented his PhD project that developed a Laboratory Information Management System called Baobab LIMS.



B3Africa site visit to Tygerberg Medical School. LIMS training for African partners.

Invitations to Lecture

From time to time, SANBI academics are invited to present lectures at a variety of events. From 26 – 29 January the **Forum of University Nursing Deans in South Africa (FUNDISA)** held a workshop at the Lord Charles Hotel, Somerset West. This NGO receives funding for capacity development in nursing scholarships. Alan Christoffels taught a module on 'Managing your Research Project'. 17 Participants attended from SA institutions.

Interns hosted at SANBI

Internships provide students the opportunity to gain hands-on experience at SANBI.

Student Name	Institution	Date	Purpose
Johan Sträßer	University of Applied Science in Gießen, Germany	January – March	Undertook his final year research project at SANBI
Yamkela Mgwatyu	DST-NRF Internship	April – December	Student training in plant sampling and molecular biology techniques
Tyler N Hartel	3rd year BTY student	September	Student training in isolation of fungal endophytes from rooibos
Danielle A Davids	3rd year BTY student	September	Student training in isolation of fungal endophytes from rooibos
Min-Ghah Kariem	3rd year BTY student	September	Student training in isolation of fungal endophytes from rooibos

Computational Resources

In 2015, we started to benchmark state of the art scalable storage solutions in an attempt to cope with the ever-increasing collaborative computational biology projects. With much excitement we have rolled out a storage solution at SANBI that will undoubtedly be the reference standard in the South African bioinformatics community. The impact of our customizable environment has lead to an invitation for Alan Christoffels to visit the Wellcome Trust-Funded Malaria Training initiative in Mali to explore how SANBI can contribute to African initiatives aimed at building similar bioinformatics capacity.

Our sophisticated computational environment comprise a compute cluster, virtual machines, storage and backup, and technical personnel as described below.

SANBI's core server infrastructure consists of a set of Dell blade servers with a total of 232 CPU cores.

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

In addition to the blade servers, a R710 rack mounted server, with 12 CPU cores and 24 GB of RAM, is used as a storage controller. Storage is provided by a Dell EqualLogic storage array that hosts 30 TB of SATA disk based storage and a Ceph storage cluster which hosts 72 TB distributed storage. The servers and the storage array are interconnected with 10 Gb Ethernet. Internet connectivity is provided by a 30 Mb/s broadband connection, and SANBI is also connected to the SANReN network that provides 10Gb/s connections between South African universities and research centers.

SANBI Compute Cluster

We continued to grow our computing capabilities in 2016, with the addition of 2 new servers that provide an extra 32 CPU cores and 192 GB of RAM to our cluster. We now have 232 CPU cores and 1952 GB of RAM on our Dell cluster.

Virtual Machine Infrastructure

The SANBI VM cloud has been migrated to a new virtual machine management platform developed by Long Yi. SANBI currently hosts 113 virtual machines using 184 virtual CPU cores and 468 GB of RAM and 6.7 TB of disk space, spread across 9 virtual machine hosts. This is used to host key SANBI infrastructure as well as web applications presenting SANBI research to the outside world.

Networking and Internet

We have upgraded our broadband connection to 30 Mbit/s, doubling our internet speed. We also completed the installation of the 10GbE fibre link to the main UWC ICS datacenter.

Storage and backup

The Ceph storage array we had been planning since 2014 arrived in 2015, providing 72 TB of extra storage. This storage array will make our disk usage as scalable as our computing, and we have already added a third storage server to the initial two, which we expect will provide another 60 TB of disk space. The storage expansion has boosted our ability to do science as we no longer need to worry so much about the disk space an analysis will consume. The new storage has also relieved pressure on our older Dell EqualLogic SAN, ensuring that we don't have work-interrupting full disks. We also installed a Backblaze storage server, providing 60 TB of disk based backup to keep SANBI data safe.

Software Development



Software developers at SANBI have been active contributors to the Galaxy project. Galaxy, which promises “data intensive science for everyone” is an open source science gateway and workflow engine for bioinformatics. SANBI has embraced Galaxy as part of the COMBAT-TB project, and contributed to both code and documentation for the project. We have also added support for working with Neo4j Graph databases within Galaxy.

Peter van Heusden and Thoba Lose also presented our work on a Graph database model for genome annotation to the Galaxy Community Conference (GCC) and GMOD community meetings in Bloomington, Indiana, USA. Through this engagement our development team has established a collaboration with others working in the field of graph based knowledge representation including members of the SciGraph team.

Whilst at GCC Peter and Thoba gave talks on Graph databases for genome annotation: Peter gave a talk at GCC on the transition from RDBMS to Graph databases, delving on the lessons learnt from previous projects (Seabass). Thoba gave a talk at the GMOD meeting on Graph databases for genome annotation, demo-ing the Galaxy IE and the COMBAT-TB Graph database built at SANBI.

The team has also established a working relationship with the team working on SciGraph and the Monarch Initiative to establish a standard graph-based genome annotation model.

Mentorship

SANBI computing staff helped with mentoring the UWC Student Cluster Team that went on to win the Centre for High Performance Computing national Student Cluster Competition in East London in December 2016, securing their entrance to the International Student Cluster Competition in Germany in 2017.

As of 2016, two out of four students that we mentored for the 2014 Student Cluster Competition team were SANBI students, showing the long term benefit of our mentoring activities.

Awards and Honours

Fulbright Visiting Scholar Award

From February – September Alan Christoffels took up this prestigious award at the Broad Institute-MIT in Boston, USA.

Innovation Prize for Africa

Imogen Wright won 2nd prize in the Innovation Prize for Africa competition for her work in the development of the HIV drug resistance testing pipeline exatype. Of the 985 applications, the 10 finalists travelled to Botswana in June for the prize giving.



Innovation Prize of Africa winners Dr Imogen Wright, Dr Eddy Agbo and Dr Valentin Agon, with African Innovation Foundation founder Jean Claude Bastos de Morais, Botswana President President Seretse Khama Ian Khama and Botswana Minister of Infrastructure, Science and Technology Nonofa Molefhi.

Photo credit: <https://www.forbes.com/sites/tobyshapshak/2016/06/23/malaria-hiv-aids-solutions-win-big-at-innovation-prize-for-africa/#67ff7ca9779a>

TechWomen Mentorship Programme

Imogen Wright is one of the winners of the programme which had more than 2300 applicants from Africa, Central Asia and the Middle East . The programme aims to empower, connect and support the next generation of female leaders in Science, Technology Engineering and Mathematics (STEM).

TW Kambule-NSTF Award

Alan Christoffels received the award for an outstanding contribution to Science, Engineering and Technology and innovation by an individual over a period of up to 15 years after completion of a PhD or equivalent, predominantly in South Africa.

SANBI in the Media

African Research Cloud takes shape

ITWEB, 03 November

South Africa's Inter-University Institute for Data-Intensive Astronomy (IDIA) has become the first African institute to launch a cloud-based data centre.

Full article at:

http://www.itweb.co.za/index.php?option=com_content&view=article&id=157246:African-Research-Cloud-takes-shape

Analysing Scientific Data: Many opportunities for bioinformaticists)

CAREER TIMES, 31 October

Prof Junaid Gamielien was featured in the 'Real jobs, real people' column. The article focused on how to become a bioinformaticist and covered, amongst others: details of the job, an average working day, best/worst parts of the job, and qualifications needed.

Exatype: Cloud for HIV Drug Resistance Testing

AMAZON WEB SERVICES, 26 September

Professor Simon Travers and his team of researchers based at the South African National Bioinformatics Institute (SANBI) at the University of the Western Cape (UWC) have launched the Exatype platform, which runs on AWS. The Exatype platform provides rapid, accurate HIV drug resistance analysis at affordable rates. Users upload HIV sequence data files directly from high-throughput sequencing machines and receive a detailed drug resistance genotype report for each sample.

Full article at:

<https://aws.amazon.com/blogs/publicsector/exatype-cloud-for-hiv-drug-resistance-testing/>

SA scientist, engineer pioneer tech in TB research

FIN24, 05 September

Johannesburg - A simple tag pioneered in South Africa could soon be used to curb the spread of tuberculosis in Africa.

In efforts to better understand how the disease is spread, IBM engineer Toby Kurien and research scientist Darlington Mapiye have developed the concept of a tracking device which measures the proximity of TB patients.

Full article at:

<http://www.fin24.com/Tech/News/sa-scientist-engineer-pioneer-tech-in-tb-research-20160902>

Cloud tech aids HIV drug resistance testing

ITWEB, 25 July

Local scientists have developed a technological innovation, underpinned by cloud computing, which will assist in testing for HIV drug resistance in SA.

Professor Simon Travers and his team of researchers based at the South African National Bioinformatics Institute (SANBI) at the University of the Western Cape (UWC) have launched the Exatype platform, which runs on Amazon Web Services (AWS).

Full article at:

http://www.itweb.co.za/index.php?option=com_content&view=article&id=154531:Cloud-tech-aids-HIV-drug-resistance-testing&catid=86

Malaria, HIV/Aids Solutions Win Big At Innovation Prize for Africa

FORBES.COM, 23 June

A ground-breaking anti-malarial drug made from natural plant extracts has won the 2016 Innovation Prize for Africa (IPA); while software to combat HIV drug-resistance medication and a cheap, simple urine test for malaria were the other big winners.

Full article at:

<http://www.forbes.com/sites/tobyshapshak/2016/06/23/malaria-hivaids-solutions-win-big-at-innovation-prize-for-africa/#283043356f35>

UWC team eyes drug innovation prize

CAPE TIMES, 15 May

A UWC team's development of a programme which rapidly, accurately and cost-effectively tests HIV drug resistance is a top 10 finalist for the Innovation Prize for Africa (IPA).

Widely recognised as the premier award for African innovation, the IPA has attracted 958 submissions from 46 African countries which have been considered for the 2016 prize.

Full article at:

<http://www.iol.co.za/capetimes/uwc-team-eyes-drug-innovation-prize-2021884>

Exatype assists in efficiency of HIV AIDS antiretrovirals

ANN7, 13 May

Full interview at:

<http://www.ann7.com/exatype-assists-in-efficiency-of-hiv-aids-antiretrovirals/>

SA's breakthrough medical tech makes African cut

ITWEB, 10 May

The African Innovation Foundation (AIF) yesterday named two South African doctors among the top 10 nominees for its landmark programme, the Innovation Prize for Africa (IPA). The South African nominees are Dr Imogen Wright, co-founder of Hyrax Biosciences, and Dr Kit Vaughan, chief executive of Cape Ray.

Full article at:

http://www.itweb.co.za/index.php?option=com_content&view=article&id=152360:SA-s-breakthrough-medical-tech-makes-African-cut&catid=86

10 nominees announced for AIF's Innovation Prize for Africa 2016

INNOVATION PRIZE FOR AFRICA, 9 May

The African Innovation Foundation (AIF) today announced the top 10 nominees for its landmark programme, the Innovation Prize for Africa (IPA). Now celebrating its 5th year under the theme "Made in Africa", IPA is the premier innovation initiative in the African continent, offering a grand share prize of US\$150 000 and incentives to spur growth and prosperity in Africa through home-grown solutions.

Full article at:

<http://innovationprizeforafrica.org/10-nominees-announced-for-aifs-innovation-prize-for-africa-2016-2>

Exclusive: Real Solutions To Real Problems Named In Innovation Prize For Africa 2016 Finalists

FORBES, 7 May

An anti-malaria drug made from a local plant, a 25-minute test for malaria, an organic low-cost fertiliser, software to determine which ARV drugs will be most effective, imaging technology to dramatically improve breast cancer detection and solar-powered water heating are among the 10 finalists named for the 2016 Innovation Prize for Africa (IPA).

Full article at:

<http://www.forbes.com/forbes/welcome/?toURL=http://www.forbes.com/sites/tobyshapshak/2016/05/07/exclusive-real-solutions-to-real-problems-named-in-innovation-prize-for-africa-2016-finalists/&refURL=&referrer=>

UWC HIV drug resistance test innovation

CAPE TIMES, 18 February

Cape Town. Prof Simon Travers and his team of researchers based at the South African National Bioinformatics Institute (SANBI) at the University of the Western Cape (UWC) developed a programme called Exatype. The DNA of the HIV virus analysed using Exatype to quickly and accurately produce an easy-to-interpret report showing the clinician which drugs would be most effective for that individual patient at that time.

Full article at:

<http://www.iol.co.za/capetimes/uwc-hiv-drug-resistance-test-innovation-1986693>

Research Outputs

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

Journal Publications

Congratulations to the **first authors** on their publications!

#	Publication (SANBI contributors)	Published Date
1.	<i>Exploration of Scaffolds from Natural Products with Antiplasmodial Activities, Currently Registered Antimalarial Drugs and Public Malarial Screen Data.</i> Egieyeh S. , Syce J, Christoffels A , Malan SF. Molecules. 2016 Jan 16;21(1):104. doi: 10.3390/molecules21010104	16 January
2.	<i>Semantic interrogation of a multi knowledge domain ontological model of tendinopathy identifies four strong candidate risk genes.</i> Saunders CJ. Jalali Sefid Dashti M, Gamielidien J. Scientific Report 2016 Jan 25;6:19820. doi: 10.1038/srep19820	25 January
3.	<i>Prioritization of anti-malarial hits from nature: chemo-informatic profiling of natural products with in vitro antiplasmodial activities and currently registered anti-malarial drugs.</i> Egieyeh SA. Syce J, Malan SF, Christoffels A . Malaria Journal. 2016 Jan 29;15:50. doi: 10.1186/s12936-016-1087-y.	29 January
4.	<i>The Development of Computational Biology in South Africa: Successes Achieved and Lessons Learnt.</i> Mulder NJ, Christoffels A , de Oliveira T, Gamielidien J , Hazelhurst S, Joubert F, Kumuthini J, Pillay CS, Snoep JL, Tastan Bishop Ö, Tiffin N. PLoS Computational Biology. 2016 Feb 4;12(2):e1004395. doi: 10.1371/journal.pcbi.1004395.	04 February
5.	<i>Resistance related metabolic pathways for drug target identification in Mycobacterium tuberculosis.</i> Cloete R. Oppon E, Murungi E, Schubert WD, Christoffels A . BMC Bioinformatics. 2016 Feb 8;17:75. doi: 10.1186/s12859-016-0898-8	08 February
6.	<i>Genome-Wide Comparative Analysis of Chemosensory Gene Families in Five Tsetse Fly Species.</i> Macharia R. Mireji P, Murungi E, Murilla G, Christoffels A , Aksoy S, Masiga D. PLoS Negl Tropical Diseases. 2016 Feb 17;10(2):e0004421. doi: 10.1371/journal.pntd.0004421.	17 February

#	Publication (SANBI contributors)	Published Date
7.	<p>NetCapDB: measuring bioinformatics capacity development in Africa.</p> <p>Bendou H, Entfellner JB, van Heusden P, Gamielidien J, Tiffin N; H3ABioNet Consortium, as members of the H3Africa Consortium.</p> <p>BMC Research Notes. 2016 Mar 5;9:144. doi: 10.1186/s13104-016-1950-5</p>	05 March
8.	<p>Molecular characterization and prevalence of two capulaviruses: Alfalfa leaf curl virus from France and Euphorbia caput-medusae latent virus from South Africa.</p> <p>Bernardo P, Muhire B, François S, Deshoux M, Hartnady P, Farkas K, Kraberger S, Filloux D, Fernandez E, Galzi S, Ferdinand R, Granier M, Marais A, Monge Blasco P, Candresse T, Escriu F, Varsani A, Harkins GW, Martin DP, Roumagnac P.</p> <p>Virology. 2016 Jun;493:142-53. doi: 10.1016/j.virol.2016.03.016.</p>	31 March
9.	<p>Chromosomal-Level Assembly of the Asian Seabass Genome Using Long Sequence Reads and Multi-layered Scaffolding.</p> <p>Vij S, Kuhl H, Kuznetsova IS, Komissarov A, Yurchenko AA, Van Heusden P, Singh S, Thevasagayam NM, Prakki SR, Purushothaman K, Saju JM, Jiang J, Mbandi SK, Jonas M, Hin Yan Tong A, Mwangi S, Lau D, Ngoh SY, Liew WC, Shen X, Hon LS, Drake JP, Boitano M, Hall R, Chin CS, Lachumanan R, Korlach J, Trifonov V, Kabilov M, Tupikin A, Green D, Moxon S, Garvin T, Sedlazeck FJ, Vurture GW, Gopalapillai G, Kumar Katneni V, Noble TH, Scaria V, Sivasubbu S, Jerry DR, O'Brien SJ, Schatz MC, Dalmay T, Turner SW, Lok S, Christoffels A, Orbán L.</p> <p>PLoS Genetics. 2016 Apr 15;12(4):e1005954. doi: 10.1371/journal.pgen.1005954.</p>	15 April
10.	<p>A new tool for prioritization of sequence variants from whole exome sequencing data.</p> <p>Glanzmann B, Herbst H, Kinnear CJ, Möller M, Gamielidien J, Bardien S.</p> <p>Source Code Biol Med. 2016 Jul 1;11:10. doi: 10.1186/s13029-016-0056-8.</p>	01 July
11.	<p>Identification of phenotype-relevant differentially expressed genes in breast cancer demonstrates enhanced quantile discretization protocol's utility in multi-platform microarray data integration.</p> <p>Mapiye DS, Christoffels AG, Gamielidien J.</p> <p>J Bioinform Computational Biology. 2016 Jul 13:1650022.</p>	14 July
12.	<p>Genome-Wide DNA Methylation in Mixed Ancestry Individuals with Diabetes and Prediabetes from South Africa.</p> <p>Matsha TE, Pheiffer C, Humphries SE, Gamielidien J, Erasmus RT, Kengne AP.</p> <p>International Journal Endocrinology. 2016;2016:3172093. doi: 10.1155/2016/3172093.</p>	31 July
13.	<p>Mycobacterium tuberculosis strains exhibit differential and strain-specific molecular signatures in pulmonary epithelial cells.</p> <p>Mvubu NE, Pillay B, Gamielidien J, Bishai W, Pillay M.</p> <p>Dev Comp Immunol. 2016 Dec;65:321-9. doi: 10.1016/j.dci.2016.07.022.</p>	03 August

#	Publication (SANBI contributors)	Published Date
14.	<p>Computational characterization of Iron metabolism in the Tsetse disease vector, <i>Glossina morsitans</i>: IRE stem-loops.</p> <p><u>Dashti ZJ</u>, Gamielidien J, Christoffels A. BMC Genomics. 2016 Aug 8;17:561. doi: 10.1186/s12864-016-2932-7</p>	08 August
15.	<p>Structure of an N276-Dependent HIV-1 Neutralizing Antibody Targeting a Rare V5 Glycan Hole Adjacent to the CD4 Binding Site.</p> <p>Wibmer CK, Gorman J, Anthony CS, Mkhize NN, Druz A, York T, Schmidt SD, <u>Labuschagne P</u>, Louder MK, Bailer RT, Abdool Karim SS, Mascola JR, Williamson C, Moore PL, Kwong PD, Morris L. Journal of Virology 2016 vol: 90 (22) pp: 10220-10235 doi:10.1128/JVI.01357-16.</p>	31 August
16.	<p>Divergent evolutionary and epidemiological dynamics of cassava mosaic geminiviruses in Madagascar.</p> <p>De Bruyn A, Harimalala M, Zinga I, <u>Mabvakure BM</u>, Hoareau M, Ravigné V, Walters M, Reynaud B, Varsani A, <u>Harkins GW</u>, Martin DP, Lett JM, Lefeuvre P. BMC Evol Biol. 2016 Sep 6;16:182. doi: 10.1186/s12862-016-0749-2.</p>	06 September
17.	<p>Ongoing geographical spread of Tomato yellow leaf curl virus.</p> <p><u>Mabvakure B</u>, Martin DP, Kraberger S, Cloete L, van Brunschot S, Geering AD, Thomas JE, Bananej K, Lett JM, Lefeuvre P, Varsani A, <u>Harkins GW</u>. Virology. 2016 Nov;498:257-64. doi: 10.1016/j.virol.2016.08.033.</p>	15 September
18.	<p>SNiPhunter. A SNP based search engine.</p> <p><u>Werner Veldsman</u> and Alan Christoffels. Data, 1, 17 doi:10.3390/data1030017</p>	29 September
19.	<p>Identification and characterization of microRNAs expressed in the African malaria vector <i>Anopheles funestus</i> life stages using high throughput sequencing.</p> <p>Mushal Allam, Belinda L. Spillings, Hiba Abdalla, <u>Darlington Mapiye</u>, Lizette L. Koekemoer and <u>Alan Christoffels</u> Malaria Journal 2016 15:542 DOI: 10.1186/s12936-016-1591-0</p>	8 November
20.	<p>Chinks in the armor of the HIV-1 Envelope glycan shield: Implications for immune escape from anti-glycan broadly neutralizing antibodies.</p> <p>Thandeka Moyo, <u>Roux-Cil Ferreira</u>, Reyaz Davids, Zarinah Sondag, Penny L. Moore, <u>Simon A. Travers</u>, Natasha T. Wood, Jeffrey R. Dorfman Virology Journal doi: 10.1016/j.virol.2016.10.026.</p>	13 November

Published Conference Proceeding

Supporting the development of biobanks in low and medium income countries.

Tomas Klingstrom, Maimuna Mendy, Dominique Meunier, Anouk Berger, Jane Reichel, [Alan Christoffels](#), [Hocine Bendou](#), Carmen Swanepoel, Lemoene Smit, Campbell McKellar-Basset, Erik Bongcam-rudloff, Jonas Soderberg, Roxana Merino-Martinez, Suyesh Amatyia, Absolomon Kihara, Steve Kemp, Robert Reihns and Heimo Muller (2016).

IST-Africa 2016 Conference Proceedings.

Paul Cunningham and Miriam Cunningham (Eds).

ISBN: 978-1-905824-55-7.

Computational Tool Developed

Peter van Heusden, Thoba Lose, Alan Christoffels

COMBAT-TB: A tuberculosis analysis work bench

<https://zenodo.org/record/51638#.Vz3A17p97CJ>

Patent Registered

Simon Travers, Imogen Wright

A worldwide patent was submitted for our novel sequence alignment tool, the second patent to arise from the exatype project.

Invited Talks

Presenter	Month	Title	Details
Simon Travers	June	Highly sensitive, easy-to-use analysis tools for NGS-based HIV drug resistance genotyping	Stanford University Medical School, California, USA
Alan Christoffels	August	COMBAT-TB: Computational bacterial analytical toolkit for Tuberculosis research.	Johns Hopkins University, Maryland, USA
Simon Travers	October	Tools and pipelines for the analysis of NGS data in HIV vaccine research.	HVTN/CHIL NGS Sequencing Workshop, Cape Town

Scientific/Research Meeting

Attendee	Month	Meeting Name	Purpose of meeting
Peter van Heusden	September	Bioinformatics Service Platform (BSP) stakeholders meeting	The current state of, as well as the vision for, the BSP was discussed with the goal of determining how the BSP can best serve the needs of end users. This included the services and infrastructure available, collaboration and funding models, and training activities.

Conference Participation

Presenter	Month	Conference Details	Title	Type
Warren Jacobus	March	Second Tunisia/South Africa Workshop on Advanced Models in QoS and Security Engineering of Hybrid Cloud Networking: Applications to Public Safety, UWC Computer Science Department, Cape Town	Lightweight Cloud Computing for Big Data Processing.	Talk
Junaid Gamielidien	April	13th International Congress of Human Genetics in Kyoto, Japan	The BioOntological Relationship Graph (BORG) Database - a Novel Concept for Prioritizing Candidates from High Throughput Genomics Studies.	Talk
Mahjoubeh Jalali	April	13th International Congress of Human Genetics in Kyoto, Japan	Trio-based exome sequencing case study in identifying the underlying genetic factors associated with progressive flail arm syndrome.	Poster
Roux-Cil Ferreira	April	23rd International HIV Dynamics and Evolution meeting, Woods Hole, Massachusetts, USA	Molecular dynamics simulation of glycosylated HIV-1 envelope trimers with different neutralisation profiles.	Poster
Phillip Labuschange	April	23rd International HIV Dynamics and Evolution meeting, Woods Hole, Massachusetts, USA	Motifbinner: Sensitively processing next-generation sequence data produced using the Primer ID approach.	Poster
Simon Travers	April	23rd International HIV Dynamics and Evolution meeting, Woods Hole, Massachusetts, USA	Evaluating the suitability of various next-generation sequence platforms for the establishment of high-throughput, highly sensitive HIV drug resistance genotyping in resource-limited settings with a high burden of HIV.	Talk
Imogen Wright	April	23rd International HIV Dynamics and Evolution meeting, Woods Hole, Massachusetts, USA	Realistic NGS drug resistance simulation with application to the evaluation of mutation-calling software.	Talk
van Heusden P, Vij S, Orban L and Christoffels A.	April	ISCB Next Generation Sequencing Conference (NGS) 2016, Barcelona, Spain	An extensible genome annotation workbench based on the Galaxy Platform.	Poster
Lose T, van Heusden P and Christoffels A.	June	Galaxy Community Conference, Indiana University, USA	A Galaxy interactive environment for exploring the Neo4j graph database.	Poster

Presenter	Month	Conference Details	Title	Type
Simon A. Travers	June	2016 BIO International Convention, San Francisco, USA	Highly sensitive, easy-to-use analysis tools for DNA-based diagnostics.	Poster
Natasha T. Wood.	June	Virus Genomics and Evolution Conference, Cambridge, UK	Development and validation of next-generation sequencing analysis software for HIV drug resistance genotyping.	Poster
Natasha T. Wood.	July	ISCB/ISMB 2016, Florida, USA	Molecular dynamics simulations of glycosylated HIV-1 GP120 trimers in the context of viral coreceptor tropism.	Poster
Timothy Reid, Carmen Swanepoel, Catherine Roussouw, Lemoene Smit, Alan Christoffels, Akin Abayomi	August	Stellenbosch University Medicine and Health Sciences 60th Annual Academic Day, Tygerberg Campus, Parow, Cape Town	The learning curve for laboratory information management system (LIMS) implementation: A case study at the NHLS-Stellenbosch Biobank (NSB).	Poster
Natasha T. Wood.	August	Warren Workshop VI 2016, Sapporo, Hokkaido, Japan	Molecular dynamics simulations of two glycosylated HIV-1 gp160 trimers with different neutralisation profiles.	Poster
Werner Veldsman, Alan Christoffels	September	15th International Conference on Bioinformatics (InCOB) 2016, Singapore	SNP based literature and data retrieval.	Poster
Imogen Wright	October	PyCon ZA 2016, The Riverclub, Cape Town	Simulating drug resistance in HIV DNA using Python.	Talk
Simon A. Travers	October	Gates Grand Challenges Annual Meeting, London, UK	Software solutions for routine multi-disease drug resistance detection & surveillance.	Poster
Imogen Wright	October	TedXCapeTown Women 2016	Can we slow the spread of drug resistance, before we run out of effective drugs to treat common infections?	Talk
Samuel Egieyeh	October	All Africa Congress on Pharmacology and Pharmacy, Misty Hills Hotel and Conference Centre in Muldersdrift, Gauteng, SA	Novel antimalarial from nature: evolutionary generated virtual compound libraries from natural products with antiplasmodial activities.	Talk
Phillip Labuschagne	October	Cape Town HVTN Next Generation Sequencing Workshop, Cape Town	Processing Primer ID Data.	Talk
Simon Travers	November	Bioinfect Conference 2016, Alderley Park Conference Centre, Cheshire, UK	Panel member for session entitled: How diagnostic tests are critical to the global AMR challenge.	Panel Member

Thesis Examination at Other Institutions

PI Name	Institution	Degree
Alan Christoffels	Stellenbosch University	1 PhD 1 MSc

External Moderation

PI Name	Institution	Course
Junaïd Gamielidien	University of KwaZulu-Natal	Honours Degree Module: Methods in Systems Biology
Simon Travers	University of Cape Town	Honours degree in Bioinformatics
	University of Pretoria	Honours degree in Bioinformatics

Journal Editing and Reviews

PI Name	Journal
Alan Christoffels	Editorial Board Member: Data Journal Reviewer: BMC Evolutionary Biology, BMC Genomics, BMC Bioinformatics, Bioinformatics, Journal of Parallel and Distributed Computing, Trends in Genetics, PLoS ONE, Genome Research
Junaïd Gamielidien	Reviewer: Nature Scientific Reports, PLoS ONE, Molecular Genetics and Genomic Medicine, Clinical Genetics, Computer Methods and Programs in Biomedicine, BMC Bioinformatics
Gordon Harkins	Reviewer: Diversity and Distributions

Expert Panel or Committee Memberships

PI Name	Membership
Alan Christoffels	Co-Chair of ISMB: Special Interest Group (SIGs): 24th Annual International Conference on Intelligent Systems for Molecular Biology, July 7-12, Orlando, USA Conference organising committee member: International Conference on Bioinformatics InCoB Official annual conference of the Asia-Pacific Bioinformatics network (APBioNet) Singapore, September
Simon Travers	Organising committee, abstract reviewer and session chair: 23rd International HIV Dynamics and Evolution, Massachusetts, USA, April

Research Projects

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

PI Name	Theme	Projects
Alan Christoffels	Host-pathogen interactions	<ul style="list-style-type: none"> • Virulence mutations: In collaboration with the Tygerberg MRC Unit, we are developing methods to analyse high throughput sequencing data for mycobacterial genomes. • Identification of novel drug targets in pathways known to contain TB drug resistant genes. • Predicting the interaction networks between human and mycobacteria. • In collaboration with the National Institute for Communicable Diseases (NICD), we are investigating miRNA targets in <i>Anopheles funestus</i> to understand regulation of mosquito development. • Characterising olfactory genes in tsetse.
	Biobanking	<ul style="list-style-type: none"> • Our efforts to develop biobank standards continue through the NIH funded H3Africa project. New funding was obtained in partnership with European Union researchers. • Identifying biomarkers in HIV lymphomas.
Junaid Gamielidien	Cloud computing	<ul style="list-style-type: none"> • Development of cloud based whole exome, genome and RNAseq pipelines that incorporate international best practices and in-house SOPs to simultaneously reduce cost of hardware ownership while enabling fast and parallel processing of any number of samples.
	Tools and SOPs for variant prioritization	<ul style="list-style-type: none"> • Continual development and improvement of rules and associated pipelines for identifying likely function impacting and phenotype modifying genetic variants from whole exome and genome sequencing experiments.
	Cognitive computing	<ul style="list-style-type: none"> • Development of a large NoSQL biomedical knowledgebase and associated search strategies for rapidly prioritising candidate disease-causing/modifying genetic variants or differentially expressed genes through simulating the cognitive processes applied by clinical geneticists and biomedical researchers.
	Applied whole exome and genome sequencing	<ul style="list-style-type: none"> • Collaborative projects on: distal and limb-girdle muscular dystrophy, ophthalmic complications of myasthenia gravis, motor neuron disease (ALS), Alzheimer's disease, sports-related tendinopathy and monogenic diabetes.
Uljana Hesse	Rooibos Transcriptome	<ul style="list-style-type: none"> • Sequencing the transcriptomes of rooibos (<i>Aspalathus linearis</i>) and eukaryotic micro-symbionts.

PI Name	Theme	Projects
Gordon Harkins	Molecular evolution of viruses	<ul style="list-style-type: none"> • In collaboration with researchers from Belgium and South Africa we have received funding from the National Research Foundation (NRF) to investigate using a combination of laboratory experimentation and state-of-the-art computational analyses, how the traits of virulence and transmission rates are evolving in viral pathogens that cause important crop diseases (Maize Streak Disease, Cassava Mosaic Disease and Tomato Yellow Leaf Curl Disease). • In collaboration with researchers from CIRAD France, the State University of Arizona and the University of Cape Town, we have adopted a “geo-metagenomics” approach to investigate whether ecological disturbances such as intensive agriculture, cause measurable changes in the spatial and temporal diversity, demographics and evolutionary dynamics of viral communities inhabiting natural ecosystems. • We are investigating the degree to which RNA viral evolution is constrained by secondary structure within the genomes of positive sense single-stranded RNA viruses in the order Picornavirales: the most common cause of infections in humans in developed countries. • In collaboration with researchers from the University of KwaZulu Natal and the University of Cape Town we have computationally reconstructed, the early spatio-temporal history of the HIV-1M epidemic to better understand the evolutionary dynamics of this group of viruses before the onset of the global pandemic. • In collaboration with researchers from France and South Africa we are studying endogenous viral elements (EVE's) and their role in virus evolution and their functions in plants.
Simon Travers	HIV Dynamics	<ul style="list-style-type: none"> • HIV drug resistance testing: We have established a spin-off company (Hyrax Biosciences) that supports the establishment of routine HIV drug resistance testing in South Africa and other resource limited countries with a high burden of HIV using the exatype computational pipeline developed at SANBI. • Research investigating the role of N-linked glycosylation in the escape of HIV from, and susceptibility to, neutralizing antibodies. • Investigating the role of N-linked glycosylation on protein structure and function in relation to host cell recognition and entry. • A number of studies as part of the HIV vaccine trials network (HVTN) analysing sequence data generated in multiple projects including sieve analysis for vaccine trials, ontogeny of B cell responses, transcription profiles. • A number of studies that are exploring the use of next-generation sequencing for HIV drug resistance genotyping through the optimisation of sample preparation and data analysis processes.
Peter van Heusden	Cloud Computing	<ul style="list-style-type: none"> • We have initiated the testing of various private cloud computing solutions for biomedical research applications and to enhance our training laboratory.

Research Laboratories

Research Laboratory of Alan Christoffels



Prof Alan Christoffels.

Highlights of the lab:

Fulbright Visiting Scholar Program

I was awarded a Fulbright fellowship that funded my sabbatical in 2016 to spend 6 months at the Broad Institute. I was hosted by a bacterial genomics team lead by Dr Bruce Birren. During this time I analyzed the role of small RNAs to drug resistance in *M.tuberculosis*.

Publications

A total of 11 papers were published in 2016 and another 4 papers submitted for peer-review.

Graduations

Congratulations to the two PhD graduates (Rosaline Macharia and Samuel Egieyeh), MSc graduate (Werner Veldsman) and Honours

students (Warren Jacobus and Lynley Abdoll) in my lab.

Funding

The research in my group is supported by funding from the South African National Research Foundation, SA Medical Research Council and the European Union Horizon 2020.

Research Projects:

As a DST/NRF Research Chair in Bioinformatics and Public Health Genomics, my group has been developing methods to manage, and analyze large genomic datasets. Currently, our efforts are targeted towards infectious diseases where we are developing computational tools and applying them to genomic datasets for tuberculosis, malaria and sleeping sickness. Our research findings are only meaningful if translated into a clinical intervention strategy.

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**) initiative to deploy analytic tools across the African continent.

Search for new anti-malarials

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

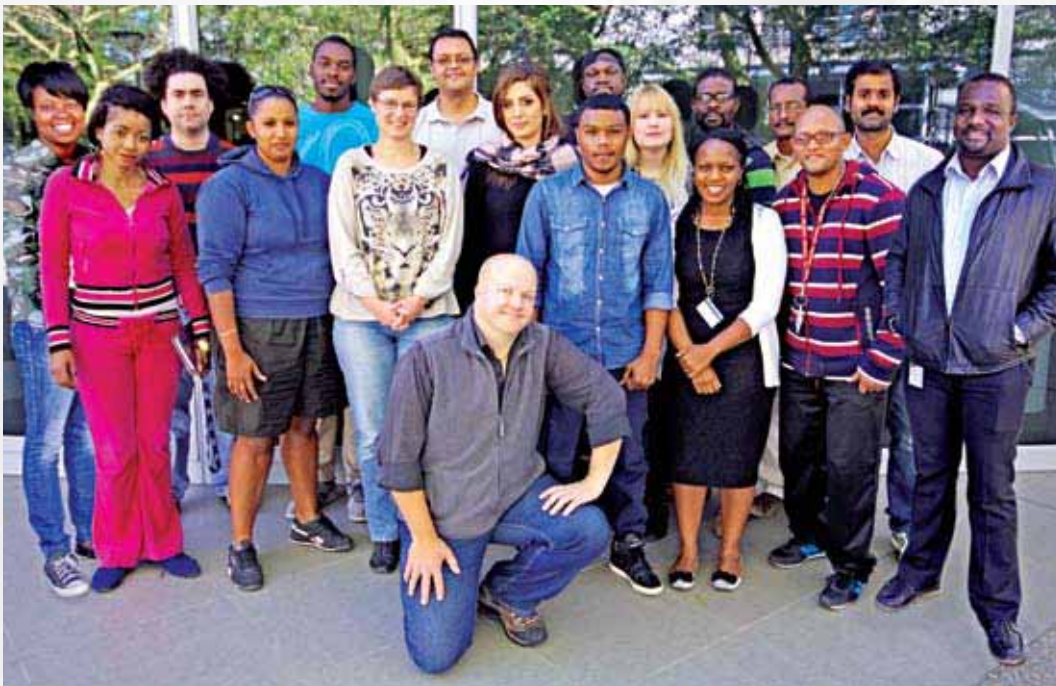
Infrastructure for biobanking

The collection and storage of biological material and its associated data is essential

to any genetics project with a long term vision. To this end we have contributed to the the development of a computational environment to manage and analyze genomic data that is linked to a biobank. Specifically, we have developed an open-source laboratory information management system – **Baobab LIMS**.

Genome Annotation

We continue to extend our genome annotation footprint through our participation in international projects to define the gene repertoire in newly sequenced genomes. More recently, we have participated in an international consortium to sequence and annotate the Asian Seabass genome.



Members of the Christoffels lab.

Research Collaborations:

1. Prediction of human-tuberculosis interaction networks

Collaborating Parties:

Prof Peter Witbooi - Mathematics Department, UWC

Prof Eileen Hoal-van Helden - Medical Biochemistry, Stellenbosch University

Nature and Purpose:

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

Output in the last 12 months:

A PhD student submitted their thesis.

2. Develop a biobank informatics management system

Collaborating Parties:

B3Africa consortium funded through Horizon2020 including:

Prof Akin Abayomi - Hematology Department, Tygerberg Hospital

Prof Erik Bongcam-Rudloff - Swedish Agriscience, Uppsala

Dr Heimo Muller - Graz University, Austria

Nature and Purpose:

Development of a laboratory management system that will integrate with other biobanking tools through an open-source platform called **Biobank in a Box (BiBox)**.

Output in the last 12 months:

We secured EU funding to support the software development.

3. Chemosensory genes in Tsetse

Collaborating Parties:

Dr Dan Masiga - ICIPE, Kenya

Nature and Purpose:

To annotate chemosensory genes in 5 tsetse species.

Output in the last 12 months:

Published two papers.

Two PhD graduations.

4. Identification of miRNAs in *An.funestus*

Collaborating Parties:

Prof Lizette Koekemoer - NICD, Pretoria

Nature and Purpose:

Mosquitoes from the Anopheline species are the principle vectors of malaria in Africa. Limited data exist for miRNA functions in Anopheline mosquitoes and few reports suggest miRNA control of plasmodium survival rate. We identified miRNAs that function in the survival and development of mosquitoes using NGS data and predicted miRNA targets (www.insectar.sanbi.ac.za).

Output in the last 12 months:

One publication.

5. Identification of novel drug targets for drug resistant tuberculosis

Collaborating Parties:

Prof Samantha Sampson - University of Stellenbosch

Prof Rob Warren - University of Stellenbosch

Dr Cedric Wereley - University of Stellenbosch

Nature and Purpose:

Current TB drugs are more than 30 years old and have unacceptable efficacy and safety profiles, emphasising the need for new drugs.

We mapped drug resistance genes, derived from comparative genome analysis of three *M. tuberculosis* strains (susceptible, multidrug, and extensively drug resistant), and from published literature to metabolic pathways and identified nine potential drug target candidate genes. These genes were ranked for further computational analyses in the quest for identifying inhibitors that could be lead compounds.

Output in the last 12 months:

Identified 2 compounds through in-silico screens that bind to drug targets. These compounds are undergoing in-vitro testing in the laboratory.

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

Collaborating Parties:

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

Nature and Purpose:

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

Output in the last 12 months:

Two papers published and one PhD graduated.

Research Laboratory of Junaid Gamiieldien



Prof Junaid Gamiieldien.

Highlights of the lab:

- 10 publications in international peer reviewed journals.
- Acceptance of our guide to prioritizing variants from whole exome sequencing studies by BioTechniques for publication in January 2017.
- Successful application of our semantic technology to identify a strong candidate for unexplained second generation symptoms in Miyoshi myopathy (manuscript in review).
- Oral presentation at the 13th International Congress of Human Genetics in Kyoto, Japan.
- Successful testing and research application of our cloud based NGS analytical services platform.
- Following the visit by the Russian Ministry of Science and Education in May, Junaid was approached for a BRICS Multilateral Joint Science and Technology Research Collaboration grant application with Russian and Chinese counterparts which is currently in review.

Research projects

Cognitive computing for disease genomics

Due to the large number of candidates that still remain after standard processing and filtering of Whole Exome Sequencing (WES) and other genomics data, cognitive computing solutions such as IBM Watson, which are patterned after human thought and harness vast amounts of extant knowledge, have emerged as important innovations to disambiguate results and generate hypotheses and to support precision medicine. Our BioOntological Relationship Graph (BORG) Database similarly assimilates and integrates multiple sources of genomic and biomedical knowledge and metadata and is able to learn rules about diseases and the phenotypes and gene functions associated with disease, which it utilizes to automatically prioritize likely candidates from genomics studies. In essence, the system uses search paradigms that simulate the cognitive processes applied by clinical geneticists and biomedical researchers. In practice, it performs a directed walk on a stored biomedical knowledge graph, which enables the identification of non-obvious yet biologically plausible links that may explain a gene's potential role in the clinical phenotype based on relevant prior annotations, or transitively via the functions of proteins that the gene's expressed product physically interacts with. This has proven especially useful in our disease variant and transcriptomics collaborative studies, where biologically and biomedically plausible links that would otherwise have been missed have been identified.

Cloud-based analytical pipelines for NGS

The cost of ownership of specialised hardware required for very large NGS projects is very high, and scaling up infrastructure is not sustainable in the current era of rapidly decreasing data cost. The primary aim of the project is to implement an effectively

infinitely scalable cloud based exome, genome and RNAseq data processing platform, which can be used to deliver cost-effective services locally and internationally, while reducing the dreaded data processing bottleneck from several months to weeks or even days.

Development of SOPs for variant prioritization

For the clinical research applications of NGS, which is the major applied research focus of our lab, it is crucial that the variant discovery process is carefully developed and rigorously optimized. The vast amount of data it produces simultaneously introduces new challenges and opportunities. While workflows and associated software to process the raw data and to produce high-confidence variant calls have significantly improved, filtering the tens of thousands of candidates to produce a subset relevant to the study at hand is still complex. We are thus continually developing rules and SOPs for identifying functional variants from whole genome and exome variant calls and prioritizing those potentially associated with a phenotype of interest. Used in conjunction with our cognitive analytics approach, we have been able to identify strong candidate causal variants in several in-house and collaborative disease studies that would otherwise have been overlooked or discarded due to the unobvious links to the phenotype.

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

Among the monogenic forms of diabetes mellitus (DM), Mature Onset Diabetes of the Young (MODY), which has an age of onset of less than 25 years, has been the most intensively investigated in recent years. MODY is a group of clinically heterogeneous, often non-insulin-dependent forms of DM that are defined at the molecular genetic level by deleterious mutations in different genes. As it is estimated that up to 80% of MODY cases go undiagnosed and therefore sub-optimally treated, we aim to develop a monogenic diagnostic instrument based on WES combined with our bespoke causative variant prioritization system as part of our MRC Strategic Health Innovation Partnerships grant. Exome sequencing data for suspected MODY cases will be delivered in February 2017.

Clinical exome sequencing projects

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

- Muscular dystrophy
- Tendinopathy
- Early onset motor neuron disease



Junaid Gamieldeen (centre) and Mahjoubeh Jalali (right) at the Kyoto, Japan Conference.

Research Collaborations

1. Exome sequencing of sporadic motor neuron disease: flail arm ALS

Collaborating Parties:

Dr Mahjoubeh Jalali – SANBI, UWC
Prof Jeanine Heckmann – University of Cape Town

Nature and Purpose:

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

Output in the last 12 months:

Participants have been exome sequenced and using our semantic model of motor neuron disease, strong candidate causative variants have been identified and Sanger confirmed in an affected-child + unaffected parents trio.

Future Direction:

Manuscript in preparation.

2. Exome sequencing of myasthenia gravis-related ophthalmoparesis

Collaborating Parties:

Dr Mahjoubeh Jalali - SANBI, UWC
Prof Jeanine Heckmann (PI), Dr Melissa Nel (PhD candidate) - University of Cape Town

Nature and Purpose:

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

Output in the last 12 months:

Manuscript under review.

Future Direction:

None yet planned.

3. Exome sequencing of atypical diabetes

Collaborating Parties:

Dr Mahjoubeh Jalali - SANBI, UWC
Prof Alison September (co-PI) - University of Cape Town
Prof Naomi (Dinky) Levitt - University of Cape Town

Nature and Purpose:

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

Output in the last 12 months:

Cases of possible MODY have been exome sequenced.

Future Direction:

Variant calling and candidate causative mutation prioritization using cognitive database.

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

Collaborating Parties:

Dr Colleen Saunders - SANBI, UWC
Prof Alison September and Malcolm Collins - University of Cape Town

Nature and Purpose:

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

Output in the last 12 months:

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

Future Direction:

Several manuscripts in preparation.
Genotyping additional candidates.

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

Collaborating Parties:

Dr Kareemah Gamielidien (PI) - Cape

Peninsula University of Technology

Dr Wayne Kleintjies - Head of the Tygerberg Hospital Burns Unit

Nature and Purpose:

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

Output in the last 12 months:

Sample collection has been initiated.

Future Direction:

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

6. Understanding the SHARED ROOTS of Neuropsychiatric Disorders and Modifiable Risk Factors for Cardiovascular Disease

Collaborating Parties:

A South African Medical Research Council Flagship project - multiple collaborating institutions and departments, with Prof Soraya Seedat of Stellenbosch University as PI. Our group is the bioinformatics lead on the project.

Nature and Purpose:

The overarching aim of SHARED ROOTS is to use a 'whole systems' biology approach to identify biological, environmental and behavioural signatures of disease comorbidity in neuropsychiatric disease in order to direct future prevention and treatment. It will combine genomic, transcriptomic, epigenetic, and complementary phenotypic and multimodal neuroimaging data, to disentangle mechanistic pathways that lead to the development of comorbidity of these disorders.

Output in the last 12 months:

mRNA isolated from blood has been profiled using RNAseq and candidate markers prioritized using a BORG cognitive model of PTSD.

Future Direction:

A larger cohort of PTSD and control patients will be studied.

Research Laboratory of Gordon Harkins



Dr Gordon Harkins

My research primarily focuses on the evolution and molecular epidemiology of single-stranded DNA and RNA viral pathogens of animals and plants. I am a member of a highly productive, plant-virus epidemiology network that seeks to determine the evolutionary underpinnings of the emergence and spread of the numerous viral diseases that seriously threaten the health and food security of Africa and the rest of the developing world. The rapid rate of evolutionary change in viruses means that the epidemiological and ecological processes that shape their genetic diversity act on approximately the same timescale as mutations that become fixed within viral populations. Consequently, the patterns of genetic variation present in viral genomes can be used to infer the processes underlying the dynamics of viral evolution, providing a unique molecular perspective on their ancestry and mechanisms of change.

A summary of some of the research projects that members of my research group have been involved in 2016 is presented below.

Plant-infecting viral pathogens

1. *Virulence Evolution*

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

2. *Viral Emergence*

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

complement the data on contemporary viral diversity derived from field studies, we are sequencing “ancient” viral genomes present in fynbos plant specimens stored within French and South African herbaria to more accurately estimate when and where, key ancestral plant viruses existed. By producing infectious clones of both these ancient genomes and re-synthesized computationally inferred ancestral genomes, it will be possible to pinpoint when and where over the past few centuries, major changes in virus virulence have occurred. It is hoped that this project will complement the results obtained from our previous work on virulence evolution in Maize streak virus and increase our understanding of the evolutionary and epidemiological processes by which members of this important group of crop pathogens (i) emerged during the colonization era from their indigenous African hosts, (ii) began causing diseases in crop species introduced by Europeans and (iii) disseminated across the continent from sites of initial emergence to become what are today among the most important biotic threats to African food security.

3. Viral dynamics

The ongoing global spread of pathogenic crop-infecting (tomatoes, maize, and cassava) viruses within the family *Geminiviridae*, represent a serious looming threat to food production in many regions of the world. We applied Bayesian phylogeographic inference and recombination analyses to all available viral sequences and reconstructed a plausible history of ongoing diversification and movements throughout the world and identified a range of predictor variables that influence the source-sink dynamics of these pathogens using Generalized Linear Models (GLM).

Human–infecting viral pathogens

RNA viruses

We are investigating the degree to which RNA viral evolution is constrained by secondary

structure within the genomes of positive sense single-stranded RNA viruses in the order Picornvirales (one of the most genetically diverse of the positive-sense single-stranded RNA viral orders and the most common cause of infections in humans in developed countries. Among the most notable serious health threats include clinical syndromes such as HIV/AIDS, aseptic meningitis, paralytic poliomyelitis, SARS and hepatitis.

Human immunodeficiency virus (HIV)

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

Highlights of the lab:

The following **three journal articles were published** in 2016 with an average impact factor of 3.3:

The Divergent Evolutionary and Epidemiological Dynamics of Cassava Mosaic Geminiviruses in Madagascar.

BMC Evolutionary Biology: 2016 16:182.
DOI: 10.1186/s12862-016-0749-2.

Alexandre De Bruyn, Mireille Harimalala, Innocent Zinga, Batsirai M. Mabvakure, Murielle Hoareau, Virginie Ravigné, Matthew Walters, Bernard Reynaud, Arvind Varsani, Gordon W. Harkins, Darren P. Martin, Jean-Michel Lett and Pierre Lefevvre.

Molecular characterization and prevalence of two capulaviruses: Alfalfa leaf curl virus from France and Euphorbia caput-medusae latent virus from South Africa.

Virology 493 (2016)142–153.
<http://dx.doi.org/10.1016/j.virol.2016.03.016>.

Pauline Bernardo, Brejnev Muhire, Sarah François, Maëlle Deshoux, Penelope Hartnady, Kata Farkas, Simona Kraberger, Denis Filloux, Emmanuel Fernandez, Serge Galzi, Romain Ferdinand, Martine Granier, Armelle Marais, Pablo Monge Blasco, Thierry Candresse, Fernando Escriu, Arvind Varsani, Gordon W. Harkins, Darren P. Martin, Philippe Roumagnac.

The Ongoing Geographic Spread of Tomato yellow leaf curl virus.

Virology, 498, 257-264, 2016.
<http://dx.doi.org/10.1016/j.virol.2016.08.033>

Batsirai Mabvakure, Darren P. Martin, Simona Kraberger, Leendert Cloete, Sharonvan Brunschot, Andrew D.W. Geering, John E. Thomas, Kaveh Bananej, Jean-Michel Lett, Pierre Lefeuvre, Arvind Varsani, Gordon W.Harkins

One MSc (cum laude) student graduated in 2016. The MSc student Eugene Madzokere **obtained his first coauthored publication** entitled “The role of Kenya in the trans-African spread of Maize streak virus strain A” in 2017 and will submit a second as a primary author for review in 2017.

A second SANBI MSc graduate Batsirai Mabvakure obtained his **third publication**

entitled “The Ongoing Geographic Spread of Tomato yellow leaf curl virus” as a primary author in 2016.

In **recognition of their productivity** in terms of scientific publications (34 in five years), Gordon Harkins along with his collaborators from CIRAD, Montpellier (Dr. Philippe Roumagnac) the University of Cape Town (Associate Prof Darren Martin) and Stanford University (Dr. Brejnev Muhire) were invited to present their research findings at SANBI on the evolution and molecular epidemiology of important crop infecting viral pathogens to the French ambassador to South Africa, Mrs. Elisabeth Barbier.

Philippe Roumagnac spent two years in South Africa on a Marie Curie fellowship to conduct research on the ecological and molecular mechanisms of viral emergence. The purpose of the ambassador’s visit was to showcase the research that has emanated from this **highly productive international collaboration** and to determine how to ensure that future bioinformatics research initiatives between the respective French and South African scientific research institutes may be sustained after Philippe Roumagnac returned to France at the end of 2016. To this end, a further meeting was held in April 2016 in Cape Town with Daniel Barthélémy (Director of Cirad System Biology Department) and **funding was subsequently secured** to collect and sequence “ancient” viral genomes present in fynbos plant specimens stored within French and South African herbaria to more accurately estimate when and where, key ancestral plant viruses existed.

Research Collaborations:

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

Collaborating Parties:

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

Nature and purpose:

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

Output in the last 12 months:

None

Future Direction:

A manuscript reporting the findings of this study will be submitted for publication in early 2017.

2. Virulence Evolution

Collaborating Parties:

Philippe Roumagnac - CIRAD, Montpellier, France.

Darren P. Martin, Adérito Luis Monjane, Dionne Natalie Shepherd - UCT, South Africa.

Philippe. Lemey and Simon Dellicour - Department of Microbiology and Immunology, Katholieke Universiteit, Leuven, Belgium.

Pierre Lefeuvre and Jean-Michel Lett - CIRAD, Reunion Island.

Arvind Varsani - University of Canterbury, Christchurch, New Zealand.

Nature and purpose:

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

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

Output in the last 12 months:

None

Future Direction:

Several manuscripts reporting the findings of this study will be submitted for publication in 2017.

3. Molecular Mechanisms of Viral Emergence

Collaborating Parties:

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

Brejnev Muhire, Darren P. Martin - Institute of Infectious Disease and Molecular Medicine, University of Cape Town.

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

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

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

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

Nature and purpose:

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

Output in the last 12 months:

This work has resulted in one published paper and a second manuscript that was submitted for publication to the ISME journal *Nature* in 2016.

Future Direction:

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

4. Viral dynamics

Collaborating Parties:

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

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

Darren P. Martin - University of Cape Town.

Virginie Ravigné - CIRAD, Montpellier, France

Simona Kraberger - Colorado State University, Fort Collins, USA

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

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

Nature and purpose:

To apply recombination analyses and Bayesian phylogeographic inference methods to all available TYLCV and CMV viral sequences and reconstruct a plausible history of ongoing diversification and movements throughout the world.

Output in the last 12 months:

This work has resulted in two published papers in 2016 and a third that was submitted in 2016 and accepted for publication in 2017.

Future Direction:

To collect and sequence “ancient” viral genomes present in fynbos plant specimens stored within French and South African herbaria to more accurately estimate when and where, key ancestral plant viruses existed.

5. Endogenous Viral Elements (EVE's)

Collaborating Parties:

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

Darren Martin - University of Cape Town.

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

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

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

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

Andrew Geering - University of Queensland, Australia.

Nature and purpose:

To study endogenous viral elements (EVE's) and their role in virus evolution and their functions in plants to address fundamental biological issues that are critical to the understanding of plant virus evolution and the contribution of plant viruses to both the host genome evolution and biology. This

involves approaches in virology, bioinformatics, genomics, proteomics, evolution and synthetic biology, and aims at disentangling key mechanisms of adaptive plasticity of plant and viral genomes such as gene domestication, genetic and epigenetic regulation of plant gene expression and deciphering mechanisms that are central to molecular plant-virus interactions such as recombination, transcriptional or post transcriptional gene silencing.

Output in the last 12 months:

None

Future Direction:

A funding proposal to support this research has been re-submitted to the French National Research Foundation in November 2016.

Research Laboratory of Uljana Hesse



Members of the lab: Uljana Hesse, Wesley William, Yamkela Mgwatyu and Emily Stander.

My primary research interests focus on the genomics of South African medicinal plants and their symbiotic microorganisms. South Africa is home to ~3000 plant species with health promoting properties that are actively used across the country, with nearly 70% of the population relying on herbal remedies for medication. Yet, plant production and breeding of medicinal plants in South Africa are virtually missing. Understanding the genetic background of medicinal compound biosynthesis and plant adaptation mechanisms to stress would not only improve agricultural aspects of plant production, but would also allow to develop bio-engineering of selected compounds through in-vitro expression.

Highlights of the lab:

Funding had been secured in 2015 through successful NRF project applications for rooibos transcriptome analyses (Research and Technology Fund; first successful application at UWC) and rooibos genome sequencing (Competitive Funding for Unrated Researchers).

In 2016, three students have joined my research team:

- Yamkela Mgwatyu, who was conducting a DST-NRF internship in 2016 and is continuing with her MSc in 2017;
- Emily Stander, who has started her PhD in 2016, and
- Wesley William, who successfully applied for a scarce-skills postdoctoral fellowship from the NRF, and will initiate his studies in 2017.

Fruitful collaborations have been established with Prof Jeanine Marnewick and Dr Marilize Le Roes-Hill from the Cape Peninsula University of Technology (CPUT), Rooibos LTD (the largest rooibos processing company in South Africa) and small-holder farmers in the Cederberg Mountain region (supported by Northern Cape Department of Agriculture), as well as with Dr Ferreira from WestCape Biotech.

Rooibos sampling, conducted in spring and autumn 2016, resulted in the **identification of 43 morphologically diverse rooibos genotypes** from 19 locations in the Cederberg Mountains, that will be monitored and re-sampled throughout the study period (2016-2018).

Subsequent polyphenol profiling analyses of these plants revealed that some do not produce aspalathin (a rooibos-specific polyphenol that is associated with diverse health promoting properties). Such plants are essential for comparative analyses in search of genes involved in aspalathin production. In the process, my team has **developed a biochemical screening protocol** using thin layer chromatography (TLC) that reproducibly visualizes aspalathin. Considering that currently aspalathin identification in plant material requires HPLC (i.e. expensive equipment, laboratorial environment and knowhow), the TLC method represents a comparatively cheap and simple alternative for plant screening, which could potentially be employed by rooibos producers on site. These results are currently being prepared for publication.

In collaboration with Dr Ferreira, we have identified the rooibos genome size using flow cytometry.

Furthermore, we have benchmarked RNA and DNA extraction procedures in **preparation for rooibos transcriptome and genome sequencing**.

In preparation for sequencing data analysis, all members of my research team received training in biocomputational data analyses, including Lynux, Python programming, and biostatistics using R.

Accepted for publication in 2017:

Hesse, U., van Heusden, P., Kirby, B.M., Olonade, I., van Zyl, L.J. and Trindade, M., 2017. Virome assembly and annotation: a surprise in the Namib Desert. *Frontiers in Microbiology*, 8, p.13.

Community engagement during 2016 included, seven sampling trips where the projects were explained to the respective farmers and a meeting was organized at Rooibos LTD in Clanwilliam to report to the farmers on the progress of the projects.



Emily Stander and farmer Adam Christian



Freezing rooibos sample, Uljana Hesse and Jan Perang of Rooibos LTD.



Sampling rooibos in the Cederberg

Research Laboratory of Simon Travers



Some of the members of the lab involved in the development of the exatype drug resistance testing pipeline: Baruch Lubinsky, Simon Travers, Kirsten Miller-Duys, Imogen Wright and Andrew Patterson.

Year in review

2016 proved to be a very exciting year for my research group. In early January 2016 our spinoff company (**Hyrax Biosciences**) was invited to be a signatory on the Declaration by the Pharmaceutical, Biotechnology and Diagnostics Industries on Combating Antimicrobial Resistance that was launched at the World Economic Forum in Davos, Switzerland. We were the only African company to sign the declaration. Further, I was a contributor to the final report of the **UK Review on Antimicrobial Resistance** that was launched in May 2016.

In February 2016 we officially launched **exatype**, an online solution that facilitates the use of next-generation sequencing technologies for HIV drug resistance testing. There are now research and diagnostic

laboratories from all over the world using exatype routinely for their analysis. This was an incredibly proud moment for me as this launch was the result of almost five years of research by an excellent and dedicated team. In June 2016 the exatype team, represented by Dr Imogen Wright, was awarded second prize (US\$25,000) in the Innovation Prize For Africa at an award ceremony held in Botswana.

In July 2016 Dr Wright and myself travelled to Durban where we met with **Bill Gates and his global health team**. In this very productive meeting we presented exatype and described how we are expanding into new disease lines and moving towards a point of care TB diagnostic and drug susceptibility test. Further, we discussed our plans to develop a real-time disease incidence and drug resistance surveillance platform.

In September 2016 we received notification that we had been awarded a new **Strategic Health Innovation Partnerships (SHIP) award** by the South African Medical Research Council (in partnership with the South African Department of Science and Technology) to support the development of a TB drug susceptibility testing pipeline and a real-time disease incidence and drug resistance surveillance platform. Further, we received **funding for a further two years on the HIV SHIP award** to enable the development of a standalone version of exatype that will be capable of performing drug resistance testing locally on a point of care device.

We have also submitted a **worldwide patent** for our novel sequence alignment tool, the second patent to arise from the exatype project.

2016 was a very productive year for our HIV vaccine research projects. The project exploring the role of sugars bound to the surface of the HIV virion in vaccine susceptibility and escape has yielded fascinating results that will be published in early 2017.

Phillip Labuschagne was invited to present at the **HIV vaccine trial network workshop** held in Cape Town in October 2016 on his work developing software tools to process the large volumes of sequence data that are being generated as part of a number of HIV vaccine trials currently being run in South Africa. Further, Phillip and Roux-Cil Ferreira both presented at the **International HIV Dynamics and Evolution** meeting held in Woods Hole, Massachusetts in April 2016.

Research projects:

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

One of the biggest challenges with NGS is the complexity involved in the analysis of the data. This is particularly difficult for infectious

diseases, such as HIV, where it is important to be able to identify pertinent mutations in the infection in the presence of a lot of noise introduced by sample preparation and the sequencing process.

Our exatype™ platform is capable of identifying these mutations even when they are present at very low levels within a dataset. This capability enables us to identify drug resistance in viral populations with unrivaled sensitivity and has potential applications across a number of disease lines. 2016 saw the launch of the online version of exatype that facilitates the use of next-generation sequencing technologies for HIV drug resistance testing.

Further, we continued research in 2016 to expand the range of diseases that exatype can be applied to as well as increasing the number of sequencing platforms that can be supported. We also started the development of a standalone version of exatype for use in settings that cannot access a cloud based version. We have also started to develop a system that enables real-time surveillance of disease incidence and drug resistance using data that has been analysed through the exatype platform.

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

While NGS approaches are immensely powerful for sequencing the entire breadth of a whole genome (e.g. human genome) in a cost-effective and efficient manner, they also provide considerable power for generating 'ultra-deep' data enabling us to identify low abundance viral variants infecting individuals. We are involved in a number of research projects as part of the HIV vaccine trials network (HVTN) to develop analysis pipelines that will be used as part of HIV vaccine trials to study large volumes of sequence data that will be produced to evaluate the efficacy of the vaccines.

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

Recent studies have identified antibodies that are capable of neutralising a wide range of HIV strains, prompting interest in eliciting them as part of a vaccine response. These studies have shown that carbohydrates bound to the surface of HIV's gp120 surface protein form all, or part, of the epitope for many of these antibodies. Very little is known, however, about the structural dynamics of these carbohydrates

and how they may affect susceptibility to, and escape from, neutralisation by these carbohydrate binding antibodies.

We have developed an approach that enables us to model a fully glycosylated HIV gp120 trimer. Using this we can study the movement and interactions of these carbohydrates across time within a HIV infected individuals thereby providing valuable insights into the role that these carbohydrates play in the susceptibility to antibody-induced neutralization. This, therefore, allows us to explore the potential response within individuals to HIV treatments that target the gp120 trimer that is exposed on the surface of the HIV virion.

Research Collaborations:

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

Collaborating Parties:

Dr Oliver Grant - University of Georgia

Prof Robert Woods - University of Georgia

Dr Elisa Fadda - Maynooth University, Ireland

Prof Penny Moore - NICD, South Africa

Dr Jeffrey Dorfman - University of Cape Town, South Africa

Dr Natasha Wood - University of Cape Town, South Africa

Nature and purpose:

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

Output in the last 12 months:

Chinks in the armor of the HIV-1

Envelope glycan shield: Implications for immune escape from anti-glycan broadly neutralizing antibodies.

Thandeka Moyo, Roux-Cil Ferreira, Reyaaz Davids, Zarinah Sunday, Penny L. Moore, Simon A. Travers, Natasha T. Wood, Jeffrey R. Dorfman 2016. *Virology*.doi:10.1016/j.virol.2016.10.026

A number of presentations (oral and poster) at the 23rd International HIV Dynamics and Evolution meeting, Woods Hole, Massachusetts, US, the Warren Workshop that was held in Sapporo Japan as well as at the ISCB/ISMB conference held in Orlando.

Future Direction:

This research is continuing as part of the DST-NRF centre of excellence in HIV prevention awarded to CAPRISA. We are

currently advertising for a postdoctoral research fellow who will work on this team.

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

Collaborating Parties:

Prof Maria Papathanasopoulou - University of the Witwatersrand Medical School.

Dr Robert Shafer - Stanford University

Nature and purpose:

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

Output in the last 12 months:

A large number of presentations (oral and poster) at the 23rd International HIV Dynamics and Evolution meeting, Woods Hole, Massachusetts, US, the PyCon ZA 2016 held in Cape Town, the Gates Grand Challenges annual meeting, held in London, UK., BIO 2016, in San Francisco, and the Virus Genomics and Evolution meeting held in Cambridge UK.

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

Collaborating Parties:

Prof Lynn Morris - NICD, South Africa

Prof Carolyn Williamson - UCT, South Africa

Dr Paul Edlefsen - Fred Hutchinson Cancer Research Center, Seattle

The HIV vaccine trials network (HVTN)

Nature and Purpose:

We are involved in a number of research projects that are harnessing the power of NGS to enable us to fully understand the diversity of the viral quasispecies in HIV infected individuals. The vast majority of this work is focused on the identification

of the potential of individuals to produce broadly cross-neutralising antibodies as well as exploring the viral diversity of breakthrough infections following vaccination with HIV vaccine candidates.

Future Direction:

The development of pipelines and analysis of data will continue through 2017 as we move towards a number of HIV vaccine trials starting in South Africa over the next number of years.

Output in the last 12 months:

Structure of an N276-Dependent HIV-1 Neutralizing Antibody Targeting a Rare V5 Glycan Hole Adjacent to the CD4 Binding Site.

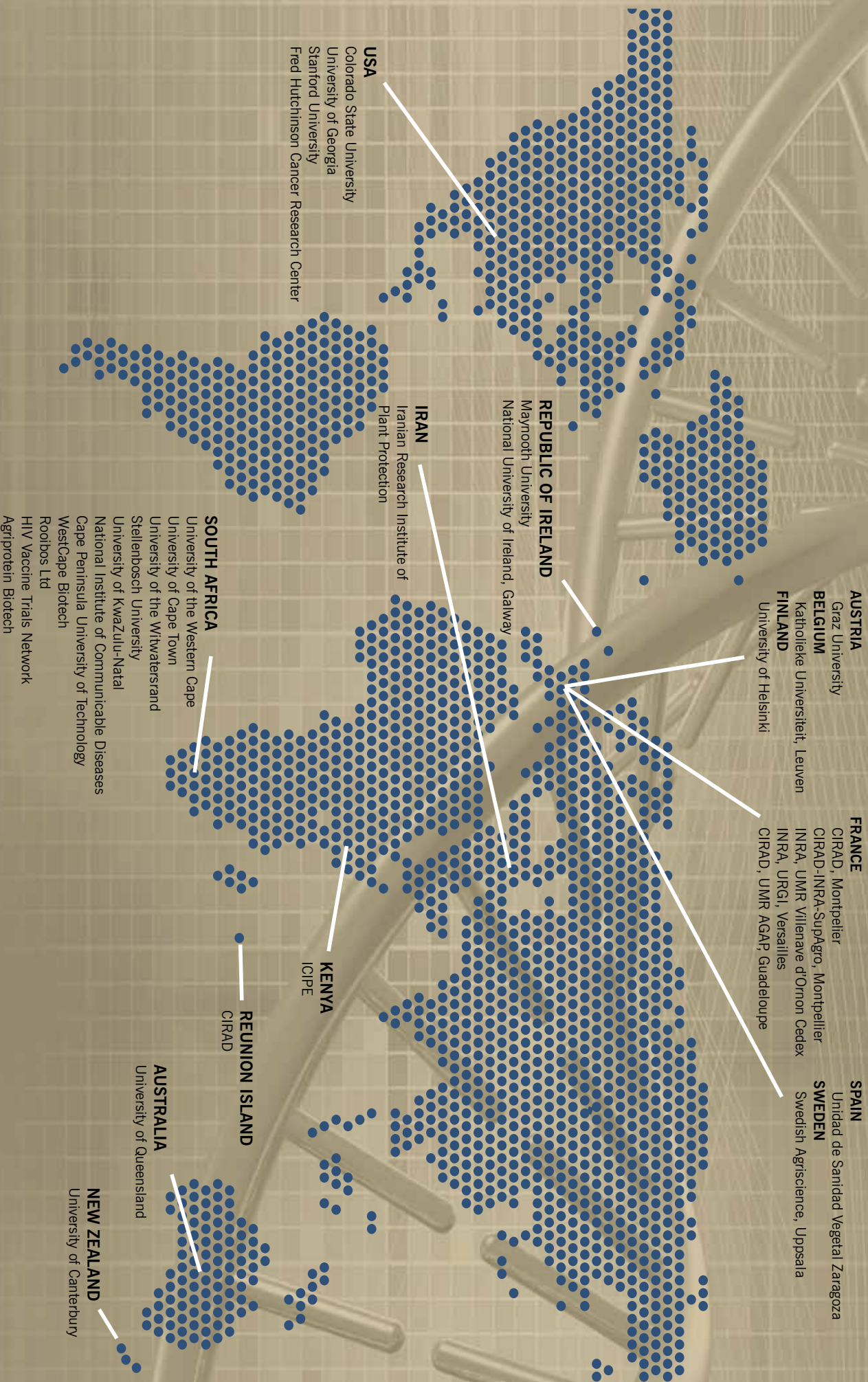
Wibmer C, Gorman J, Anthony C, Mkhize N, Druz A, York T, Schmidt S, Labuschagne P, Louder M, Bailer R, Abdool Karim S, Mascola J, Williamson C, Moore P, Kwong P, Morris L.

Journal of Virology 2016 vol: 90 (22) pp: 10220-10235.

A number of presentations (oral and poster) at the 23rd International HIV Dynamics and Evolution meeting, Woods Hole, Massachusetts, US as well as the HVTN Cape Town Next Generation Sequencing Workshop held in Cape Town.

Research Collaborations

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



2016 Financials

SANBI's income and expenditure trends for 2016 are shown in this section.

Figure 1. Distribution of income received from all sources.

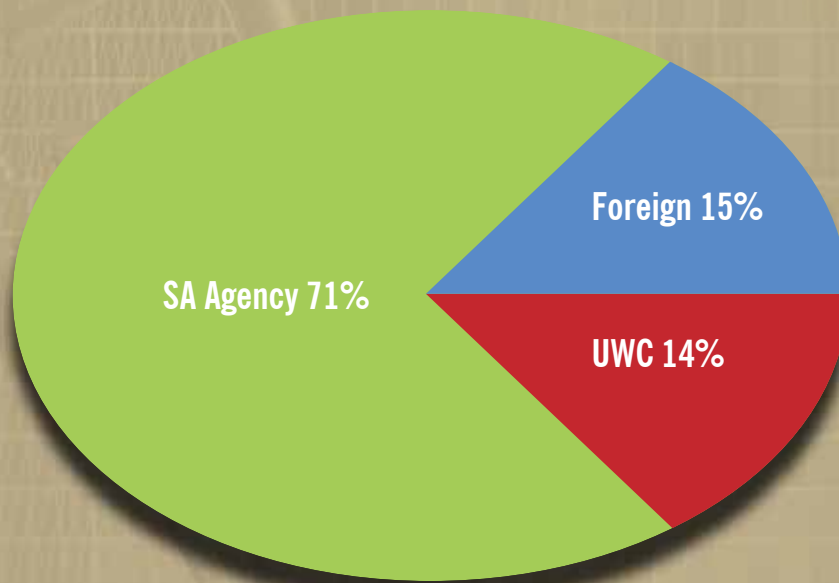


Figure 2. Income received from SA sources.

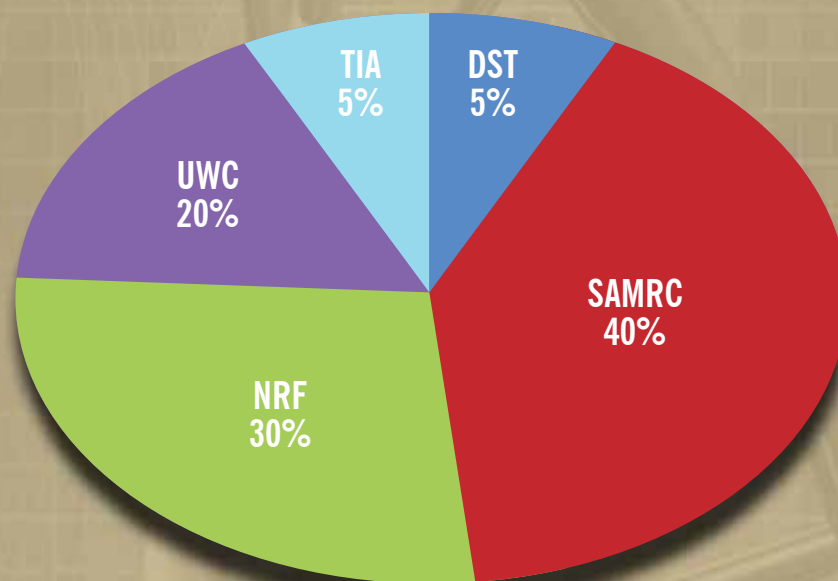


Figure 3. Income vs. Expenditure 2010 – 2016.

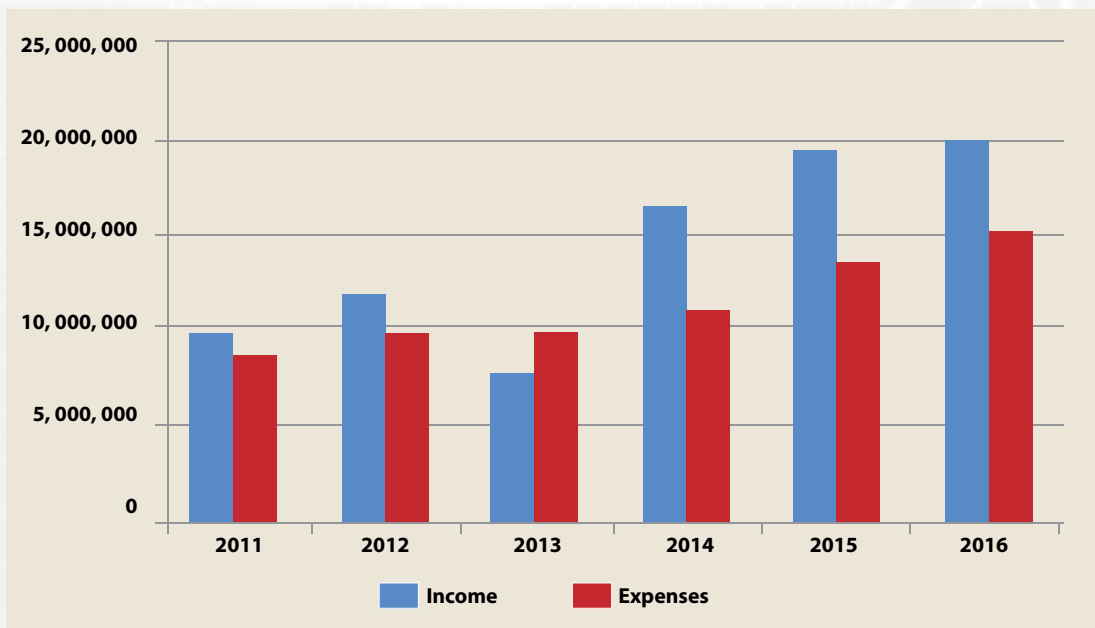
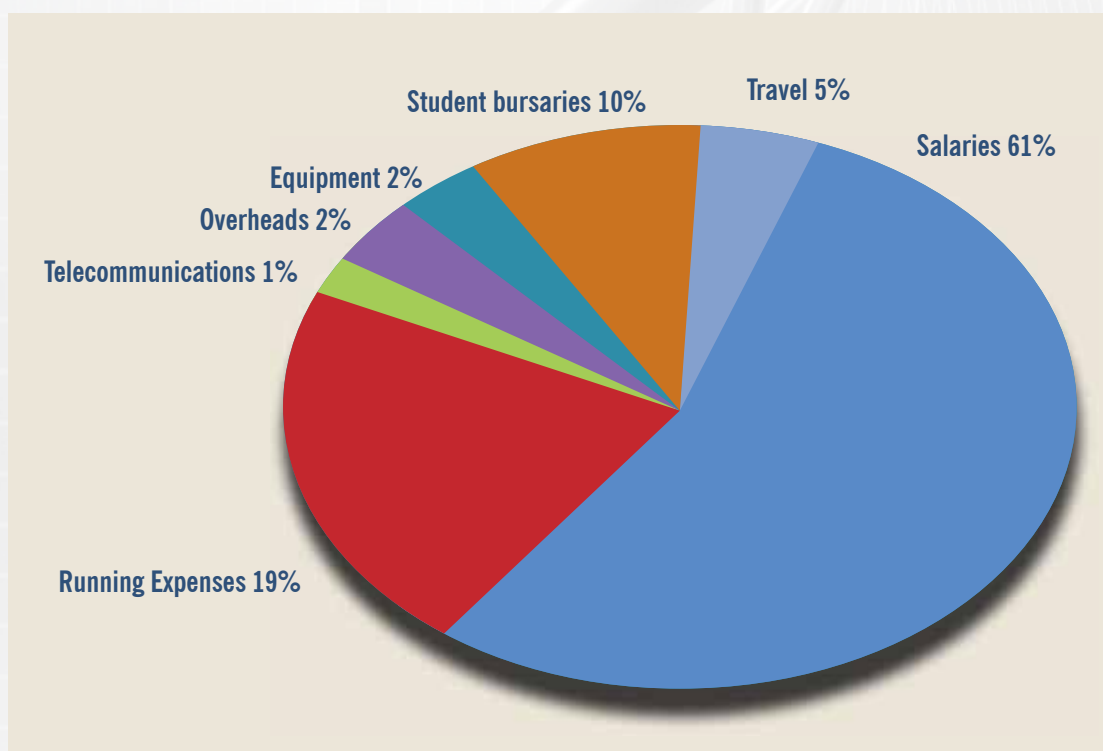


Figure 4. Expenditure in 2016.



Alumni

Staff:

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

Postdoctoral Fellows:

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

Adele Kruger	2010	Wayne State University, Detroit, Michigan
Oliver Hofmann	2010	Affiliated Faculty, Harvard Stemcell Institute, Associate Director at Harvard School of Public Health
Sundarajan Seshadri	2010	Nanyang Technology University, Singapore
Ashley Pretorius	2010	Senior Lecturer, Biotechnology, UWC
Sumir Panji	2012	H3ABioNet Project Manager, UCT
Samson Muyanga	2012	
John Pool	2012	University KwaZulu-Natal
Uljana Hesse	2013	Senior Lecturer, SANBI
Barbara Picone	2013	Italy
Monique Maqungo	2013	
Edwin Murungi	2013	Department of Biochemistry and Molecular Biology, Egerton University, Njoro, Kenya
Hannah Ajoge	2013	Postdoctoral Fellow, University of Western Ontario, Canada
Natasha Wood	2014	Lecturer, UCT Computational Biology Department
Sarah Mwangi	2016	Karolinska Institute, Sweden
Zahra Jalali	2016	Postdoctoral Fellow, University of British Columbia, Centre for Molecular Medicine and Therapeutics
Mahjoubeh Jalali	2016	Postdoctoral Fellow, Pacific Northwest Diabetes Research Institute, Washington
Colleen Saunders	2016	Junior Research Fellow, University of Cape Town

PhD:

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

MSc:

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

Honours:

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

Funders



UNIVERSITY of the
WESTERN CAPE

SANBI
South African National
Bioinformatics Institute

Research Institute at the
University of the Western
Cape since 1997



**National
Research
Foundation**

National Research
Foundation funding since 1998



South African Medical
Research Council
Bioinformatics Unit
since 2002





science
& technology
Department:
Science and Technology
REPUBLIC OF SOUTH AFRICA




**National
Research
Foundation**

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



National Institutes of Health



H3ABioNet Node 2012 -2016

**BILL & MELINDA
GATES foundation**

Bill and Melinda Gates
Foundation funding
since 2014



European
Commission

Horizon 2020
European Union funding
for Research & Innovation

B3 Africa funding since 2015



technology innovation
AGENCY

Technology Innovation
Agency funding since
2016



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 | South Africa**

Telephone: +27 (0)21 959-3645

Facsimile: +27 (0)21 959-2512

Email: info@sanbi.ac.za

www.sanbi.ac.za



**UNIVERSITY of the
WESTERN CAPE**