University of the Free State – Next Generation Sequencing (UFS-NGS) Unit Data and Bioinformatics Workshop 2023

03-07 July 2023

Introduction

With the emergence of the recent pandemic and epidemic diseases, bioinformatics has become a critical tool for public health response and disease surveillance. It is increasingly important for experimental scientists to gain bioinformatics skills required to analyze the large volumes of data produced by the next generation sequencers in order to inform public health policy and interventions as well as pandemic and epidemic preparedness.

Aim

The workshop is designed to equip the participants with the necessary foundational bioinformatics skills required to analyse NGS data focusing on three aspects: whole genome data, 16S/ITS metagenomics and microbial metagenomics data. The workshop will begin by giving participants a basic understanding of the Linux command line functionality that will be helpful when navigating the various stages of NGS data analysis. The workshop will combine both theoretical and practical sessions.

Intended audience

The course is aimed at individuals from a Molecular Biology background, interested in bioinformatics, and who work (or are planning to work) on genomic datasets.

Participants will be introduced to:

- Linux command lines
- 16S/ITS metagenomics data analysis aspects
- Shotgun metagenomics data analysis
- Whole genome data analysis aspects
- Enteric and respiratory meta transcriptomics and metagenomics analysis aspects
- Computational modeling for new chemical entity characterization

Host

Prof. Martin Nyaga

Venues

- Kine 5
- Computer labs - MODL LAB 3
- Faculty of Health Sciences Foyer

Guest Speakers

Dr. Shane Murray (Genomics Coordinator and Domain Expert, DIPLOMICS, South Africa)

Prof. Jason Mwenda (WHO-AFRO): The roles and responsibilities of a World Health Organization Collaborating Centre (WHO CC) on Vaccine Preventable (VPD) surveillance and pathogen genomics

University of the Free State Guest speakers

- Prof. Joyce Tsoka-Gwegeni, Vice Dean, Faculty of the Health Sciences
- Dr. Cornelius Hagenmeier, Director of the Office for International Affairs
- Dr. Glen Taylor, Senior Director, Research and Internationalisation

Local Organizing Committee and co-facilitators

Coordinator of the sub-committees

- Prof. Martin Nyaga

Planning/content sub-committee

- Dr. Peter Mwangi (UFS-NGS Unit)
- Dr. Jijoho Michel Agbla (UFS-NGS Unit)
- Mr. Milton Mogotsi (UFS-NGS Unit)

Logistics sub-committee

- Dr. Emmanuel Ogunbayo (UFS-NGS Unit)
- Ms. Hlengiwe Sondlane (UFS-NGS Unit)

Marketing sub-committee

- Mr. Teboho Mooko (UFS-NGS Unit)
- Ms. Robyn-Lee Potgieter (UFS-NGS Unit)
- Ms. Nkosazana Shange (UFS-NGS Unit)

ICT sub-committee

- Mr. Stephanus Riekert (ICT, UFS)

Instructors

- Dr. San James (Center for Epidemic Research and Innovation/KwaZulu-Natal Research and Innovation Sequencing Platform (CERI/KRISP), South Africa)
- Dr. Karla Stucker (Waterford School and Salt Lake Veterinary Services, Utah, USA)
- Dr. Celeste Donato (Murdoch Children's Research Institute, Australia)
- Prof. Errol Cason (University of the Free State, South Africa)
- Dr. Benjamin Kumwenda (Kamuzu University of Health Sciences, Malawi)
- Dr. Kwe Yinda (Rocky Mountains Laboratories, National Institute of Health, MT, USA)
- Dr. Stephanie van Wyk (CERI/KRISP), South Africa
- Prof. Saheed Sabiu (Durban University of Technology, South Africa)

Instructor Assistants

- Mr. Armand Bester
- Ms. Precious Letebele

Rapporteur

- Prof. Saheed Sabiu

Funding Organizations

- Distribution Platform in Omics (DIPLOMICS)
- Bill and Melinda Gates Foundation (BMGF)
- World Health Organization (WHO/AFRO)

Media coverage/marketing

- Mr. Paul Harris
- Mr. Louis van Wyk

Participating Organizations

Local (South Africa) organizations

- University of the Free State (Bloemfontein campus), South Africa
- Center for Epidemic Research and Innovation/KwaZulu-Natal Research and Innovation Sequencing Platform (CERI/KRISP), South Africa
- University of Witwatersrand
- University of Pretoria
- University of Cape Town
- University of KwaZulu-Natal
- University of South Africa
- University of Fort Hare
- Sefako Makgatho Health Sciences University
- University of Limpopo
- National Institute of Communicable Diseases
- University of Venda
- North-West University

International organizations

- Cameroon: University of Yaounde, Centre for Research in Infectious Diseases
- Ghana: Noguchi Memorial Institute for Medical Research
- Kenya: The University of Embu
- Lesotho: National University of Lesotho
- Malawi: Public Health Institute of Malawi, Kamuzu University of Health Sciences
- Mali: National Institute for Public Health
- Mozambique: Manhica Health Research Center
- Namibia: Namibia University of Science and Technology
- United Kingdom: University of Liverpool
- Zambia: Churches Health Association of Zambia, Centre for Infectious Disease Research in Zambia
- Zimbabwe: National Biotechnology Authority

Number of participants for practical training (~100)

Programme Schedule and Outline

Day 1: Monday 03 July 2023

Time	Description	Venue	Responsible person(s)
08:00 - 09:00	Registration	Foyer	All facilitators
09:00 - 09:15	Welcome message: Prof. Martin Nyaga	Kine 5	Prof. Martin Nyaga
09:15 - 09:25	Opening of the workshop by Prof. Joyce Tsoka-Gwegeni (Vice Dean (Research) Faculty of Health Sciences		
09.25 - 10:00	*Guest speaker: Prof. Jason Mwenda (WHO AFRO)*		
10:00 - 11:00	Tea/Coffee break and orientation to the venue(s)	Foyer	All facilitators
	Satellite meetings with guests (Dr. Shane Murray and Prof. Jason Mwenda) and UFS management (Prof. Joyce Tsoka-Gwegeni, Dr. Glen Taylor, Dr. Cornelius Hagenmeier and Mr. Tebogo Machethe)	Boardroom	Prof. Martin Nyaga
11:00 - 12:30	Theoretical talk on Introduction to bioinformatics and command lines for NGS data analysis	Kine 5	Dr. San James, Dr Stephanie van Wyk, and Dr. Karla Stucker supported by the rest of the trainers
12:30 - 13:30	Lunch break	Foyer	Kovsie Inn
13:30 - 15:00	Hands-on training on command lines for NGS data analysis	MODL LAB 3	Dr. San James, Dr Stephanie van Wyk and Dr. Karla Stucker supported by the rest of the trainers
15:00 - 16:30	Hands-on training on command lines for NGS data analysis	MODL LAB 3	Dr. San James, Dr Stephanie van Wyk and Dr. Karla Stucker supported by the rest of the trainers
17:00 - 17:10	Welcoming dinner Prof. Martin Nyaga introduces Dr. Cornelius Hagenmeier, Director of the Office for International Affairs	Foyer	Dr. Cornelius Hagenmeier
17:10 - 17:20	Welcoming of international guests by Dr. Cornelius Hagenmeier		
17:20 - 17:30	Dr. Glen Taylor, Senior Director, Research and Internationalisation *Talk on Research, Innovation, Partnerships & Internationalisation agenda of UFS Directorate of Research & Development* Welcome Dr. Shane Murray		
17:30 - 18:00	Dr. Shane Murray (Genomics Coordinator and Domain Expert, DIPLOMICS-SA) *Talk by Shane Murray from DIPLOMICS-SA*		
18:00 - 20:00	Dinner		

Day 1 topics

Introduction to Linux and command lines for NGS data analysis

- Basic commands: ls, cd, pwd, mkdir, rm, cp, mv, cat, less, head, tail
- File manipulation: touch, nano, chmod, chown
- Searching and Retrieving Data: grep, find, curl, wget
- Retrieving Biological Databases: NCBI, UniProt, Ensembl, etc.
- Manipulating Biological Data: FASTA, FASTQ, SAM, BAM
- Basic Bioinformatics Tools: BLAST, Bowtie, BWA, Samtools, bedtools, etc.
- Data Processing and Manipulation with Command Line Tools
- Sequence Analysis with Command Line Tools
- Genome assembly using SPAdes
- Basic Data Visualization in the Command Line Interface
- Command-line plots using R

Day 2: Tuesday 4 July 2023

Theoretical talks/ Practical sessions

Time	Description	Venue	Responsible person(s)
08:30 - 09:30	Theoretical talks on 16S metagenomics aspects	Kine 5	Prof. Errol Cason
09:30 - 10:30	Theoretical talks on whole genome data analysis aspects	Kine 5	Dr. Benjamin Kumwenda
10:30 - 11:00	Tea/Coffee break	Foyer	Kovsie Inn
11:00 - 12:00	Theoretical talks on whole genome data analysis aspects	Kine 5	Dr. Celeste Donato
12:00 - 13:00	Lunch break	Foyer	Kovsie Inn
13:00 - 13:45	Theoretical talks on microbiome aspects	Kine 5	Dr. Kwe Yinda
14:00 - 16:30	Hands on training on command lines continues.	MODL LAB 3	Dr. San James, Dr Stephanie van Wyk, and Dr. Karla Stucker supported by the rest of the trainers

Day 3: Wednesday 5 July 2023

16S metagenomics aspects

Time	Description	Venue	Responsible person(s)
08:30 - 10:30	Hands on training on 16S metagenomic analysis	MOD LAB 3	Prof. Errol Cason supported by the rest of the trainers
10:30 - 11:00	Tea/Coffee break	TBA*	Kovsie Inn
11:00 - 12:30	Hands on training on 16S metagenomic analysis continues	MOD LAB 3	Prof. Errol Cason supported by the rest of the trainers
12:30 - 13:30	Lunch break	TBA*	Kovsie Inn
13:30 - 15:00	Hands on training on 16S metagenomic analysis continues	MOD LAB 3	Prof. Errol Cason supported by the rest of the trainers
15:00 - 16:30	Molecular modeling for new chemical entity bioprospection for rotavirus (theory)	MOD LAB 3	Prof. Saheed Sabiu
16:30 - 19:00	Bloemfontein excursion	Naval Hill Planetarium	Prof. Martin Nyaga

Day 3 topics

- Importing, Quality Control and Pre-processing of Targeted Metagenomics Data in QIIME2: Hands-on exercises in importing different types of data into QIIME2 as well as quality assessment of sequence data.
- Sequence Processing and Denoising: Hands-on exercises in denoising data using DADA2 in QIIME2 as well as evaluating outputs and problem solving.
- Taxonomic Classification in samples: Training of taxonomic classifiers and the corresponding taxonomic classification.
- Exporting Data for downstream analysis: Summary of files needed for downstream diversity and statistical analysis.
- Diversity and Community Analysis: Community analysis using R packages such as Phyloseq and Vegan. Hands-on exercises in alpha and beta diversity analysis.
- Tools and Databases: Overview of commonly used bioinformatics tools and databases for 16S metagenomics analysis and hands-on exercises in accessing and using these resources.

Day 4: Thursday 6 July 2023

Whole genome and phylogenetics aspects

Time	Description	Venue	Responsible person(s)
08:30 - 10:30	Hands on training on whole genome and phylogenetics aspects	MOD LAB 3	Dr. Benjamin Kumwenda and Dr. Celeste Donato supported by the rest of the trainers
10:30 - 11:00	Tea/Coffee break	TBA*	Kovsie Inn
11:00 - 12:30	Hands on training on whole genome and phylogenetics aspects continues.	MOD LAB 3	Dr. Benjamin Kumwenda and Dr. Celeste Donato supported by the rest of the trainers
12:30 - 13:30	Lunch break	TBA*	Kovsie Inn
13:30 - 15:00	Hands on training on whole genome and phylogenetics aspects continues.	MOD LAB 3	Dr. Benjamin Kumwenda and Dr. Celeste Donato supported by the rest of the trainers
15:00 - 16:30	Hands on training on whole genome and phylogenetics aspects continues.	MOD LAB 3	Dr. Benjamin Kumwenda and Dr. Celeste Donato supported by the rest of the trainers
17:00 - 20:00	Closing dinner for the trainers and facilitators with remarks from Dr. Cornelius Hagenmeier (Director of the Office for International Affairs), Prof. Joyce Tsoka-Gwegeni (Vice Dean, Faculty of Health Sciences) and Mr. Tebogo Machethe (Director for Innovation, Contracts, Tech Transfer & IP Division)	Longhorn Steak Ranch	Prof. Martin Nyaga

Day 4 topics

- Quality Control and Pre-processing of Whole-Genome Data: Hands-on exercises in quality control of whole-genome sequencing data using different tools
- Read Mapping: Hands-on exercises in read mapping using tools
- Sequence Alignment for Phylogenetics: Hands-on exercises in sequence alignment
- Model Selection: Overview of model selection for phylogenetic analysis, including the use of likelihood-based methods and model comparison tools such as jModelTest and ModelTest.
- Phylogenetic Tree Construction: Hands-on exercises in phylogenetic tree construction using software such as RAxML and PhyML.
- Tree Visualization and Annotation: Tree visualization and annotation techniques, including the use of software such as FigTree and iTOL.
- Phylogenetic Data Management and Sharing: Data management and sharing strategies, including the use of public databases.

Day 5: Friday 7 July 2023

Viral metagenomics/microbiome aspects

Time	Description	Venue	Responsible person(s)
08:30 - 10:30	Hands on training on viral/microbiome metagenomics aspects.	MOD LAB 3	Prof. Benjamin Kumwenda/Dr. San James supported by the rest of the trainers
10:30 - 11:00	Tea/Coffee break	TBA*	Kovsie Inn
11:00 - 12:30	Hands on training on viral/microbiome metagenomics aspects continues.	MOD LAB 3	Prof. Benjamin Kumwenda/Dr. San James supported by the rest of the trainers
12:30 - 13:30	Lunch break	TBA*	Kovsie Inn
13:30 - 15:00	Hands on training on molecular modeling for new chemical entity bioprospection for rotavirus	MOD LAB 3	Prof. Saheed Sabiu supported by the rest of the trainers
15:00 – Departure time	Enjoy Bloemfontein hospitality		

Day 5 topics

- Quality Control and Pre-processing of Metagenomics Data: Hands-on exercises in quality control of metagenomics sequencing data using different tools
- Read Mapping and Assembly: Hands-on exercises in read mapping and assembly of metagenomics data using different tools.
- Taxonomic Classification: Hands-on exercises in taxonomic classification of metagenomics data using different tools.
- Alpha and beta diversity analysis
- Differential expression
- Functional Annotation: Hands-on exercises in functional annotation of metagenomics data using different tools.
- Tools and Databases: Overview of commonly used bioinformatics tools and databases for metagenomics analysis and hands-on exercises in accessing and using these resources.
- Tools and Databases: Commonly used computational tools and databases for molecular modeling and hands-on exercises in molecular bioprospection for NCE