SASBI | SAGS

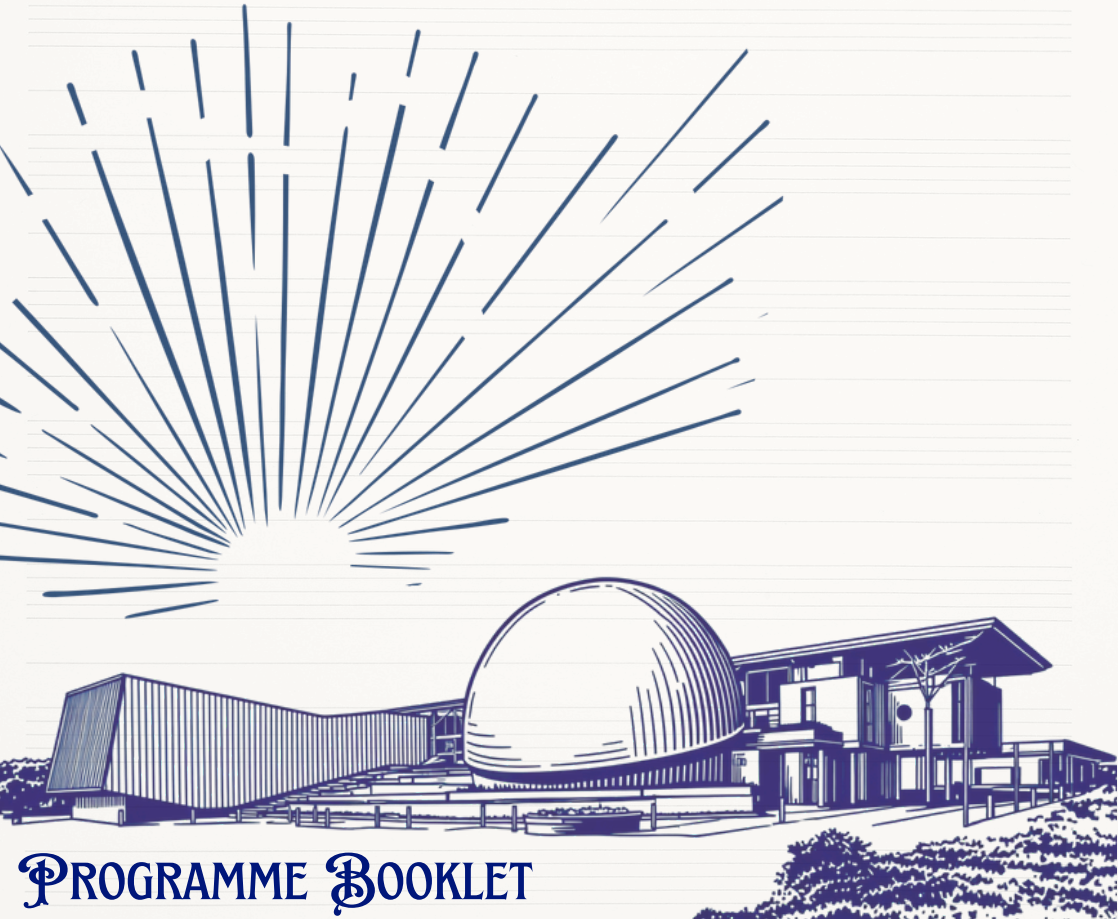
BIO 2026

STUDENT SYMPOSIUM

Building An Inclusive Bioinformatics Ecosystem

Nelson Mandela University, Gqeberha

31 August 2026



PROGRAMME BOOKLET

NELSON MANDELA UNIVERSITY

South African Society for Bioinformatics

SAGS South African Genetics Society

DIPLOMICS CONNECTING RESEARCHERS & WORLD CLASS OMICS FACILITIES

KSA Decoding South Africa's Biodiversity

ABI African Bioinformatics Institute

A WARM WELCOME FROM THE ORGANISING COMMITTEE

Dear attendees, colleagues and guests,

It is our distinct privilege to welcome you to the SASBi | SAGS BIO2026 STUDENT SYMPOSIUM. This year, under our umbrella theme, "African Genomes in the Global Village: Landscapes to celebrate diversity" we gather in the beautiful coastal city of Gqeberha (Port Elizabeth) to explore cutting-edge genomic advancements, foster cross-discipline collaborations and celebrate diversity inherent to African genomic landscapes.

This booklet serves as your roadmap for the event, inside you find the session schedules, bios of our esteemed keynotes and panelists, and local insights to help you enjoy your stay.

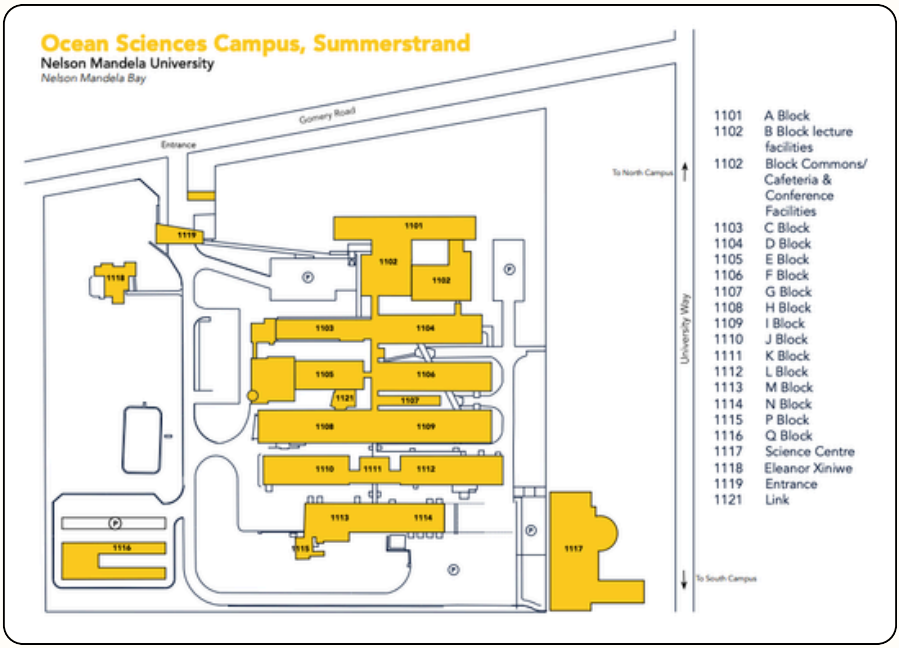
Thank you for bringing your expertise and passion to this African dialogue.

Sincerely,

Caivil Ndobela and Joshua Sampson
Chairperson, Organising Committee

VENUE INFORMATION

The SASBi | SAGS BIO2026 Student Symposium will be held at the Digital Dome on Ocean Science Campus, Summerstrand, Nelson Mandela Bay, Nelson Mandela University.



CONNECTIVITY

Wi-fi Network Name: wifi05@mandela.ac.za (wifi-zero-five)
Wi-Fi password: IneAqu#99

1. Launch the GetEduroam app
2. In the Search Bar, search for Nelson Mandela University and click on Connect
3. You will be prompted for your username and password in the format username@mandela.ac.za (wifi01@mandela.ac.za in this case).
4. Click Log In
5. You willbe prompted to add a Wi-Fi hotspot to your device. Click Allow to let the app create the eduroam Wi-Fi profile.
6. You will then be informed that the GetEduroam app wants to join the eduroam Wi-Fi network (provided you are within range of an eduroam hotspot). Select Join to connect to the
eduroam Wi-Fi network.

The steps above are generic and may differ slightly from device to device

EMERGENCY AND HEALTH CONTACTS

On-site first aid:
Report directly to the registration desk or call
Closest Hospital: Life St. George's Hospital.
General Emergency Numbers:
Police: 10111
Ambulance: 10177 or 112

CONNECT WITH US!!

#BIO26inNMU #BIOINFORMATICS #GENETICS

WHERE TO FIND US

LinkedIn: SASBi | SAGS Student Bio-Symposium 2026

Instagram: @student.biosymposium

BlueSky: @biosymposium2026.bsky.social

KEYNOTE PRESENTATIONS

Never Stop Evolving: Lessons from a Life in Bioinformatics

Bioinformatics is a field that never stands still and neither can the people working in it. From an unlikely journey into science to a career shaped by constantly evolving technologies, this keynote reflects on the unexpected turns, challenges and opportunities that have shaped one life in bioinformatics. More than a story of where we have come from, it is an invitation to embrace curiosity, adapt to change and never stop evolving.

Dr Rian Pierneef

Senior Lecturer in Genetics and Bioinformatics, University of Pretoria.
President of the South African Society for Bioinformatics.

Building the Future of African Bioinformatics: Innovation, Collaboration, and Impact

Bioinformatics in Africa has advanced significantly over the last 2 decades. In addition to capacity developed in individual countries, H3ABioNet had an enormous impact, not only for bioinformatics scientists and users within the consortium, but well beyond the project through its novel and wide-reaching training program. This pan African bioinformatics network built data and computing infrastructure, tools and databases, and most importantly, a community of highly skilled bioinformaticians. However, the scientific landscape in Africa and globally is changing, with the increasing generation of complex datasets and the emergence of evermore sophisticated AI techniques. These provide both challenges and opportunities for our next generation of bioinformatics scientists. In this talk, I will reflect on the trajectory of bioinformatics in Africa, what has been achieved, and how we can address the remaining challenges to make the most of the new opportunities. I will also discuss the role of emerging researchers in shaping the future of bioinformatics on the continent.

Prof Nicola Mulder

Head of the Computational Biology Division at the University of Cape Town (UCT).

PROGRAMME

MONDAY, 31 AUGUST 2026
THEME: AFRICAN GENOMES IN THE GLOBAL VILLAGE: LANDSCAPES TO CELEBRATE DIVERSITY

08:00 - 09:00	REGISTRATION/LOGIN Link: TBC	
09:00 - 09:10	Welcome and housekeeping	**Caivil Ndobela**, President of SASBi-SC
09:10 - 09:45	**Opening Keynote address:** Never Stop Evolving: Lessons from a Life in Bioinformatics **Dr Rian Pierneef** President of the South African Society for Bioinformatics **Venue:** Digital Dome and Teaching	
09:45 - 10:30	**Session 1: Oral Presentations**	
	Session 1A: Human Landscapes: Population Genomics, Immunity, and health diversity **Focus:** Population-level genetic diversity, host-immune responses, and molecular signatures associated with infectious disease susceptibility. **Session chair: Nozipho Magagula** **Co-Facilitator: Odireleng Mosuwe** **Venue:** Digital Dome and Teaching	**Session 1B:** Pathogen Landscapes: Host-Pathogen Interactions, Infectious Disease, and Genomic Diversity **Focus:** Molecular and genomic approaches to understanding host–pathogen interactions, pathogen diversity, and mechanisms underlying infectious disease. **Session chair: Sandile Ntloko** **Co-Facilitator: Thulani Nkosi** **Venue:** Training laboratory
09:45 - 10:00	Phylogenetic age is associated with mitochondrial heteroplasmy in African and diaspora populations - **Dayna Adrienne Croock**	Characterisation of the Subcellular Localisation of African Horse Sickness Virus Non-structural Protein NS5 - **Nipho Mtambo**
10:00 - 10:15	CCR5 regulatory variants and novel loci associated with susceptibility to HIV-1 infection in Sub-Saharan African populations - **Isabel du Randt**	Genetic Diversity of Ehrlichia Ruminantium Based on Multilocus Sequence Typing of Isolates from Indigenous Goats and Amblyomma Hebraeum Ticks - **Xolile Nuse**
10:15 - 10:30	Sex-Specific Gene Expression Signatures as Biomarkers in predicting Tuberculosis Relapse - **Bongani Mnyandli**	Transcriptomic Characterisation of Differentiation-driven Priming of the Interferon-a Response in THP-1 macrophages - **Boithuso Phale**

10:30 - 10:50	**Tea Break**	
10:55 - 11:55	**Panel discussion:** Building an inclusive and equitable bioinformatics ecosystem **Session Chair**: **Joshua Sampson**, President of the SAGS-Student Society **Panel Speakers: Prof. Scott Hazelhurst (WITS), Prof. Tastan-Bishop (RU), Lebogang Siviya(NMU), Mr. Siyanda Mazibuko (NMU), Prof. Carminita Frost (NMU)** **Venue:** Digital Dome and Teaching Link: TBC	
11:55 - 12:05	**Picture Moment #BIO26inNMU #Bioinformatics #Genetics**	
12:05 - 12:50	**Lunch**	
12:50 - 13:50	**Session 2: Oral Presentations**	
	Session 2A: Computation Landscapes: Machine Learning, Bioinformatic Pipelines, and Data Discovery **Focus:** Advanced algorithmic approaches, neural networks, automated screening workflows, and sequence classifiers applied to oncology and genomic tracking **Session chair: Sandile Ntloko** **Co-Facilitator: Odireleng Mosuwe** **Venue:** Digital Dome and Teaching	**Session 2A:** Ecological Landscapes: Wildlife conservation, Environmental Health, and Ecosystem Diversity **Focus:** Tracking diversity across diverse biomes, spanning African wildlife conservation, river systems, aquaculture and agricultural ecology **Session chair: Thulani Nkosi** **Co-Facilitator: Nozipho Magagula** **Venue**: Digital Dome and Teaching
12:50 - 13:05	Graph Neural Network-Based Discovery of miRNA-mRNA Regulatory Interactions in Breast Cancer - **Nothando Gama**	Surveillance and Monitoring of Antimicrobial-Resistant Human Pathogenic and Zoonotic Bacteria in the Ngamiland and Okavango Delta - **Patience Motshosi**
13:05 - 13:20	Reconstructing Cervical Cancer Tumour Microenvironment States from DNA Methylation Profiles Using a Machine Learning Framework - **Saltiel Hamese**	Temporal Dynamics of the Gut Microbiome During Smoltification in Atlantic Salmon (Salmo Salar) - **Muktaar Badrorin**
13:20 - 13:35	DockVina: An Automated Molecular Docking Pipeline for High Throughput Virtual Screening - **Siphamandla Mnakekeli Dlamini**	Unravelling the Elusive Mating Strategy of Dematophora Necatrix: A Combined Genomic and Physiological Approach - **Annabel Norval**

13:35- 13:50	Taxonomic Binning of Long-Read DNA Sequences by K-mer Composition using Decision Tree Classifiers - **Bruhan Kyomuhendo**	
13:50- 14:00	**Comfort Break**	
14:00- 15:00	**Session 3:** Flash talks **Session Chair: Joshua Sampson**, President of SAGS SC and **Sandile Ntloko**, Deputy Development of SASBi-SC **Facilitators: Odireleng Mosuwe** and **Caivil Ndobela** **Venue:** Digital Dome and Teaching	
14:00 - 14:07	Advancing avian genomics: A de novo assembly pipeline to generate a high-quality reference genome for the Southern Bald Ibis (Geronticus calvus) - **Ashley Colleen Mortimer**	
14:07 - 14:14	Does multi-task supervision teach a classifier where to look? A quantitative attention-localization study in brain tumor MRI - **Tisetso Letuka**	
14:14 - 14:21	The Role of the Gut Microbiome in Inflammation and Multimorbidity in African Populations: Evidence from AWI-GEN - **Aaron Berkman**	
14:21- 14:28	Impact of Reference Genome, Alignment Strategy, and Quantification Method on Locus-Specific Human Endogenous Retrovirus (HERV) Detection During Monocyte-to-Macrophage Differentiation - **Tebogo Phago**	
14:28- 14:35	Integrated Phenotypic Profiling and Therapeutic Target Prioritization of Clinical Multidrug-Resistant Pseudomonas aeruginosa Exhibiting Biofilm-Driven Persistence - **Sinethemba Yakobi**	
14:35- 14:42	Benchmarking DNBSEQ-T7 Sequencing Performance Using Genome in a Bottle Reference Genomes: Impact of Library Preparation Strategies - **Maano Malima**	
14:42- 15:00	**Session Wrap Up Q&A** **Session Chair: Nozipho Magagula,** Deputy secretary & Treasurer of SASBI-SC	

<table>
<tr><td>15:00 - 15:30</td><td colspan="2">Closing Keynote address: Building the Future of African Bioinformatics: Innovation, Collaboration, and Impact

Prof Nicola Mulder
Head of the Computational Biology
(CBIO) division at the University of Cape Town, Principal investigator of H3ABioNe, and ABI Founding member

Venue: Digital Dome and Teaching

Link: TBC</td></tr>
<tr><td>15:30- 15:50</td><td colspan="2">Sponsor talk: Diplomics</td></tr>
<tr><td>15:50- 16:30</td><td colspan="2">Prize Giving</td></tr>
<tr><td>16:30 - 17:00</td><td>Annual general meeting (AGM) SASBi
Session chair: Caivil Ndobela, President of SASBi SC

Venue: Auditorium 1

1. Welcome and Opening
2. Confirmation of Quorum
3. Approval of Previous AGM Minutes
4. Chairperson’s Report
5. Introduction of the Society’s Goals and Strategic Direction
6. Introduction of Council Roles and Responsibilities
7. Constitutional and Governance Matters
7.1. Identification of amendments required to align the Society’s Constitution with its intended NPO registration
7.2. Recommendation for the incoming Council to review and propose the necessary constitutional amendments
8. Call for Nominations
9. Election of New Council
10. Incoming Chairperson’s Address
11. Closing</td><td>Annual General Meeting (AGM) SAGS
Session chair: Joshua Sampson, President of SAGS SC

Venue: Auditorium 2

1. Welcome and opening
2. Confirmation of quorum
3. Approval of previous AGM minutes (if available)
4. Chairperson’s report
5. Treasurer’s report
6. Introduction of society’s goal and message
7. Introduction of council roles and responsibilities
8. Call for nominations
9. Election of new council
10. Incoming chairperson’s address and closing</td></tr>
</table>

MEET THE KEYNOTE SPEAKERS

OPENING KEYNOTE SPEAKER

Dr Rian Pierneef

Senior Lecturer in Genetics and Bioinformatics, University of Pretoria.

Dr. Rian Pierneef is a Senior Lecturer in the Department of Biochemistry, Genetics and Microbiology at the University of Pretoria and is affiliated with the Centre for Bioinformatics and Computational Biology (CBCB). His teaching and research focus on bioinformatics, genomics, and computational biology, with particular interests in microbial genomics, metagenomics, transcriptomics, and the development of reproducible bioinformatics workflows for next-generation sequencing data analysis. He is passionate about interdisciplinary research and the application of computational approaches to address challenges in human health, agriculture, and environmental microbiology. Dr. Pierneef is dedicated to mentoring students and advancing bioinformatics education and research in South Africa.

CLOSING KEYNOTE SPEAKER

Prof Nicola Mulder

Head of the Computational Biology Division at the University of Cape Town (UCT).

Prof Mulder heads the Computational Biology Division at UCT. She is PI of the DS-I Africa eLwazi Open Data Science Platform and co-leads a Sickle Cell Disease Data Coordinating Centre and a Wellcome Trust Discovery Data Integration Platform at UCT. She led H3ABioNet, a Pan African Bioinformatics Network which developed bioinformatics capacity to enable genomic data analysis for more than a decade. Her research focuses on genetic determinants of susceptibility to disease, African genome variation, and microbial genomics and infectious diseases. Her group provides bioinformatics services and training and develops new algorithms and resources for African genomics. Prof Mulder is actively involved in capacity development in Bioinformatics and genomics. She also sits on a number of international scientific advisory boards and is currently the interim Director for the African Bioinformatics Institute.

Prof Carminita Frost

Professor of Biochemistry at Nelson Mandela University.

Carminita Frost is a Professor of Biochemistry at Nelson Mandela University (NMU, Gqeberha, South Africa) and a scientist in bioprospecting and phytomedicine translation. Specializing in the biochemical evaluation and functional validation of indigenous South African flora, her research platform bridges traditional ethnobotanical knowledge with modern drug discovery. Prof. Frost's current program focuses on biotech innovation in the medicinal cannabis sector, where her group identifies bioactive pipelines and elucidates their molecular mechanisms of action against chronic metabolic and targets, including diabetes and cancer. Notably, her laboratory is work on Cannabis sativa, investigating novel, synergistic cannabinoid formulations engineered to induce programmed cell death in breast cancer lines via paraptosis—a unique, non-apoptotic pathway with significant therapeutic potential for overcoming drug resistance. An active leader in the scientific community, Prof. Frost has served on the South African Society of Biochemistry and Molecular Biology (SASBMB) council and maintains a robust network of national and international academic and industry collaborations. The work completed in her group includes publications and she enjoys mentoring the next generation of biotech talent.

Prof Özlem Tastan-Bishop

Full Professor of Structural Bioinformatics at Rhodes University, South Africa, and a Distinguished Adjunct Professor at Saveetha University, Chennai, India.

She obtained her BSc degree in Physics from Boğaziçi University, Istanbul, Turkey, before completing an MSc in Molecular Biology and Genetics at the same institution. She received her PhD in 2003 from the Max Planck Institute for Molecular Genetics and the Free University of Berlin, Germany. During her doctoral studies, she developed a strong interest in structural biology, which she further pursued through postdoctoral appointments at the University of Texas, USA, the University of the Western Cape, and the University of Pretoria, South Africa, where she gained expertise in structural biology and structural bioinformatics. In 2009, Prof. Tastan Bishop joined Rhodes University as an academic staff member. In 2013, she founded the Research Unit in Bioinformatics (RUBi), which has since become a leading research group in structural bioinformatics.

Since joining Rhodes University, she has supervised and graduated 29 PhD and 42 MSc students. In recognition of her research excellence, she received the Rhodes University Vice-Chancellor's Distinguished Senior Research Award in 2020 and the South African Society for Bioinformatics (SASBi) Silver Award in 2022.

Prof. Tastan Bishop is a Fellow of the Royal Society of South Africa (FRSSAf), where she serves as a Council Member and Eastern Cape Representative. She is also a Member of the Academy of Science of South Africa (MASSAf) and an Associate of the National Institute for Theoretical and Computational Sciences (NITheCS). She serves on the Editorial Boards of PLOS ONE, Frontiers in Molecular Biosciences, and Frontiers in Applied Mathematics and Statistics (Biological Modeling and Simulation section). In addition, she has acted as a reviewer for numerous leading international peer-reviewed journals.

Her research focuses on structural bioinformatics and its applications to drug discovery and development. Her recent work centres on understanding protein allostery, elucidating the structural and functional effects of nonsynonymous single nucleotide variants, and integrating artificial intelligence with structural bioinformatics to advance drug discovery and pharmacogenomics. She has published over 100 peer-reviewed research articles.

Prof Scott Hazelhurst

Professor of Bioinformatics in the School of Electrical and Information Engineering at Wits.

Scott Hazelhurst is professor of bioinformatics in the School of Electrical & Information Engineering and Senior Scientist at the Sydney Brenner Institute for Molecular Bioscience at the University of the Witwatersrand, Johannesburg. He is is Co-PI of the Data Science for Health Discovery and Innovation in Africa MADIVA Research Hub, and a co-investigator on the eLwazi Open Data Science Platform, and collaborated on several H3Africa projects, co-leading the AWI-Gen Microbiome Project. His areas of interests include bioinformatics, health informatics and high performance computing. Scott received his honours and MSc degrees from Wits and his PhD from the University of British Columbia.

Ms Lebogang Siviya

PhD Candidate, Nelson Mandela University.

Lebogang Siviya is a PhD candidate in Botany at Nelson Mandela University, where her research focuses on using multi-omics approaches to understand the molecular responses of sweet potato to abiotic stress. She holds an MSc in Molecular and Cell Biology (with distinction) from the University of the Witwatersrand, where she investigated virus–host interactions in tomatoes. Her work aims to contribute to crop improvement and food security by advancing our understanding of the molecular mechanisms underlying crop adaptations to the environment. Her research interests include plant biology, biotechnology and bioinformatics with a focus on integrating these disciplines to address agricultural challenges. Beyond her research, Lebogang is actively involved in mentoring students in molecular biology, genetics, plant biotechnology and research.

Mr Siyanda Mazibuko

Associate Lecturer, Nelson Mandela University.

Siyanda Mazibuko is an Associate Lecturer in the Department of Biochemistry and Microbiology at Nelson Mandela University. His MSc research focuses on antimicrobial resistance (AMR) surveillance in wastewater using molecular microbiology, microbial genomics, and bioinformatics. He employs PCR, Sanger sequencing, and Oxford Nanopore sequencing to investigate antimicrobial resistance determinants in clinically relevant bacteria. Prior to his current appointment, Siyanda served as a Postgraduate Scholarship and Ethics Administrator within Nelson Mandela University's Department of Research Development. He also completed specialised training in digital PCR through the Centre for Epidemic Response and Innovation (CERI) and QIAGEN, strengthening his expertise in molecular diagnostics and genomic technologies. In addition to his research, Siyanda teaches undergraduate biochemistry and has been recognised for his research presentations at both university and departmental level, including receiving an award in the university wide Three Minute Thesis (3MT) competition. Through his research, teaching, and continued professional development, he is committed to advancing genomic technologies to strengthen antimicrobial resistance surveillance and improve public health.

PRESENTATION ABSTRACTS

PHYLOGENETIC AGE IS ASSOCIATED WITH MITOCHONDRIAL HETEROPLASMY IN AFRICAN AND DIASPORA POPULATIONS

Dayna Adrienne Croock*, Elouise Kroon[1], Devon Allies[1], André G. Loxton[1], Marlo Möller[1,2,3,4], Caitlin Uren[1,2,3,4], Desiree C. Petersen[1,2,3,4]

[1]South African Medical Research Council (SAMRC) Centre for Tuberculosis Research (CTR), Division of Molecular Biology and Human Genetics (MBHG), Department of Biomedical Sciences, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa, [2]Centre for Bioinformatics and Computational Biology, Stellenbosch University, Cape Town, South Africa, [3]Genomics for Health in Africa (GHA), Africa-Europe Cluster of Research Excellence (CoRE), South Africa, [4]National Institute for Theoretical and Computational Sciences (NITheCS), South Africa, *dayna@sun.ac.za

Background: Mitochondrial DNA (mtDNA) heteroplasmy—the co-existence of multiple mtDNA variants within an individual—has been linked to ageing, disease susceptibility, and cellular dysfunction. However, the evolutionary and population-level factors governing heteroplasmy burden remain poorly understood. Recent work in green sea turtles demonstrated that individuals carrying phylogenetically derived mtDNA haplotypes accumulate more heteroplasmic variants than those carrying ancestral haplotypes, suggesting that lineage age may influence heteroplasmy burden. Here, we test whether a similar phylogenetic age effect exists in humans using whole mitochondrial genome sequencing data from 1,306 individuals representing African and diaspora populations. Using negative binomial and Poisson generalised linear models with L0—the most ancestral human mtDNA macrohaplogroup—as the reference, we find that carriers of L1, L2, and L3 haplogroups exhibit significantly higher heteroplasmy counts (incidence rate ratios 2.71–3.41, all FDR-adjusted $p < 10^{-39}$). This pattern persists after controlling for broad geographic region and is replicated in a diaspora-only subset with shared demographic history, suggesting that the association is not solely explained by population structure. Importantly, using calibrated sub-haplogroup age estimates, we identify a significant negative relationship between sub-haplogroup age and heteroplasmy burden within L0 itself (IRR = 0.978 per thousand years, $p = 3.8 \times 10^{-14}$). Each additional 1,000 years of lineage age is associated with approximately 2.2% fewer heteroplasmic variants. This within-lineage gradient, independent of between-haplogroup differences, provides strong evidence that phylogenetic age contributes to shaping mtDNA heteroplasmy burden in humans, suggesting that long-term evolutionary history influences contemporary patterns of mitochondrial variation.

TRANSCRIPTOMIC CHARACTERISATION OF DIFFERENTIATION-DRIVEN PRIMING OF THE INTERFERON-A RESPONSE IN THP-1 MACROPHAGES

Boithuso Phale[1*], Prumohdya D. Muamba[1], Vanessa Meyer[1], Nikki L. Gentle[1]

[1]School of Molecular and Cell Biology, University of the Witwatersrand, *2451444@students.wits.ac.za

Monocytes and macrophages are important coordinators of the innate response to infectious disease. During infection, monocytes are recruited to tissues where they differentiate into macrophages, which mount interferon (IFN)-driven antiviral responses through the expression of IFN-stimulated genes (ISGs). While ISG expression is classically attributed to IFN signalling, the contribution of monocyte-to-macrophage differentiation to priming this response remains poorly understood. The aim of this study was to investigate how differentiation primes the IFN response in macrophages, using THP-1 cells as a model system. To define the macrophage IFN response, differential gene expression (DGE) was performed with DESeq2 on RNA-sequencing data from THP-1 macrophages differentiated with PMA for 24 hours and stimulated with IFN-α2 for four hours. This identified 978 significantly differentially expressed ISGs (padj < 0.05, |log2FC| > 1). To assess ISG dynamics during differentiation, Cap Analysis of Gene Expression sequencing (CAGE-seq) data spanning 96 hours of THP-1 monocyte-to-macrophage differentiation were analysed using CAGEfightR. Promoters were defined as transcription start site clusters supported in at least two samples within 20 bp, while enhancers were defined as bidirectional clusters supported in at least two samples. Differential expression analysis of promoter- and enhancer-associated genes revealed progressive ISG induction, increasing from 24 differentially expressed ISGs during early differentiation to 173 ISGs by 24 hours without cytokine stimulation. Hierarchical clustering of ISG expression revealed distinct temporal patterns, including early-induced and sustained, late-induced, and transiently induced ISGs. Motif-enrichment analysis with motifmatchr and JASPAR identified EGR1, EGR2, SP1, and IRF8 as potential key regulators of this ISG priming. Together, these findings suggest that differentiation progressively rewires ISG expression through distinct regulatory waves that prime macrophages for a robust IFN response.

CCR5 REGULATORY VARIANTS AND NOVEL LOCI ASSOCIATED WITH SUSCEPTIBILITY TO HIV-1 INFECTION IN SUB-SAHARAN AFRICAN POPULATIONS

Isabel du Randt[1,2,3*] Wenlong C. Chen[1,2,7,21] Irene Namara[4,5,6], Luicer A. Ingasia Olubayo[1], Alexander J. Mentzer[13,14], Mazvita Muchengeti[7,8,9,21], Chantal B. de Villiers[3], Rob Newton[19,14,] Freddy Sitas[15,16,17], Debbie Bradshaw[15], Tim Waterboer[20], June Fabian[18,19], Segun Fatumo[4,19,11], Michèle Ramsay[1,3], Christopher G. Mathew [1,3,21], Jean-Tristan Brandenburg[1,2,21]

[1]Sydney Brenner Institute for Molecular Bioscience, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa, [2]Strengthening Oncology Services Research Unit, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa, [3]Division of Human Genetics, School of Pathology, NHLS and Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa, [4]MRC/UVRI and LSHTM Uganda Research Unit, Entebbe, Uganda, [5]Makerere University, Kampala, Uganda, [6]Institute of Translational Genomics, Helmholtz Zentrum München, German Research Center for Environmental Health, 85764 Neuherberg, Germany, [7]National Cancer Registry, a Division of the NICD, NHLS, Sandringham, Johannesburg, South Africa, [8]School of Public Health, University of the Witwatersrand, Johannesburg, South Africa, [9]South African DSI-NRF Centre of Excellence in Epidemiological Modelling and Analysis (SACEMA), Stellenbosch University, Stellenbosch, South Africa, [10]Medical Research Council/ Uganda Virus Research Institute/London School of Hyglene and Tropical Medicine (MRC/UVRI/LSHTM) Uganda Research Unit, Entebbe, Uganda, [11]The Department of Non-communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine London, London, UK, [12]The Centre for Human Genetics, University of Oxford, Oxford, UK, [13]Chinese Academy of Medical Science Oxford Institute, University of Oxford, Oxford, UK, [14]University of York, YO10 5DD York, UK, [15]Burden of Disease Research Unit, South African Medical Research Council, Cape Town 7505, South Africa, [16]International Centre for Future Health Systems, University of New South Wales, Sydney Australia, [17]School of Population Health, University of New South Wales, Sydney, Australia, [18]Medical Research Council/Wits University Rural Public Health and Health Transitions Research Unit (Agincourt), School of Public Health, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, Gauteng, South Africa, [19]Wits Donald Gordon Medical Research Institute, Faculty of Health Sciences, University of Witwatersrand, Johannesburg, Gauteng, 2193, South Africa, [20]Infections and Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany, [21]Network for Oncology Research in Africa (NORA), Global Health Working Group, Martin-Luther-University, Halle-Wittenberg, Germany *fourielsabet1@gmail.com

Background: Human immunodeficiency virus type 1 (HIV-1) infection and sequelae remain a major global health challenge, with Sub-Saharan Africa (SSA) bearing the highest disease burden. Despite evidence that host genetic factors influence HIV-1 susceptibility, most genome-wide association studies (GWAS) have been conducted in non-African populations, leaving the genetic basis of susceptibility in SSA largely unexplored. High HIV-1 prevalence in SSA increases the likelihood that controls were exposed but uninfected, improving the interpretability of case-control analyses compared with studies in low-prevalence populations. **Methods:** We conducted the largest GWAS to date of susceptibility to HIV-1 infection in individuals of SSA ancestry, including 3,969 HIV-1 positive cases and 14,291 HIV-1 negative controls, from multiple study populations from South and East Africa. Fine-mapping and colocalisation analysis with blood expression quantitative trait loci (eQTLs) from the Southern African Blood Regulatory (SABR) resource were used to evaluate functional mechanisms underlying associated loci. **Results:** Three independent loci reached genome-wide significance for HIV-1 infection: rs2040388 and rs2227010 on chromosome 3 (in CCR5AS and upstream of CCR5), rs112358207 on chromosome 1 (in ESRRG), and rs113298445 on chromosome 8 (in INTS9). The chromosome 3 signal colocalised with eQTL variants for CCR5AS expression. Although the protective CCR5Δ32 mutation is absent in African populations, our findings identified alternative variants in the CCR5 regulatory region that may modulate susceptibility to HIV-1 infection. Conclusions: This study revealed both shared and population-specific genetic determinants of susceptibility to HIV-1 infection, underscoring the central role of CCR5 regulation in host defence. By expanding genomic studies to African populations, we identify novel loci and refine our understanding of HIV-1 infection in the populations most affected by the epidemic. Further replication and functional analyses are needed to establish causal mechanisms and evaluate translational relevance.

SEX-SPECIFIC GENE EXPRESSION SIGNATURES AS BIOMARKERS IN PREDICTING TUBERCULOSIS RELAPSE

Bongani Mnyandli[1*], Anandi Bierman[1], Gerhard Walzl[1]

[1]Stellenbosch University, Stellenbosch, South Africa, *nzimandebongani95@gmail.com

Background: Tuberculosis (TB) relapse, defined as the recurrence of active disease after treatment completion, is a major challenge in the treatment of the disease. The lack of reliable biomarkers, along with the sex-dependent biological differences in disease. Highlight the need for the identification of minimal, sex-specific transcriptional signatures for TB relapse to guide treatment monitoring strategies. **Methods:** We analysed gene expression data from the PREDICT-TB cohort, employing a robust three-stage machine learning feature selection pipeline (filtering, feature importance, and wrapper optimization) on the male subset of the dataset. The final minimal gene panel was evaluated using Random Forest (RF) and eXtreme Gradient Boosting (XGBoost) classifiers, with metrics such as using metrics including Area Under the Curve (AUC), Sensitivity, and Specificity. Performance was assessed via repeated k-fold cross-validation and subsequently tested on independent external validation cohorts. **Results:** We successfully identified a 7-gene signature (ENTPD2, KANK1, ANAPC13, PI4K2A, INTS1, ZNF439 and NFATC2IP-AS1) capable of discriminating relapsed from cured patients in the internal cross-validation set. The RF model yielded the highest predictive performance (AUC = 0.9656), validating the concept of a potent sex-specific predictive signature. However, predictive performance declined significantly when tested on external validation cohorts, indicating a failure of generalizability to the broader population.Conclusion: The 7-gene signature serves as a strong mechanistic hypothesis for male-specific relapse susceptibility, identifying biological pathways that may be therapeutically targetable. The failure of the model to generalize highlighted its sensitivity to cohort-specific batch effects. Future work must prioritize cross-cohort data harmonization and the development of a signature for the female population for clinical translation.

CHARACTERISATION OF THE SUBCELLULAR LOCALISATION OF AFRICAN HORSE SICKNESS VIRUS NON-STRUCTURAL PROTEIN NS5

Nhlakanipho Mtambo[1*], Eshchar Mizrachi[1], Vida van Staden[1]

[1]Department of Biochemistry, Genetics and Microbiology, *nipho.mtambo@up.ac.za

African horse sickness virus (AHSV) is an economically important orbivirus affecting equids throughout sub-Saharan Africa. The virus has a segmented double-stranded RNA (dsRNA) genome. Genome segment 10 (Seg-10) encodes the well-characterised NS3 protein, involved in viral egress. It also contains an alternative +1 open reading frame (ORF) predicted to encode the non-structural protein NS5. Despite preliminary evidence of NS5 expression and a proposed role in host-virus interactions, its intracellular distribution remains poorly characterised. This study aimed to investigate the subcellular localisation and targeting motifs of AHSV NS5. To achieve this, AHSV Seg-10 nucleotide sequences were retrieved from NCBI GenBank and analysed, yielding a dataset of 105 unique NS5 ORFs. In silico bioinformatic analysis predicted conserved N-terminal features associated with mitochondrial targeting, alongside a putative nuclear localisation signal. Further, mutagenesis analysis demonstrated that N-terminal truncation or the substitution of arginine residues at positions 21 and 22 reduced predicted mitochondrial targeting scores across all sequences. To experimentally investigate this, recombinant baculovirus constructs encoding wild-type and mutant NS5 fused to a Strep-tag, a Twin-Strep-tag, or eGFP were expressed in Sf9 insect cells and subsequently visualised. Confocal microscopy revealed distinct subcellular distributions: Strep-tagged and Twin-Strep-tagged NS5 displayed predominant nuclear localisation, while eGFP fusions exhibited mixed nuclear and partial mitochondrial localisation. These findings align with the bioinformatic predictions for dual organelle localisation and suggest the potential involvement of N-terminal arginine residues in mitochondrial targeting. Overall, these results provide a foundation for future functional characterisation of NS5 in the context of AHSV infection. Understanding these host-virus interactions may ultimately aid in developing novel antiviral strategies.

GENETIC DIVERSITY OF EHRLICHIA RUMINANTIUM BASED ON MULTILOCUS SEQUENCE TYPING OF ISOLATES FROM INDIGENOUS GOATS AND AMBLYOMMA HEBRAEUM TICKS

Xolile Nuse[1, 2*], H.C. Steyn[4], M.A. Van Der Nest[3], K. Hadebe[2], F.C. Muchadeyi[2], E.F. Dzomba[1]

[1]Discipline of Genetics, School of Life Sciences, University of KwaZulu-Natal, South Africa. [2]Agricultural Research Council - Biotechnology Platform, Onderstepoort, Pretoria, South Africa. [3]Department of Biochemistry, Genetics and Microbiology, University of Pretoria, Pretoria, South Africa. [4]Department of Molecular Biology, Agricultural Research Council - OVR, Onderstepoort 0110, South Africa. *xolilenuse@gmail.com

Heartwater, an infectious tick-borne disease caused by the intracellular bacterium Ehrlichia ruminantium and transmitted by Amblyomma ticks, remains a major constraint on ruminant production in sub-Saharan Africa due to high mortality. The main challenge in controlling the disease is the lack of a broadly effective vaccine, which is thought to be related to the high genetic variation among pathogen strains circulating in the same area. This study investigated the genetic diversity and phylogenetic structure of E. ruminantium isolates circulating in South African indigenous goats and A. hebraeum ticks using multilocus sequence typing (MLST). A total of 381 goat blood samples and 70 tick samples were screened using pCS20 quantitative real-time PCR, yielding 56 (14.7%) and 31 (44.3%) positive samples, respectively. Positive samples were characterized by sequencing four loci (gltA, groEL, sucA2, and pCS20), followed by locus-specific and concatenated phylogenetic analyses. MLST identified 11 unique sequence types among 44 sequences and a haplotype diversity of 0.77, demonstrating considerable genetic heterogeneity within the pathogen population. Phylogenetic analyses revealed multiple co-circulating lineages, including clusters related to reference strains such as Gardel, Kümm2, and Grootvallei, alongside divergent genotypes suggestive of locally evolving variants. Population genetic analyses indicated weak host-associated structuring between goat- and tick-derived isolates (FST = 0.0247; p > 0.05), while AMOVA showed that 97% of genetic variation occurred within host groups. Haplotype network and multivariate analyses further supported a shared goat-tick transmission system, with no evidence of strict host specificity. Selection analyses suggested that adaptive evolutionary processes may contribute to the observed diversity. These findings demonstrate that E. ruminantium populations in South Africa are highly diverse and weakly structured by host, providing important insights into pathogen evolution, transmission dynamics, and future heartwater vaccine development.

Keywords: E. ruminantium, heartwater, multilocus sequence typing, genetic diversity, phylogenetic, indigenous goats, A. hebraeum

GRAPH NEURAL NETWORK-BASED DISCOVERY OF MIRNA-MRNA REGULATORY INTERACTIONS IN BREAST CANCER

Nothando Gama[1*], Hocine Bendou[1]

[1]South African National Bioinformatics Institute (SANBI), SA Medical Research Council Bioinformatics Unit, University of the Western Cape, Private Bag X17, Bellville, Cape Town 7535, South Africa. *nothandograce82@gmail.com

Breast cancer remains one of the leading causes of cancer-related mortality among women worldwide, with dysregulation of microRNA (miRNA)-mediated gene regulation contributing to tumor initiation, progression, and metastasis. MicroRNAs are small non-coding RNAs that regulate gene expression through post-transcriptional repression of target messenger RNAs (mRNAs), making the identification of miRNA-mRNA interactions critical for understanding disease mechanisms and discovering therapeutic targets. Despite advances in computational prediction, many biologically relevant miRNA-mRNA interactions remain unidentified. Heterogeneous graph-based deep learning provides an opportunity to model complex regulatory networks and identify novel interactions. In this study, we developed a Heterogeneous Graph Transformer (HGT)-based link prediction framework to discover miRNA-mRNA regulatory interactions associated with breast cancer. Experimentally validated interactions from miRTarBase were integrated with transcriptomic data from TCGA breast invasive carcinoma (TCGA-BRCA). To enhance network coverage, ovarian cancer data (TCGA-OV) were incorporated, leveraging shared regulatory mechanisms between the two cancers. A heterogeneous graph comprising miRNA and mRNA nodes was constructed using mean expression levels and node degree as node features. The model was trained using a threshold-optimized negative sampling strategy and evaluated through link prediction. The HGT framework achieved an accuracy of 82%, AUC-ROC of 90%, AUC-PRC of 80%, F1-score of 75%, and Matthews Correlation Coefficient (MCC) of 62%. High-confidence novel predictions included hsa-miR-4775 targeting SMAD4, ABCA1, and ZMAT3, and hsa-miR-520d-5p targeting BMPR2. These targets are implicated in TGF-β signalling, epithelial-mesenchymal transition, tumour suppression, and p53-mediated apoptosis implicated in breast cancer progression. Overall, this study demonstrates the potential of graph-based deep learning to accelerate the discovery of miRNA-mediated regulatory mechanisms and prioritize candidate interactions for experimental validation in breast cancer.

RECONSTRUCTING CERVICAL CANCER TUMOUR MICROENVIRONMENT STATES FROM DNA METHYLATION PROFILES USING A MACHINE LEARNING FRAMEWORK

Saltiel Hamese*, Mutsa Takundwa[1], Earl Prinsloo[2], Deepak Balaji Thimiri Govinda Raj[1]

[1]Synthetic Nanobiotechnology and Biomachines, Synthetic Biology and Precision Medicine Centre, CSIR Pretoria, South Africa. [2]Biotechnology Innovation Centre, PO Box 94, Rhodes University, Makhanda, 6140 South Africa. *hamesesaltiel@gmail.com

Background: Cervical cancer remains a major cause of cancer-related morbidity and mortality globally, particularly in low- and middle-income countries where the burden is exacerbated by persistent human papillomavirus (HPV) infection and high rates of HIV co-infection. Despite the availability of large-scale genomic, epigenomic, transcriptomic, and proteomic datasets, translating these data into biologically meaningful tumour classifications and clinically actionable insights remains challenging. A major limitation is the lack of computational frameworks capable of reconstructing tumour microenvironment (TME) states from bulk patient datasets and systematically linking these states to representative experimental model systems. **Methods:** We developed a DNA methylation-based machine learning framework for tumour microenvironment state inference and molecular stratification in cervical cancer. The framework integrates epigenomic feature selection, methylation-based deconvolution, unsupervised clustering, and predictive modelling to reconstruct biologically meaningful tumour states directly from patient DNA methylation profiles. By incorporating signals derived from malignant epithelial cells, immune populations, stromal compartments, and host-associated biological processes, the approach enables comprehensive characterisation of cervical tumour ecosystems. **Results:** Application of the framework to cervical cancer cohorts identified four distinct tumour microenvironment states characterised by differences in immune activity, stromal composition, epithelial differentiation, and proliferative signalling. These reconstructed states revealed substantial inter-tumour heterogeneity and provided insight into the biological consequences of HPV-driven carcinogenesis and HIV-associated immune modulation. Notably, HIV-associated molecular perturbations were largely restricted to immune-inflamed tumours, indicating that HIV influences specific biological contexts rather than defining a universal cervical cancer subtype. The inferred tumour states also enabled systematic alignment of patient tumours with cervical cancer cell line models, facilitating identification of experimental systems that more accurately recapitulate patient-derived phenotypes. **Conclusions:** This framework provides a reproducible approach for methylation-driven tumour microenvironment reconstruction and molecular stratification in cervical cancer. By linking patient tumour states to experimentally tractable models, it supports biomarker discovery, disease-risk stratification, therapeutic vulnerability assessment, and precision oncology research. More broadly, the framework demonstrates how machine learning applied to epigenomic data can bridge the gap between patient biology and preclinical experimentation, supporting future advances in cervical cancer prevention, screening, and treatment.

SURVEILLANCE AND MONITORING OF ANTIMICROBIAL-RESISTANT HUMAN PATHOGENIC AND ZOONOTIC BACTERIA IN THE NGAMILAND AND OKAVANGO DELTA

Patience Motshosi[1*] , Kabo Baruti [2]

[1]Botswana Harvard Health Partnership, Gaborone, Botswana. [2]Okavango Research Institute, University of Botswana, Maun, Botswana. *patiencemotshosi@gmail.com

Antimicrobial resistance (AMR) represents a critical global health threat, with projections of 10 million annual deaths by 2050, disproportionately affecting low- and middle-income countries (LMICs). Freshwater environments at high human-animal-wildlife interfaces, such as the Okavango Delta and Ngamiland, serve as key reservoirs and dissemination pathways for pathogenic bacteria and AMR genes. This environmental surveillance study collected surface water samples from eight sites: three along the Thamalakane River in Maun and four semi-permanent pools within Moremi Game Reserve. Physicochemical parameters were recorded at river sites only. Samples were concentrated by centrifugation, and total nucleic acids were extracted and sequenced using the Illumina NextSeq2000 platform. Taxonomic classification employed Kraken2, while AMRFinder identified antibiotic resistance genes (ARGs). Statistical analyses included independent t-tests, one-way ANOVA, Shannon diversity index, and PERMANOVA for bacterial alpha diversity comparisons. Seasonal variation significantly influenced salinity and electrical conductivity ($p = 0.0007$ and $p = 0.0001$, respectively), while temperature, pH, and culturable bacterial counts remained unaffected ($p > 0.05$). No significant inter-site differences were detected across any parameter. Metagenomic sequencing identified Acinetobacter lwoffii and Acinetobacter johnsonii (~1–3%) as the most abundant pathogens. Clinically significant zoonotic pathogens, including Klebsiella pneumoniae, Acinetobacter baumannii, Stenotrophomonas maltophilia, and Proteus mirabilis, were detected at low abundance. The resistome was dominated by efflux pump genes MexF (18.3%) and MexB (8.3%), alongside the clinically relevant β-lactamase TEM-116 (17.4%). Hydrological variability primarily influenced ionic composition, while bacterial load remained consistent across sites and seasons. The co-detection of TEM-116 and transposase gene TniB raises concern for horizontal gene transfer (HGT) between environmental and clinical bacteria. Even low-abundance detection of A. baumannii, a known HGT facilitator, is clinically significant. These findings underscore the value of freshwater metagenomic surveillance as an early warning system for AMR in Botswana and comparable LMICs.

TEMPORAL DYNAMICS OF THE GUT MICROBIOME DURING SMOLTIFICATION IN ATLANTIC SALMON (SALMO SALAR)

Badrodin M[1], Difford G[2] 1,Rhode C[1].

[1]Department of Genetics, Stellenbosch University, Stellenbosch. [2]Institute of Animal and Aquaculture Sciences, Norwegian University of Life Sciences, Ås, Norway. *23563036@sun.ac.za

Smoltification is a critical developmental transition in Atlantic salmon (Salmo salar) that enables seawater adaptation. Although the gut microbiome is increasingly recognized as important in host development and metabolism, its temporal dynamics during smoltification remain poorly understood. This study aimed to investigate temporal changes in the gut microbiome of Atlantic salmon (n = 162) across the smoltification period. Gut microbial communities were characterized through 16S rRNA gene sequencing followed by DADA2 preprocessing and Phyloseq-based downstream analyses to assess community composition, diversity and temporal variation. Temporal restructuring of the gut microbiome was observed. Alpha diversity varied significantly among sampling weeks with Chao1 richness ranging from 80-200 taxa (Kruskal–Wallis $\chi^2 = 43.95$, df = 7, $p = 2.19 \times 10^{-7}$), Simpson diversity exceeding 0.75 ($\chi^2 = 23.75$, df = 7, $p = 0.00126$) and Shannon diversity ranging from 2-5 ($\chi^2 = 26.954$, df = 7, $p = 0.00034$). Beta diversity demonstrated significant temporal differentiation in community composition for Bray–Curtis (PERMANOVA: $R^2 = 0.073$, F = 12.615, $p = 0.001$), suggesting ecological succession rather than a uniform increase in microbial diversity. The microbiome included groups such as Bacillota and Pseudomonadota, which may serve as candidate microbial indicators of seawater adaptation. These findings indicate that gut microbiome composition is a complex quantitative trait shaped by host developmental state and environmental conditions. Improved understanding of microbiome dynamics during smoltification may contribute to strategies to optimize salmon health and welfare at sea.

ADVANCING AVIAN GENOMICS: A DE NOVO ASSEMBLY PIPELINE TO GENERATE A HIGH-QUALITY REFERENCE GENOME FOR THE SOUTHERN BALD IBIS (GERONTICUS CALVUS)

Ashley Mortimer[*1], Anri van Wyk[1], Arrie Klopper[1], Paulette Bloomer[1]

[1]Molecular Ecology and Evolution Programme, Department of Biochemistry, Genetics and Microbiology, University of Pretoria. *u21536794@tuks.co.za

The Southern Bald Ibis (Geronticus calvus) is a near threatened southern African avian species of special interest, displaying behaviour vastly differing from other ibis species; apart from its closest relative, the Northern Bald Ibis (Geronticus eremita). Due to their unique preference for both high rainfall and altitude, their distribution is restricted to parts of Lesotho, eSwatini and South Africa. Genetically the species has received minimal attention, with a single mitochondrial DNA study comparing the congeners, of which neither has an available reference genome. A reference genome will be the first step in aiding the species' conservation, acting as a guide for future research, including population genetics/genomics, landscape genomics, and providing a baseline for investigating potential adaptation within the species due to their unique habitat preferences. Oxford Nanopore Technology (ONT) long-read sequencing was performed to generate a high accuracy reference genome. The following de novo assembly pipeline was utilised: trimming, filtering, assembly, raw read mapping, polishing, haplotig purging and scaffolding; with quality checks in-between. A final genome assembly of 1.218 Gb was produced, consisting of 464 contigs, with an N50 value of 25.9698 Mb, only containing 20 ambiguous bases and a 99.4% BUSCO completeness score. Based on available avian reference genomes, the genome produced exhibits higher BUSCO completeness scores, lower scaffold N50 values and the expected 1.2 Gb genome size of avian species. The assembly pipeline generated a reference-level genome for the species to be utilized in future population genetics and genomics studies, allowing for informed conservation efforts.

UNRAVELLING THE ELUSIVE MATING STRATEGY OF DEMATOPHORA NECATRIX: A COMBINED GENOMIC AND PHYSIOLOGICAL APPROACH

Annabel Norval[1,2*], Markus Wilken[1], Noëlani van den Berg [1,2]

[1]Forestry and Agricultural Biotechnology Institute (FABI), Department of Biochemistry, Genetics and Microbiology, University of Pretoria. [2]The Avocado Research Programme, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria. *annabel.norval@up.ac.za

Sexual reproduction in pathogenic ascomycete fungi generates genetic variation, accelerating natural selection and contributing to fungicide resistance and the ability to overcome host resistance. This process is governed by the mating-type (MAT1) locus, comprising two idiomorphs (MAT1-1 and MAT1-2), each encoding transcription factors regulating downstream sexual development genes. The genes within each idiomorph vary considerably, forming the genetic basis distinguishing heterothallic from homothallic fungi. Mating-type is conventionally determined through in silico BLAST-based searches or in vitro crossing experiments. For the avocado pathogen, Dematophora necatrix, both approaches have been challenging. Canonical MAT1 genes appear entirely absent across sequenced Xylariales genomes, and no crossing experiments have, to date, characterised mating strategy of D. necatrix. This study employed a dual genomic-physiological approach to elucidate this mating strategy. Full-genome annotation of Xylariales species using the funannotate pipeline, incorporating RNA-seq data, enabled comparative identification of candidate mating-type genes through non-conventional approaches. Candidate proteins were subjected to HMMER searches against a custom mating-type protein database, InterProScan domain analysis, maximum likelihood phylogenetic reconstruction, and three-dimensional structural prediction via AlphaFold3, with structural comparisons quantified using RMSD calculations in PyMOL — a method not previously applied to mating-type gene identification. Concurrently, controlled crossing experiments between eight single-ascospore isolates, derived from field-collected sexual structures and subjected to Illumina whole-genome sequencing, assessed mating-type segregation physiologically. Together, these approaches yielded candidate mating-type genes via a novel methodology, provided genome-wide annotations for the Xylariales, and generated foundational insights into the mating strategy of D. necatrix, aiding downstream control strategies of this devastating pathogen.

TAXONOMIC BINNING OF LONG-READ DNA SEQUENCES BY K-MER COMPOSITION USING DECISION TREE CLASSIFIERS

Bruhan Kyomuhendo[1]*, Oleg Reva[1], Rian Pierneef[1]

[1]1Centre for Bioinformatics and Computational Biology; Department of Biochemistry, Genetics and Microbiology, University of Pretoria

*bruhankyomuhendo89@gmail.com

Third-generation sequencing technologies have revolutionized metagenomics by producing long reads exceeding 10 kb, yet traditional taxonomic binning tools designed for short reads often struggle with the increased error rates and computational demands of this data. To address these challenges, this study developed a machine learning-based pipeline for the taxonomic binning of long-read DNA sequences, utilizing k-mer composition analysis paired with decision tree and neural network classifiers. Leveraging a dataset of 63,425 bacterial genomes from NCBI's RefSeq, we extracted k-mers (sizes 4–7) and generated distance matrices by normalizing frequencies. K-mers were categorized by abundance and binary-encoded, allowing for efficient Hamming distance calculations via bitwise XOR operations, significantly reducing computational overhead. We then employed spectral clustering to organize genomes into hierarchical nodes and selected informative features using chi-square statistics with False Discovery Rate (FDR) correction. At each node, we trained a hybrid model, multilayer perceptron neural networks were utilized for clusters exceeding 30 genomes, while decision trees served as robust backup classifiers for smaller nodes. Validation using synthetic reads (5–20 kb) demonstrated high performance, with the majority achieving species-level classification and identity scores consistently exceeding our 0.8 specificity threshold. While the pipeline proved robust across diverse taxa, some difficulty remained in distinguishing closely related strains. Our results demonstrate that a hybrid k-mer-based approach offers a scalable and effective solution for species-level identification within complex microbial communities. This pipeline has significant implications for pathogen surveillance and environmental metagenomics. Future research will focus on validating the system with real-world metagenomic datasets and optimizing classification thresholds to further enhance accuracy and broader database integration.

DOES MULTI-TASK SUPERVISION TEACH A CLASSIFIER WHERE TO LOOK? A QUANTITATIVE ATTENTION-LOCALIZATION STUDY IN BRAIN TUMOR MRI

Tisetso Letuka* , Amukelani Mtshabi[1]

[1]Ezintsha, Wits Health Consortium, Faculty of Health Sciences, University of the Witwatersrand. *tisetsoletuka@gmail.com

Background: Explainability methods, such as Grad-CAM, are widely trusted to reveal where an artificial intelligence model focuses when making a medical diagnosis. Multi-task learning is frequently proposed as a strategy to encourage models to attend to clinically relevant regions. However, whether explicit spatial training in a dedicated task head actually improves the internal attention of a shared classification head remains largely unproven in medical imaging. This study quantitatively evaluates whether a lesion-localization head aligns a model's diagnostic attention maps with true, radiologist-annotated tumor centers across 30 pathology classes. **Methodology:** We evaluated a multi-task deep learning architecture (EfficientNet-B4 backbone, trained from scratch) trained on identical brain MRI data to simultaneously classify pathologies and predict (x,y) lesion coordinates. Using a 1,128-image validation set encompassing 30 distinct tumor types and normal controls, we generated post-hoc Grad-CAM++ attention heatmaps for correctly classified abnormal images. We calculated the Euclidean distance between the heatmap's weighted centroid and the true radiologist-annotated tumor locations, comparing results between a multi-task model (loss weight beta=0.5) and a single-task classifier (beta=0.0). **Results:** Both architectures achieved high diagnostic performance (multitask macro-F1: 0.91; single-task: 0.91). The multi-task model's dedicated localization head demonstrated strong spatial accuracy, achieving a median error of 30.7 pixels, with 85.5% of predictions falling within 64 pixels of the true lesion center. However, despite this internal accuracy, its classification attention maps remained highly misaligned, showing a median error of 227.4 pixels (only 2.0% within 64 pixels). Crucially, the single-task model showed a statistically indistinguishable attention profile (226.8 pixel median error; 1.9% within 64 pixels), despite receiving zero spatial training. Conclusion: Explicitly teaching a model to locate a lesion yields highly accurate coordinate predictions but does not meaningfully improve the spatial alignment of its classification attention, suggesting that diagnostic decisions may rely on features that are partially decoupled from localized tumor boundaries. This has direct implications for clinical trust: post-hoc attention maps cannot be blindly relied upon for diagnostic justification. Instead, clinical decision support tools should prioritize explicitly supervised, dedicated localization outputs over passive explainability heatmaps.

THE ROLE OF THE GUT MICROBIOME IN INFLAMMATION AND MULTIMORBIDITY IN AFRICAN POPULATIONS: EVIDENCE FROM AWI-GEN

Aaron Y. Berkman[1,2], Luicer A. Olubayo[1,2], Scott Hazelhurst[1,3]

[1]Sydney Brenner Institute for Molecular Bioscience, Faculty of Health Sciences, University of the Witwatersrand. [2]Division of Human Genetics, School of Pathology, Faculty of Health Sciences, University of the Witwatersrand. [3]School of Electrical and Information Engineering, Faculty of Engineering, University of the Witwatersrand. *1831659@students.wits.ac.za

The gut microbiome is a central regulator of systemic inflammation, a pathway that links diet and environment to multimorbidity — the accumulation of chronic disease. This project aims to determine how gut-microbial function relates to inflammation and multimorbidity in African populations and to resolve the strain-level genomic basis of any associations. C-reactive protein (CRP) offers an interpretable marker of the inflammatory load that drives it. Multimorbidity is an escalating burden across sub-Saharan Africa, yet the evidence base remains overwhelmingly Western. African populations are critically under-represented despite their distinct diets, infectious exposures, and microbial ecology, and the reference databases on which standard functional profilers depend largely exclude novel African gut taxa. Quantifying these organisms, therefore, calls for de novo assembly rather than reference-based methods.The project is nested within the AWI-Gen 2 Microbiome Project, the largest African population-representative metagenomic resource to date (1,801 participants from Burkina Faso, Ghana, Kenya, and South Africa); it pairs deep shotgun (n=1,801) and long-read (n=209) metagenomic sequencing with serum CRP and harmonised multimorbidity data, offering a rare setting in which these relationships can be properly resolved.To date, metagenome-assembled genomes have been constructed de novo and their gene-level functional contributions quantified, recovering taxa that reference methods miss. A pre-specified panel of microbial functions, spanning short-chain fatty acid, bile-acid, and tryptophan–indole pathways, has been tested for association with CRP and multimorbidity. Building on this, a long-read pangenome workflow will resolve within-species structural variation in prioritised taxa, asking what strain-level genomic differences reveal about host–phenotype associations beyond community-level measures. We will present these reproducible Nextflow workflows together with preliminary findings from their application to the AWI-Gen dataset.

IMPACT OF REFERENCE GENOME, ALIGNMENT STRATEGY, AND QUANTIFICATION METHOD ON LOCUS-SPECIFIC HUMAN ENDOGENOUS RETROVIRUS (HERV) DETECTION DURING MONOCYTE-TO-MACROPHAGE DIFFERENTIATION

Tebogo Phago*, Nikki Gentle[1], Vanessa Meyer[2]

[1]Functional Genomics and Immunogenetics Research Group, School of Molecular and Cell Biology, Faculty of Science, University of the Witwatersrand. [2]ImmunOmics Research Group, School of Molecular and Cell Biology, Faculty of Science, University of the Witwatersrand. *2436993@students.wits.ac.za

Human endogenous retroviruses (HERVs) are ancient proviral sequences that comprise approximately 8% of the human genome and have been increasingly implicated in cellular differentiation, immune regulation, and disease. However, their repetitive and dispersed genomic nature complicates accurate locus-specific expression analysis, and no standardized workflow currently exists for HERV profiling. This study aimed to identify HERV loci associated with monocyte-to-macrophage differentiation while evaluating how genome assembly, alignment strategy, and quantification method influence HERV detection. THP-1 monocytic cells were used for pipeline development, while U937 cells are being analysed as an independent validation model. RNA-sequencing reads were aligned to the GRCh38 and CHM13 reference genome assemblies using HISAT2 and STAR, followed by locus-level HERV quantification using TE-locator and MAJEC, among the few established tools for locus-specific transposable element expression analysis. Differential expression analysis was performed using DESeq2 (FDR < 0.05, log2 fold change ≥ 1). Overall mapping rates were comparable across assemblies and aligners (approximately 89–93%), although HISAT2 consistently produced a higher proportion of uniquely mapped reads and reduced multimapping relative to STAR. Downstream quantification was strongly influenced by alignment strategy, with HISAT2 yielding higher read assignment to annotated features (approximately 25–40 million) compared to STAR (15–25 million; p = 0.0341), indicating substantial pipeline-dependent variation in feature-level quantification. Differential expression results also varied across workflows, with CHM13 consistently identifying more differentially expressed HERV loci than GRCh38, highlighting the impact of reference genome choice on transposable element discovery. Pipeline-specific detection was observed, including over 571 loci uniquely identified by the HISAT2–TE-locator workflow, while a smaller subset of loci (n = 253) was consistently detected across all pipelines. Notably, a differentially expressed HERV located near the IFNB1 promoter was identified, suggesting a potential regulatory relationship with interferon signalling during macrophage differentiation. Overall, these results suggest that genome assembly, alignment strategy, and quantification method substantially influence locus-specific HERV detection and downstream biological interpretation, and findings are still being validated.

INTEGRATED PHENOTYPIC PROFILING AND THERAPEUTIC TARGET PRIORITIZATION OF CLINICAL MULTIDRUG-RESISTANT PSEUDOMONAS AERUGINOSA EXHIBITING BIOFILM-DRIVEN PERSISTENCE

Sinethemba Yakobi*, Nontuthuko Maningi[1]

[1]School of Life Sciences, College of Agriculture, Engineering and Science, University of KwaZulu-Natal. *shyakobi@gmail.com

Multidrug-resistant (MDR) Pseudomonas aeruginosa poses a major clinical threat due to its capacity to form biofilms that promote persistence and therapeutic failure. In this study, we integrated phenotypic profiling of clinical isolates with structure-based computational analysis to identify and prioritize therapeutic targets associated with biofilm-driven persistence. Fifteen clinical isolates were characterized for antimicrobial susceptibility and biofilm-forming capacity, revealing a high prevalence of strong biofilm phenotypes across resistance classes and a significant positive association between resistance burden and biofilm biomass. A candidate persistence-associated target, PA0328 (AaaA), was structurally modelled and subjected to consensus binding pocket identification. Ensemble molecular docking of eight candidate ligands identified pimozide and epigallocatechin gallate (EGCG) as top-ranked compounds, with high reproducibility across replicate runs. Interaction topology analysis revealed distinct stabilization mechanisms, with pimozide exhibiting hydrophobic and π–π stacking–driven binding, and EGCG forming extensive hydrogen-bond networks. Cross-platform docking validation confirmed consistent binding site localization and ligand ranking. ADMET profiling showed contrasting translational profiles; pimozide displayed favourable drug-like properties and bioavailability, supporting its potential for repurposing, whereas EGCG exhibited strong binding stability but limited pharmacokinetic feasibility due to high polarity and poor absorption. This study establishes a scalable, simulation-independent framework for therapeutic target prioritization in MDR P. aeruginosa, integrating phenotypic, structural, and pharmacokinetic dimensions. The findings highlight PA0328 as a promising target and support prioritized repurposing candidate for further investigation.

BENCHMARKING DNBSEQ-T7 SEQUENCING PERFORMANCE USING GENOME IN A BOTTLE REFERENCE GENOMES: IMPACT OF LIBRARY PREPARATION STRATEGIES

Maano Malima[1], Ansia van Coller[1,2], Setshaba Taukobong[1,2], Samira Ghoor[1], Nganea Nangammbi[1], Enrico Roode[1], Martin Naicker[1], Victoria Cole[1], Brigitte Glanzmann[1,3], Craig Kinnear[1,3], Nadia Carstens[1,4]

[1]South African Medical Research Council (SAMRC) Genomics Platform, Cape Town, South Africa. [2]South African Medical Research Council (SAMRC) Biomedical Research and Innovation Platform (BRIP), Cape Town, South Africa. [3]SAMRC Centre for Tuberculosis Research, Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa. [4]Division of Human Genetics, National Health Laboratory Service and School of Pathology, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa. *maano.malima@mrc.ac.za

Background: Advances in sequencing technologies have improved the accuracy, throughput and completeness of human genome characterisation, enabling more reliable detection of genetic variation. Well-characterised reference genomes are essential for benchmarking sequencing platforms and bioinformatics pipelines. **Methods:** We present whole-genome sequencing datasets generated on the MGI DNBSEQ-T7 platform for the Ashkenazi Jewish trio reference samples from the Genome in a Bottle (GIAB) Consortium. Libraries were prepared using three distinct MGI workflows; PCR-free, FastFS DNA, and Universal DNA followed by sequencing to a minimum of 400 million paired-end reads per sample (≥30× mean genome coverage). Raw reads were processed with a standardised GATK bioinformatics pipeline, and sequencing performance was evaluated against the GIAB high-confidence benchmark variant sets. **Results:** All three workflows produced high-quality data with strong overall concordance to the GIAB truth sets. The PCR-free library preparation demonstrated the best indel-calling performance and the lowest Mendelian violation rates across the Ashkenazi trio. Conclusion: This publicly available dataset provides a valuable resource for benchmarking DNBSEQ-T7 sequencing performance and for assessing the influence of different library preparation strategies on whole-genome variant detection. The results support the adoption of the DNBSEQ-T7 platform in high-throughput genomic research and reinforce its utility for producing accurate variant calls in precision-medicine and bioinformatics pipelines relevant to African genomic diversity studies.

Keywords: DNBSEQ-T7; whole-genome sequencing; Genome in a Bottle; library preparation; variant calling; bioinformatics; African genomics

DOCKVINA: AN AUTOMATED MOLECULAR DOCKING PIPELINE FOR HIGH THROUGHPUT VIRTUAL SCREENING

Siphamandla Dlamini[1*], Ruben Cloete[1]

[1]South African National Bioinformatics Institute (SANBI), University of the Western Cape, Bellville, South Africa
*4161376@myuwc.ac.za

Virtual screening is an important stage of modern drug discovery, yet the technical complexity of molecular docking pipelines remains a significant barrier for researchers without advanced bioinformatics expertise. Existing tools such as DockOpt demand extensive computational knowledge, affecting accessibility in resource-limited African countries. We present DockVina, a fully automated virtual screening pipeline developed for deployment on High Performance Computing clusters. It is specifically designed for researchers with basic HPC familiarity but no advanced programming experience. DockVina accepts standard input formats (.pdb, .mol2) and is configured via a single config.txt file, with all dependencies loaded during the runtime. The pipeline leverages AutoDock Vina with OpenMPI parallelisation across multiple nodes for efficient large-scale screening. The automated workflow allows for receptor and ligand preparation via MGLTools, docking protocol validation through RMSD re-docking of a native reference ligand, and subset docking across different exhaustiveness levels using any number of actives and decoys, in a 1:50 ratio, for docking parameter optimization. Docking models are automatically ranked by Boltzmann-Enhanced Discrimination of Receiver Operating Characteristic (BEDROC), Enrichment Factor (EF), and Area Under the Receiver Operating Characteristic Curve (AUC), with the best performing model selected for full library screening. The pipeline produces ranked ligand outputs, ROC curves, and charge distribution quality control figures. The performance of DockVina was benchmarked across 15 of 43 DUDE-Z targets as well as a randomly selected Mtb target with known inhibitors were included. Preliminary results indicated for one of the DUDE-Z targets, Peptide Deformylase (DEF) achieved an AUC score of 0.866, Enrichment factor (EF1%) of 14.28, and BEDROC α160.9 score of 0.351 across all docked ligands. Ongoing work involves completing all 43 targets and systematically benchmarking DockVina against DockOpt across performance, accessibility, and usability dimensions.

THANK YOU TO OUR SPONSORS

DIPLOMICS

CONNECTING RESEARCHERS & WORLD CLASS OMICS FACILITIES

KSA | Decoding South Africa's Biodiversity

ABI African Bioinformatics Institute

MEET THE ORGANISING COMMITTEE

Caivil Ndobela Chairperson
President of SASBi Student Council,
University of Pretoria

Joshua Sampson Chairperson
President of SAGS Student Society,
Stellenbosch University

Odireleng Mosuwe Program, Design & Finance
Secretary and *Treasurer of SASBi Student Council,*
University of Pretoria

Nozipho Magagula Program, Outreach & Finance
Deputy Secretary & *Treasurer of SASBi Student Council,*
University of the Witwatersrand

Thulani Nkosi Finance
Development Officer of SASBi Student Council,
Stellenbosch University

Sandile Ntloko Design & Outreach
Deputy Development Officer of SASBi Student Council,
University of Johannesburg

Yonatan Wolberg Outreach
Media Officer of SASBi Student Council,
University of Witswatersrand

GQEBERHA TRAVEL CITY GUIDE

Welcome to the “Friendly City”

Port Elizabeth, officially renamed Gqerrha, is located on the breathtaking shores of Algoa Bay in the Eastern Cape Province. Known for its rich cultural history, pristine wind-swept coastlines, and exceptional wildlife access, it offers delegates a perfect blend of academic atmosphere and natural beauty.

Local Highlights and Attractions

The Beachfront Promenade:
Take a walk along Hobie Beach, HumeWood Beach or King's Beach right outside the central hub. Ideal for clearing your head between dense scientific tracks.

The Boardwalk Precinct:
Located in Summerstrand, this area features excellent dining, shopping, and entertainment walks surrounding an artificial lake.

Addo Elephant National Park:

Situated just 30 - 40 minutes outside the city, it is the third largest national park in South Africa and offers world-class viewing of the "Big Seven" in a malaria free environment.

Route 67:

A fascinating cultural arts heritage trail consisting of 67 public artworks symbolizing the 67 years Nelson Mandela dedicated to the freedom of South Africa.

Culinary Recommendations

Seafood:

Try the local catch (especially Kingklip or Calamari) at restaurants along the beachfront or near the harbor.

Coffee Shops:

Summerstrand and the nearby Walmer area host superb artisan cafes perfect for morning prep or informal meetings.