Increasing Bioinformatics Uptake in Africa – the H3ABioNet Experience

IUBMB Webinar Series
May 2022

Verena Ras
Training and Outreach Coordinator
Co-chair Training and Education Work Package
Email: verena.ras@uct.ac.za
Website: https://h3abionet.org/
Twitter: @RasVerena
Outline

• What is Bioinformatics?
• H3ABioNet overview
• H3ABioNet training
• Training models
• Other activities related to capacity development
What is Bioinformatics?

“Bioinformatics is an interdisciplinary field that develops and improves on methods for storing, retrieving, organizing and analyzing biological data. A major activity in bioinformatics is to develop software tools to generate useful biological knowledge”

Definition from: https://informatics.sdsu.edu/bioinformatics/

San Diego State University
What is Bioinformatics?

• The whole world is currently generating enormous amounts of biological data – particularly in fields like genomics.

• Interdisciplinary field that develops methods and software tools for understanding biological data - particularly when the data sets are large and complex.

• Combines biology, chemistry, physics, computer science, information engineering, mathematics and statistics to analyze and interpret the biological data.

• Uses computer programming as part of their methodology, as well as specific analysis "pipelines" that are repeatedly used, particularly in the field of genomics.
H3ABioNet

• **Pan African Bioinformatics Network** as part of the Human Hereditary and Health in Africa Consortium (H3Africa). Includes 28 partners in 17 countries, >200 members.

• H3ABioNet was established to develop bioinformatics capacity in Africa and specifically to enable genomics data analysis by H3Africa researchers across the continent. H3ABioNet is developing human capacity through training and support for data analysis, and facilitating access to informatics infrastructure by developing or providing access to pipelines and tools for human, microbiome and pathogen genomic data analysis.

• Major goal is to increase the number of qualified bioinformatics graduates on the continent while creating research opportunities and providing financial support for promising newly-graduated bioinformatics students in Africa, as well as attracting Africans studying abroad back to the continent.

• H3ABioNet delivers high quality training in a variety of formats.

• Increasing presence across Africa and now globally.

[https://h3abionet.org/](https://h3abionet.org/)
H3ABioNet

• Ensuring access to the **computing infrastructure** for moving, storing and analysing data
• Developing **containerized workflows** to make data analysis easy, consistent and reproducible
• Providing **support for data analysis** using existing or developing new algorithms
• **Standardizing and harmonizing** H3Africa data and mapping it to ontologies
• Preparing their data for **submission** to the EGA/ENA
• Ensuring data is **FAIR**
• Making H3Africa data & biospecimens searchable in a **catalogue**
• **Building relevant skills** for data analysis and interpretation

[https://h3abionet.org/](https://h3abionet.org/)  
Slide taken from a presentation by Prof N. Mulder
H3ABioNet

Compute Infrastructure for Genomics
- 3,348 Cores
- 8,872 RAM GB
- 3 Petabytes Storage
- ~40 Million Core Compute hours used

Archive for African Genomics Data
- Assist in preparing data for submission
- Notify intent to submit data: archive@h3abionet.org
- Register on Dashboard
- Transfer encrypted data to EGA
- Passes validation, prepare EGA XML files, re-encrypt and move to cold storage
- Move data to vault, decrypt and validate
- 14 African Genomic Datasets in the archive
- 134.9 Terabytes of African Genomics data
- 77.3 Terabytes transferred for storage at the EGA

https://h3abionet.org/
Images taken from a presentation by Prof N. Mulder
H3ABioNet

- Computing & network infrastructure
- Data management support
- Data analysis tools and workflows
- Data storage, submission and access
H3ABioNet Training

**Face to face Workshops**
- Run >30 face-to-face courses

**Train-the-Trainer**
- NGS & carpentries trainers

**Internships**
- Placed ~20 interns to learn skills

**Live Online/Blended-learning Training**
- Trained >4000 African scientists

**Very impactful**

**Developed workflows, analysed data**

**Hackathons/Data Jamborees**

**Very impactful**
Multiple-Delivery-Mode Training Model

- Local classrooms across Africa
- TA + SA for onsite support
- Managed by training coordinator
- Scalable model
- Online tools – low resource

Image from Aron, 2018

- 2 contact sessions per week (4 hrs)
- Duration of course ~ 3 months

Original model: Gurwitz et al. 2017
Adapted Model for Advanced Training

- **OER**: Tools, datasets, software, dependencies packaged in Singularity containers
- **Face-to-face**: Containers pulled by hosting classrooms
- **Distance Learning**: Classrooms connect virtually in Vula and Adobe Connect

Modified model: Ras et al. 2021

Images from Wikimedia commons
The Development of a Sustainable Bioinformatics Training Environment Within the H3Africa Bioinformatics Network (H3ABioNet)

Shaun Aron1*, Pabalile Abel Chauke2*, Verena Ras2†, Sumir Panj2†, Katherine Johnston2 and Nicola Mulder2 on behalf of the H3ABioNet Training and Education Work Package

1Sydney Brenner Institute for Molecular Bioscience, University of the Witwatersrand, Johannesburg, South Africa.
2Computational Biology Division, Institute of Infectious Disease and Molecular Medicine, CSIR Africa, Wits University, Johannesburg, South Africa

OPEN ACCESS

Bioinformatics training programs have been developed independently around the world based on the perceived needs of the local and global academic communities. The field of bioinformatics is complicated by the need to train audiences from diverse backgrounds in a variety of topics to various levels of competencies. While there have been several attempts to develop standardised approaches to provide bioinformatics training globally, the challenges encountered in resource limited settings hinder the adaptation of these approaches. H3ABioNet, a Pan-African Bioinformatics Network with 27 nodes in 16 African countries, has realised that there is no single simple solution to this challenge and has rather, over the years, evolved and adapted training approaches to create a sustainable training environment, with several components that allow for the successful...

The Numbers…


FAIR and Bioschemas

- Findable
- Accessible
- Interoperable
- Reusable

https://bioschemas.org/
FAIR and Bioschemas

Materials submitted to repository for DOI/PMID

Curate course materials and metadata (pptx, docx, pdf, etc)

Placed onto website (assigning a url) making pages and materials discoverable

Course metadata entered into web template and standardized

Define access – H3ABioNet materials created under cc - license for re-use

Design a template incorporating required fields

Image from: Aron, et al. 2021
Guides and SOP’s
Developing tools and services for H3Africa and the broader bioinformatics community.

H3ABioNet Tools and Services

H3ABioNet has developed a number of tools and services for H3Africa and the broader bioinformatics community. These tools and services have been categorized into Bioinformatics Workflows, Standard Operating Procedures (SOP), Technical Guidelines, Computing Infrastructure and Node Accreditation. To read more about these tools and services, follow the links below.

- Bioinformatics Tools
- Bioinformatics Workflows
- Technical Guidelines
- Computing Infrastructure
- Node Accreditation

https://h3abionet.org/
Developing tools and services for H3Africa and the broader bioinformatics community.

H3ABioNet Tools and Services

H3ABioNet has developed a number of tools and services for H3Africa and the broader bioinformatics community. These tools and services have been categorized into Bioinformatics Workflows, Standard Operating Procedures (SOP), Technical Guidelines, Computing Infrastructure and Node Accreditation. To read more about these tools and services, follow the links below.

- Bioinformatics Tools
- Bioinformatics Workflows
- Technical Guidelines
- Computing Infrastructure
- Node Accreditation
Where to Find More Information?
Where to Find More Information?

Tools & Services
- Bioinformatics Tools
- Bioinformatic Workflows
- Technical Guidelines
- Computing Infrastructure
- Node Accreditation

Technical Guidelines

The system administrator taskforce has developed a series of how-to documents to assist H3ABioNet system administrators with installing, configuring and securing, monitoring and managing their server hardware.

The how-to documentation has been developed to be usable by both non-IT individuals and those just new to Linux by providing detailed text-based step by step instructions supported by images where possible. These how-to's have been divided into three levels: level 1, level 2 and level 3.

- Level 1 How to install a Linux Operating System
- Level 2 How to configure a Linux Operating System
- Level 3 How to monitor and manage a Linux Operating System
Where to Find More Information?

Upcoming and Current Training

H3ABioNet organises a variety of high quality courses and training events covering various aspects of bioinformatics from general introductory topics to more specialized ones such as Next Generation Sequencing and Genome Wide Association Studies analyses. Training proposals on different topics that are pertinent to human health, have synergies with the H3Africa projects and are also in line with H3ABioNet’s vision of developing bioinformatics capacity within Africa, may be submitted here. Please also take note of the H3ABioNet training event policy available here.

Introduction to Bioinformatics Training 2022

The course aims to provide an introduction to the field of bioinformatics, with a focus on important bioinformatics tools, and resources. The course aims to use a combination of theoretical and practical sessions in order for participants to gain practical experience
Other Activities

- Train-the-trainer programme – developed trainer portal
- TtT developed a data extraction tool that can be used on publications
- HtrainDB – trainer database portal
- E-genomics catalogue
- Carpentries
- Education summits
- Training to implement

**NB** – we have 7 work packages, training is just one of them
Acknowledgements

- H3ABioNet - Funding: NIH Common Fund, NGHRI grant: U41HG006941, U24HG006941

- Training core team: (current) Sumir Panji, Shaun Aron, Nicola Mulder, Suresh Maslamoney and Gerrit Botha; (past) Paballo Chauke, Kim Gurwitz

- Local classroom staff across Africa: Too many to mention!

- H3ABioNet Training and Education Work Package members

- All course trainers and organisers