

# Module **B03**



# Bioinformatics Foundational Course

## Informatics Skills

NGS Academy for the Africa CDC

# Module B03

## Informatics Skills

 [back to the table of modules](#)

**Module last updated:**

December 2024

Suggested or approximate number of sessions	2
Suggested or approximate total learning time	6-8 hours
Target audience	Bioinformaticians and IT personnel
Delivery format	Lectures, practicals
Level of the module	Introductory



### Contributors

George Githinji, Shahiid Kiyaga and Nicola Mulder.



### Suggested prerequisite module(s)

- [Module B01. Introduction to Unix/Linux, Command Line, and Shell Scripting](#)
- [Module B02. Introduction to Version Control](#)



### Module description

The module covers basic topics in computing technology, including the foundations of computer science, memory and data structures, efficiency, software engineering, and computational biology software. It provides an introduction to computing environments, HPC, parallelization, virtualization, cloud computing, and how to choose a computing environment. In this module, participants are introduced to the following topics and/or concepts:

- 
- Terminology such as ram, cpu, hard disk, bandwidth, parallelization, virtualization
  - Configuring your network cards
  - Configuring your software repositories
  - Tools for accessing Linux e.g. Putty, MobaXterm
  - Log file management
  - How to set up secure shell (SSH) access
  - How to use nmap to scan your local server
  - The basics of data encryption
  - Overview of operating system updates
  - The basic components of a data analysis system
  - Common steps of bioinformatics analysis workflows and the use of workflow managers in pipeline development and automation
  - Setting up a scheduler e.g SLURM, SGE
  - Submitting jobs to an HPC
  - Choice of computing hardware: single machines vs HPC vs Cloud
  - Comparing the costs of computing and analysis platforms and the differences between local and cloud infrastructure
  - Procurement, service contracts, terms and conditions for cloud based infrastructure services
  - Troubleshooting and evaluating tools
  - Data security, sensitive data
  - Safeguarding digital information



## Module learning outcomes

---

On completion of this module, the participants will have a basic knowledge of, or will be able to:

- Describe the components of a computing environment (memory, cpu, etc.)
- Establish bioinformatics computing infrastructure, including selecting appropriate operating systems, queuing systems, and resource managers.
- Install and update operating systems
- Install necessary software, including outlining the use of containers and Conda environments.
- Identify the resources required to establish and maintain computing infrastructure, including human resources.
- Set up remote access
- Understand data security, protection, and governance principles.



## Module assessments

---

Module practical: Not applicable

Module quiz: Assessment questions available on the [ASLM platform](#)



## Module resources

---

- [WCS & COG-UK | Slides - Data tools and pipelines](#)
- [H3ABioNet | Tools and Services/Technical Guidelines](#)



## Acknowledgements

---

We would like to thank the following individuals, in alphabetical order of last name, for their valuable time and effort spent in designing (i.e., drafting, reviewing, and refining) this module: **George Githinji, Shahiid Kiyaga and Nicola Mulder.**

Furthermore, we would like to thank the following institutions, societies, journals and individuals from whom we sourced open-access resources, used in this module:

COVID-19 Genomics United Kingdom Consortium, Human Heredity and Health in Africa Bioinformatics Network, Wellcome Connecting Science; Amadou Diallo, Kirsty Lee Garson, Fatma Guerfali, George Githinji, Leigh Jackson, Aquillah Kanzi, Stanford Kwenda, Alice Matimba