

# 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](mailto:verena.ras@uct.ac.za)

Website: <https://h3abionet.org/>

Twitter: @RasVerena



**H3ABioNet**

Pan African Bioinformatics Network for H3Africa

# Outline

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



**H3ABioNet**

Pan African Bioinformatics Network for H3Africa

# 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*



**H3ABioNet**

Pan African Bioinformatics Network for H3Africa

# 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.

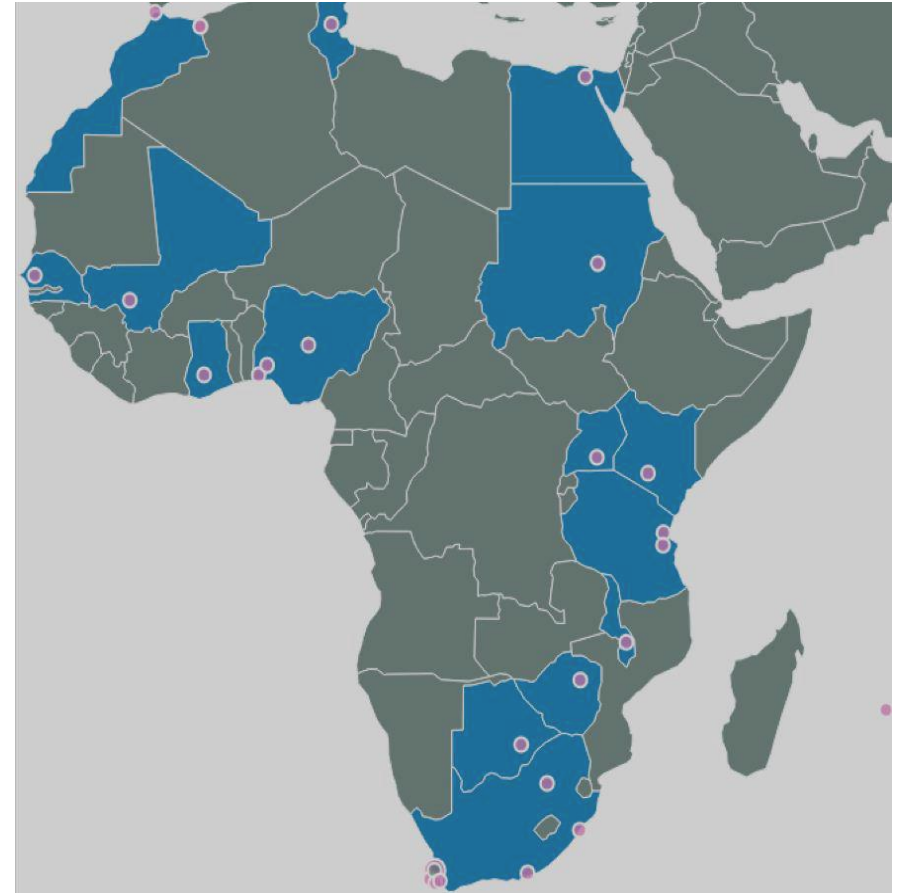


**H3ABioNet**

Pan African Bioinformatics Network for H3Africa

# 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/>



**H3ABioNet**

Pan African Bioinformatics Network for H3Africa

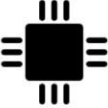

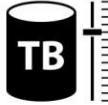

# H3ABioNet

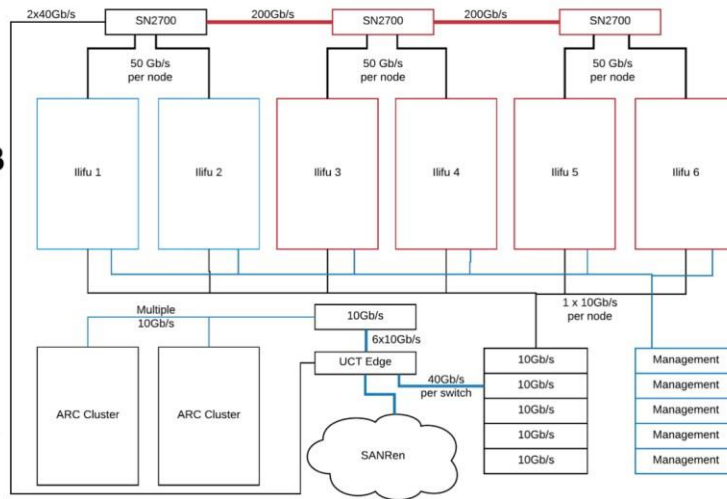
- Ensuring access to the **computing infrastructure** for moving, storing and analysing data
- Developing **containerized workflows** to make data analysis easy, consistent and reproducible
- Providing **support for data analysis** using existing or developing new algorithms
- **Standardizing and harmonizing** 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



# H3ABioNet

## Compute Infrastructure for Genomics

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



The screenshot shows the H3Africa Data & Biospecimen Catalogue interface. It includes a search bar, navigation tabs (Catalogue, About, Contact), and a list of available studies. The 'Available Studies' section displays details for several studies:




Study Name	Acronym	Design	Specimens with aliquots	Participants with data
Collaborative African Genomics Network	CaGEN	Case-Control	85	69
The Genomics of Schizophrenia in the South African Xhosa Population	SAX	Case-Control	4	4
H3Africa Diabetes Study	H3AD	Case-Control	0	0
Stroke Investigative Research & Educational Network	SIREN	Case-Control	17	17
An integrated approach to the identification of genetic determinants of susceptibility to trypanosomiasis	TrypanoGEN	GWAS	0	0

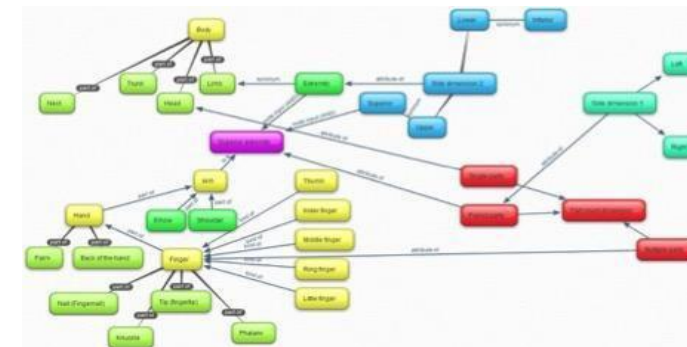
### Archive for African Genomics Data

Assist in preparing data for submission

The flowchart details the submission process:
 

- Notify intent to submit data: [archive@h3abionet.org](mailto: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
- Transfer encrypted data to the landing area

-  **14 African Genomic Datasets in the archive**
-  **134.9 Terabytes of African Genomics data**
-  **77.3 Terabytes transferred for storage at the EGA**



# H3ABioNet

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

Training



**H3ABioNet**

Pan African Bioinformatics Network for H3Africa



# 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

## Hackathons/Data Jamborees

Developed workflows, analysed data

## Live Online/Blended-learning Training

Trained >4000 African scientists

← Very impactful



**H3ABioNet**

Pan African Bioinformatics Network for H3Africa

# Multiple-Delivery-Mode Training Model

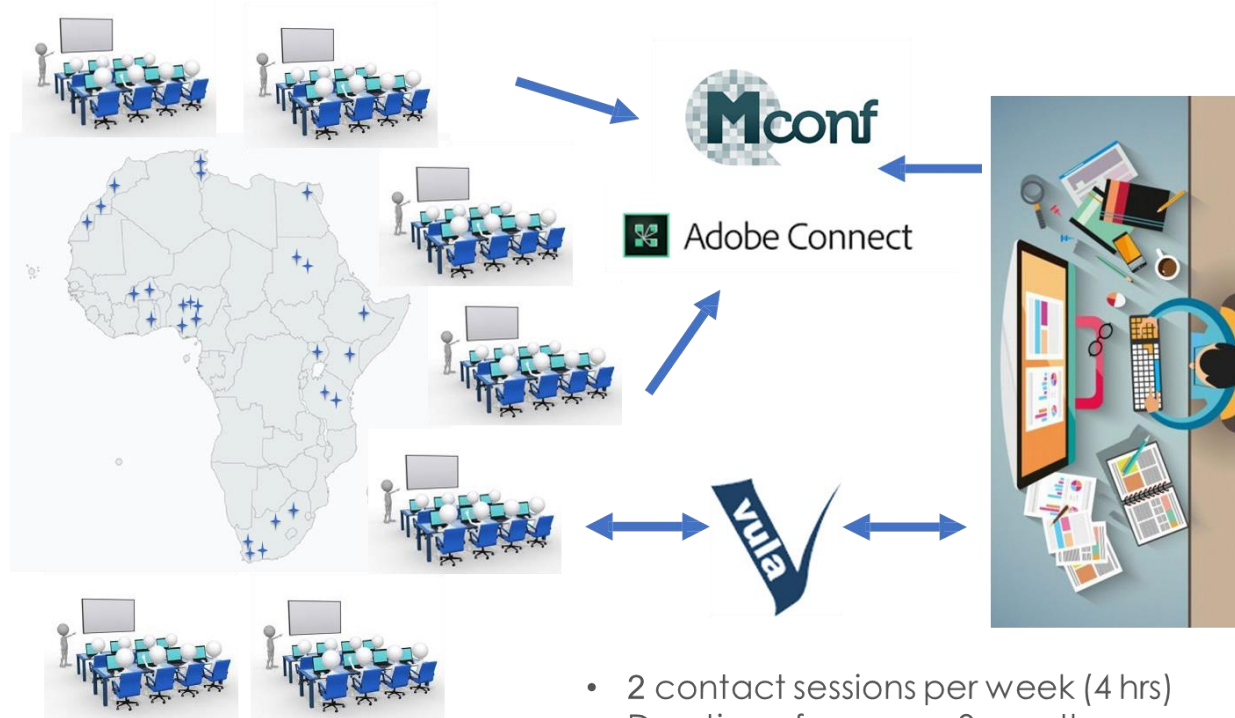


Image from Aron, 2018

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

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

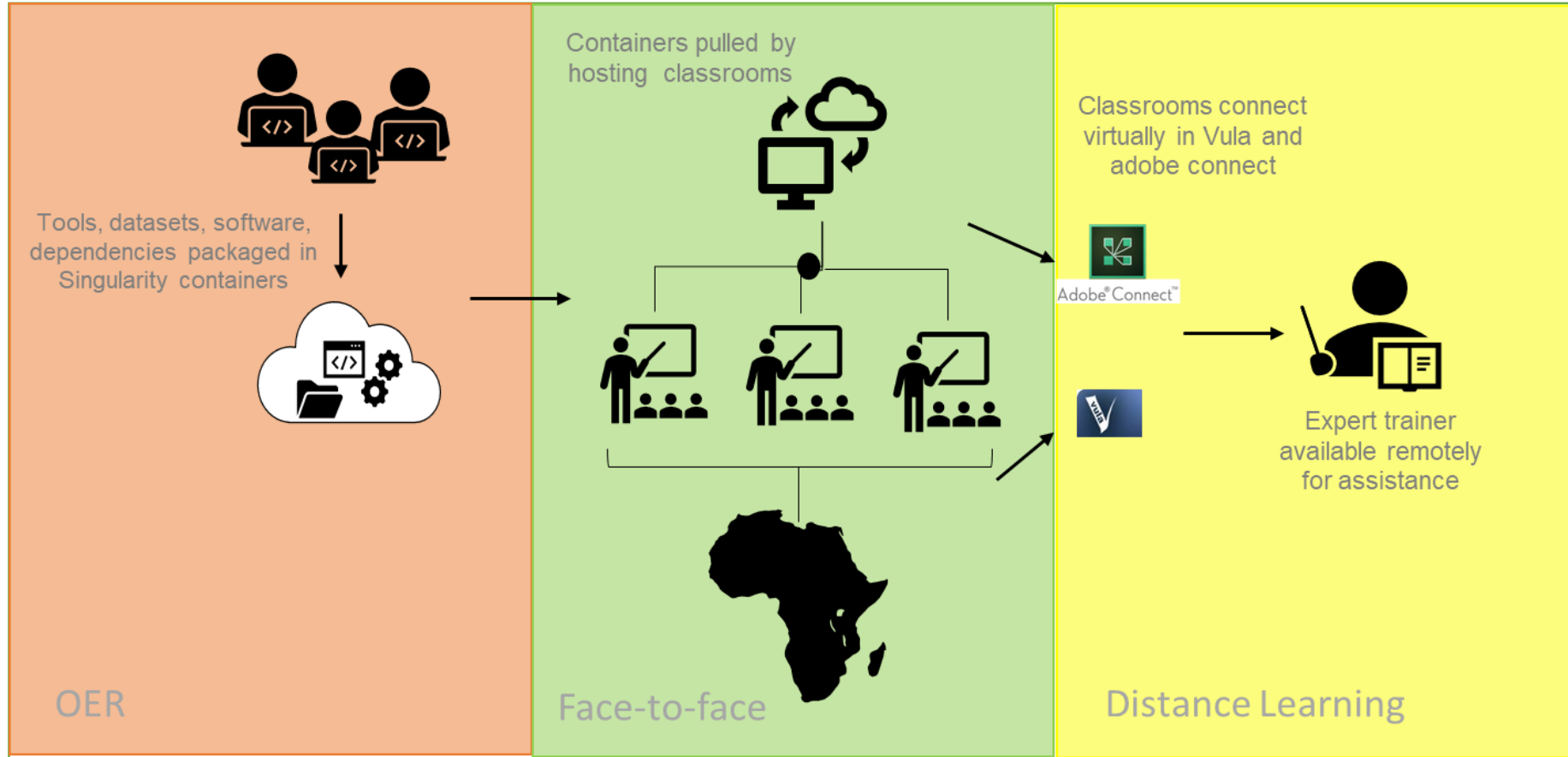
Original model: Gurwitz et al. 2017



**H3ABioNet**

Pan African Bioinformatics Network for H3Africa

# Adapted Model for Advanced Training



Modified model: Ras et al. 2021



**H3ABioNet**

Pan African Bioinformatics Network for H3Africa



# The Development of a Sustainable Bioinformatics Training Environment Within the H3Africa Bioinformatics Network (H3ABioNet)

Shaun Aron<sup>1\*</sup>, Paballo Abel Chauke<sup>2†</sup>, Verena Ras<sup>2†</sup>, Sumir Panji<sup>2†</sup>, Katherine Johnston<sup>2</sup> and Nicola Mulder<sup>2</sup> on behalf of the H3ABioNet Training and Education Work Package

<sup>1</sup>Sydney Brenner Institute for Molecular Bioscience, University of the Witwatersrand, Johannesburg, South Africa,

<sup>2</sup>Computational Biology Division, Institute of Infectious Disease and Molecular Medicine, CIDRI Africa Wellcome Trust Centre, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa

## OPEN ACCESS

### Edited by:

Hugo Verli,  
Federal University of Rio Grande do  
Sul, Brazil

### Reviewed by:

Peter Van Heusden,  
University of the Western Cape,  
South Africa  
Erik Bongcam-Rudloff,  
Swedish University of Agricultural  
Sciences, Sweden

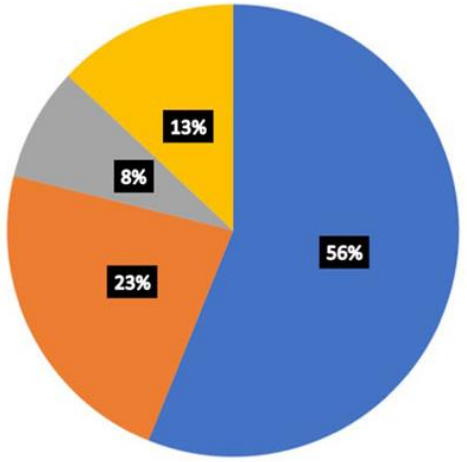
Bioinformatics training programs have been developed independently around the world based on the perceived needs of the local and global academic communities. The field of bioinformatics is complicated by the need to train audiences from diverse backgrounds in a variety of topics to various levels of 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 global 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

<https://www.frontiersin.org/articles/10.3389/feduc.2021.725702/full>



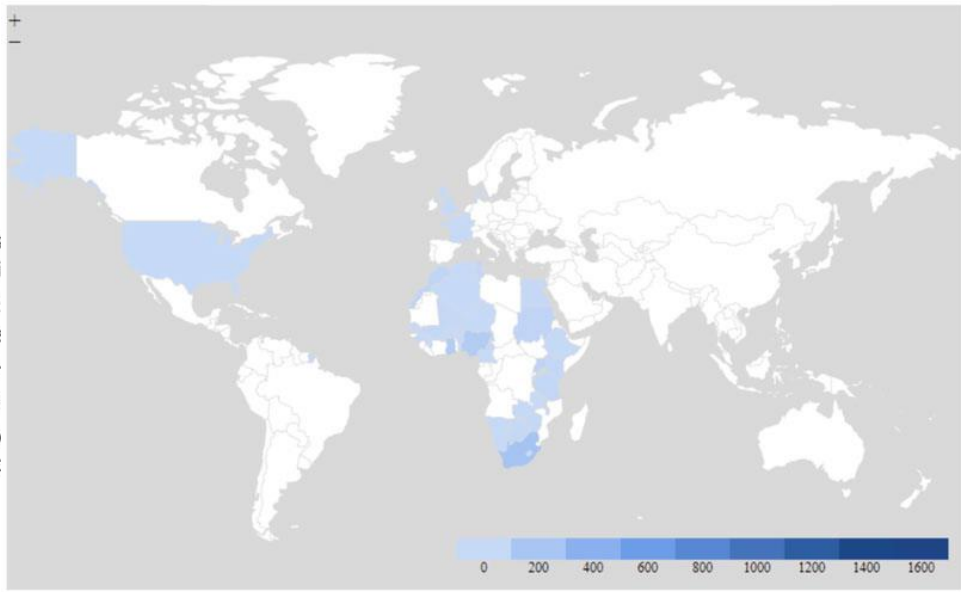
**H3ABioNet**

Pan African Bioinformatics Network for H3Africa



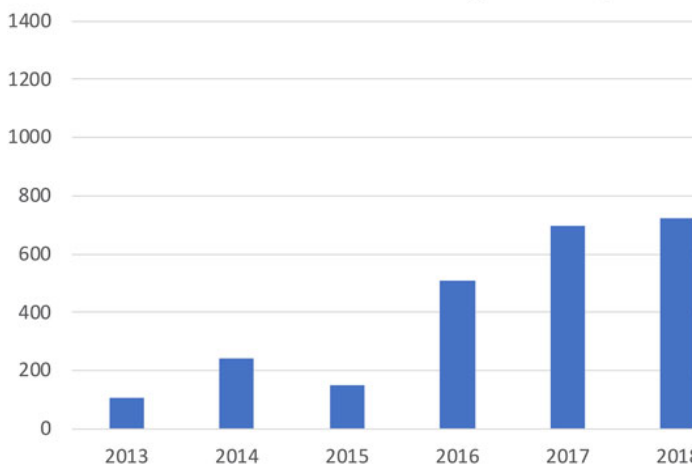
- Bioinformatics user (a wet lab biologist)
- Bioinformatics scientist (a dry lab biologist)
- Bioinformatics professional (analyst, developer)
- Person not involved in bioinformatics

### Face-to-Face Training Attendees

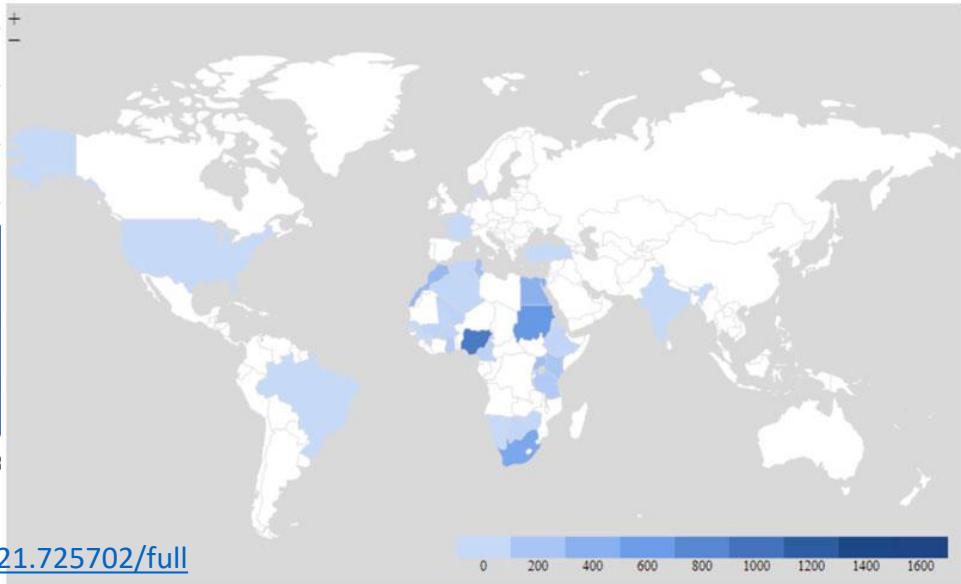


	Introductory Bioinformatics	Metagenomics	Professional development	Genomics	Technical/Software	GWAS	Data Management	Education	Intermediate Bioinformatics	NGS	SysAdmin	Advanced Bioinformatics	Research skills
Introductory Bioinformatics	3382	207	32	36	19	17	6	10	28	18	2	15	2
Metagenomics	207	492	8	16	12	9	2	5	6	9	0	1	1
Professional development	32	8	93	0	3	13	8	7	4	4	2	2	0
Genomics	36	16	0	335	5	1	2	3	0	3	0	0	2
Technical/Software	19	12	3	5	78	6	5	4	2	4	2	0	1
GWAS	17	9	13	1	6	105	12	10	10	6	2	1	2
Data Management	6	2	8	2	5	12	81	12	2	5	3	0	1
Education	10	5	7	3	4	10	12	68	1	6	2	1	4
Intermediate Bioinformatics	28	6	4	0	2	10	2	1	62	3	2	21	0
NGS	18	9	4	3	4	6	5	6	3	68	4	0	1
SysAdmin	2	0	2	0	2	2	3	2	2	4	39	0	2
Advanced Bioinformatics	15	1	2	0	0	1	0	1	21	0	0	31	0
Research skills	2	1	0	2	1	2	1	4	0	1	2	0	22

Number of Training Attendees per Year



### Online and Blended Training Attendees



Number of Attendees by Trainees

Year	Attendees
2013	1
2014	1
2015	1
2016	8
2017	9
2018	11

Training event (excl. 2021)

Images from: Aron, et al. 2021  
<https://www.frontiersin.org/articles/10.3389/feduc.2021.725702/full>

# FAIR and Bioschemas

Examples for all properties

- Findable
- Accessible
- Interoperable
- Reusable

<https://bioschemas.org/>

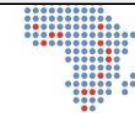
```
//Property: about
"about": [
  {
    "@type": "DefinedTerm",
    "@id": "http://edamontology.org/topic_3316",
    "inDefinedTermSet": "http://edamontology.org",
    "termCode": "topic_3316",
    "name": "Computer Science",
    "url": "https://bioportal.bioontology.org/ontologies/EDAM/?p=classes&conceptid=http%3A%2F%2Fedamontology.org%2Ftopic_3316"
  }, {
    "@id": "https://zenodo.org/record/1480624#.W-2as3r0nCM"
  }
]

//Property: abstract
"abstract": "This tutorial will take you through the basic usage of the command line shell. In it, you will discover what a shell is, skills for direc

//Property: accessibilitySummary
"accessibilitySummary": "Visual elements are not described."

//Property: audience
"audience": {
  "@type": "Audience",
  "audienceType": "PhD students"
}

//Property: author
"author": [
  {
    "@type": "Person",
    "name": "Elizabeth Windsor",
    "email": "e.r.windsor@gmail.com"
  }
]
```



**H3ABioNet**

Pan African Bioinformatics Network for H3Africa

# FAIR and Bioschemas

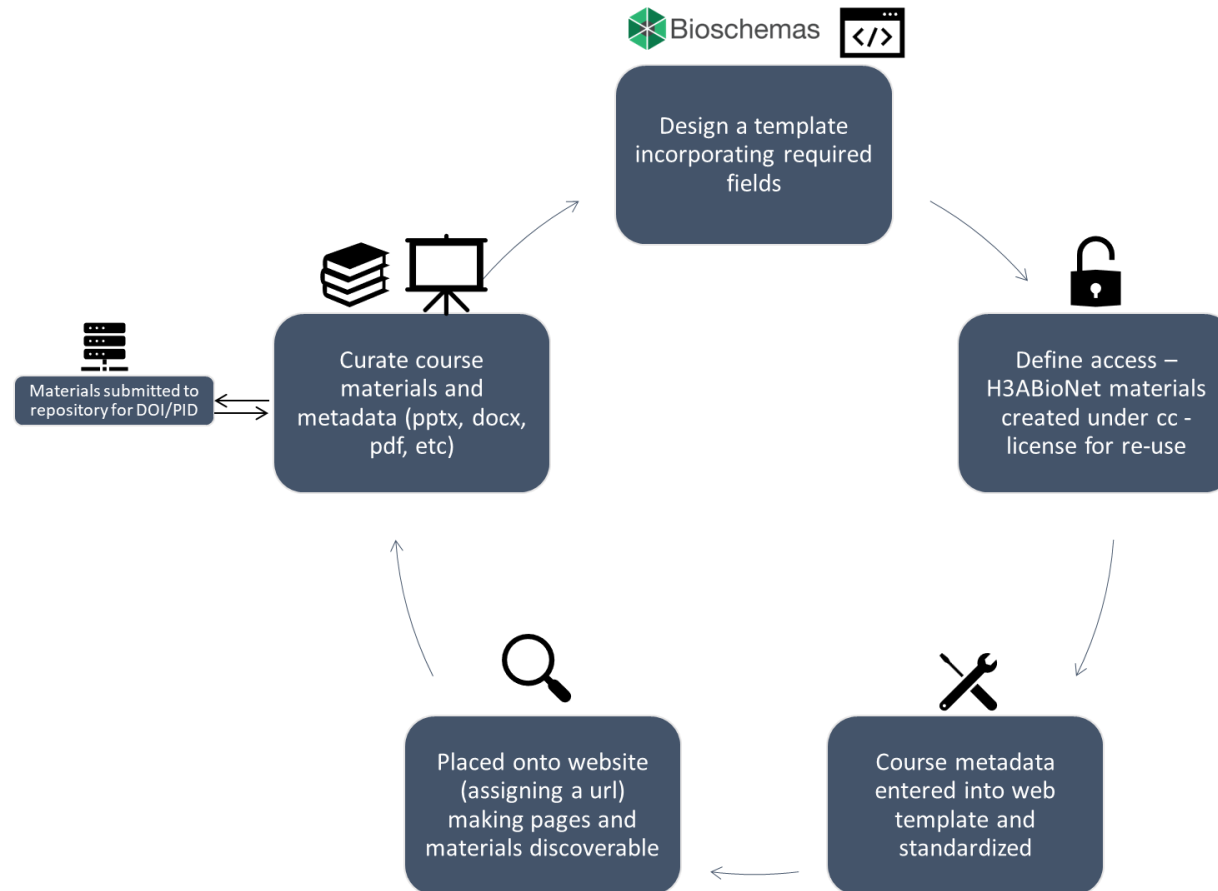


Image from: Aron, et al. 2021

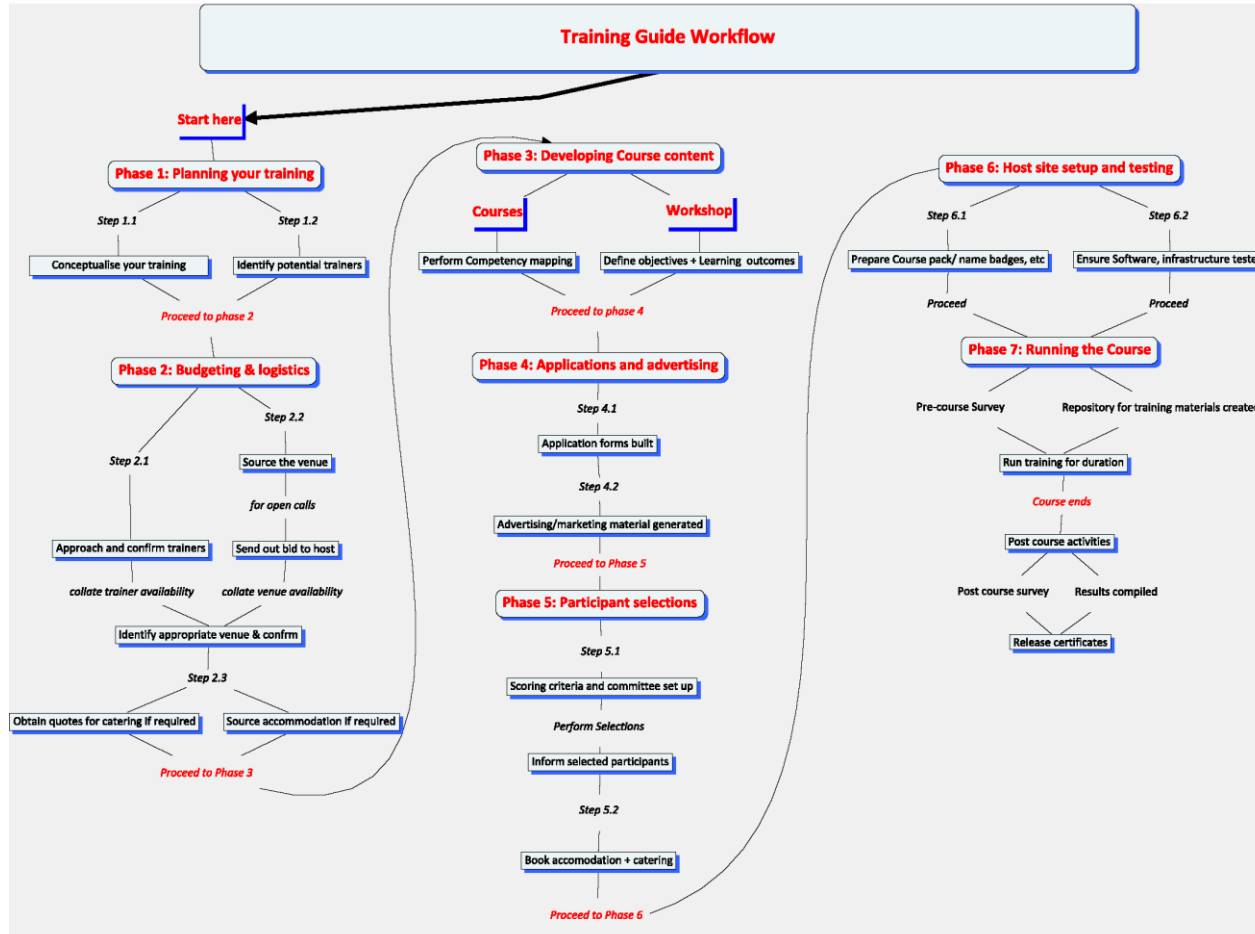
<https://www.frontiersin.org/articles/10.3389/feduc.2021.725702/full>



**H3ABioNet**

Pan African Bioinformatics Network for H3Africa

# Guides and SOP's



**H3ABioNet**  
Pan African Bioinformatics Network for H3Africa

WP\_Training Guide.pdf (1.58 MB) 1/58

### H3ABioNet TEWP Training Guide

Cite Download (1.58 MB) Share Embed + Collect

**Version 2** Educational Resource posted on 30.08.2021, 11:32 by Verena Ras, Sumir Panji, Faisal M. Fadlilmola, Shaun Aron, Lyndon Zass, Nicola Mulder

USAGE METRICS  
799 views | 226 downloads | 0 citations

This document aims to serve as a general guideline to assist in developing and running (Bioinformatics) training events. It summarizes the major steps taken by H3ABioNet to plan/organize a short-term workshop or course. H3ABioNet often tailors these approaches depending on the needs of a particular training but examples have been included where available.

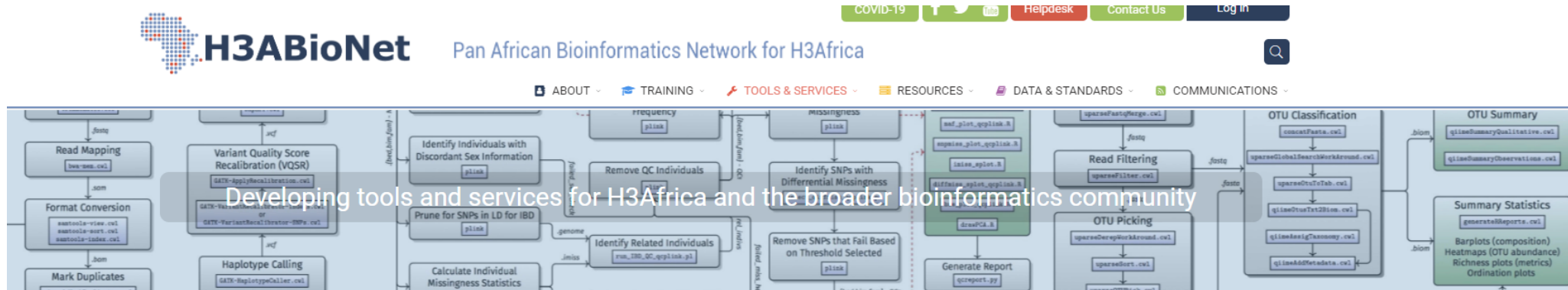
A separate archive is available containing (only) the original template documents that have been provided as part of the main guide (for easy reference) here:  
<https://doi.org/10.25375/uct.14229752>

Read the peer-reviewed publication  
H3ABioNet Training Support Pack

ZivaHub Support | Data Deposit | About Figshare | Features | FAQ | Privacy Policy | Cookie Policy | T&Cs | Disclaimer | Sitemap



# Where to Find More Information?



## Tools & Services

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

## 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.

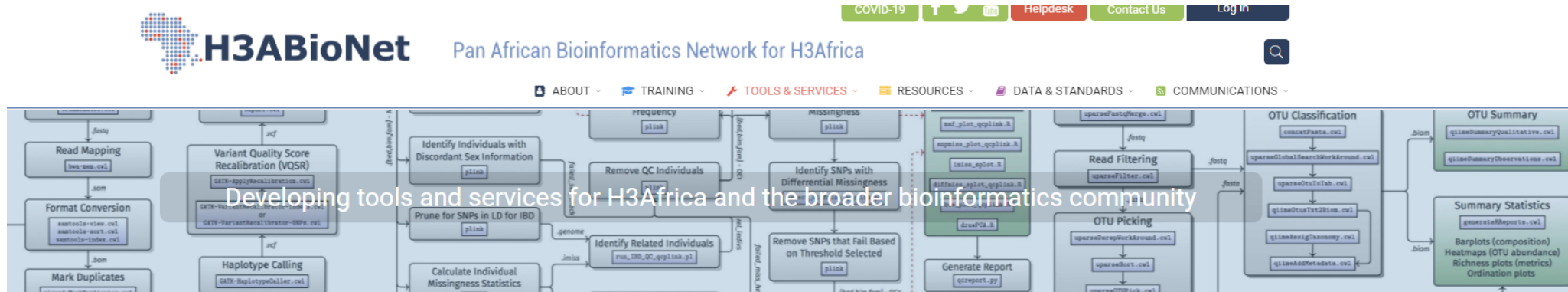
- Bioinformatics Tools +
- Bioinformatics Workflows +
- Technical Guidelines +
- Computing Infrastructure +
- Node Accreditation +



**H3ABioNet**

Pan African Bioinformatics Network for H3Africa

# Where to Find More Information?



## Tools & Services

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

## 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 +



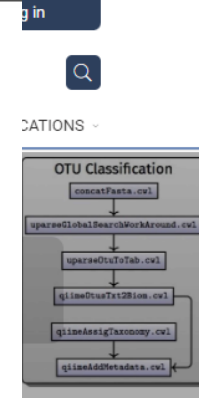
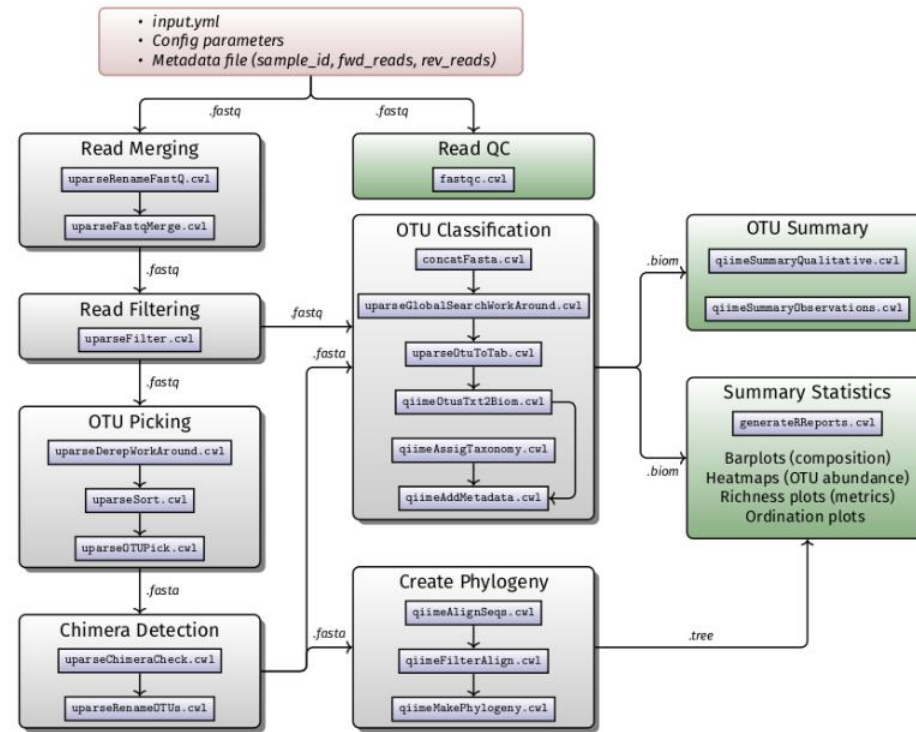
**H3ABioNet**

Pan African Bioinformatics Network for H3Africa

# Where to Find More Information?

16S rDNA data. We have implemented a workflow that performs QC of reads, generates OTUs, QC OTUS, does OTU classification and generates some descriptive statistics.

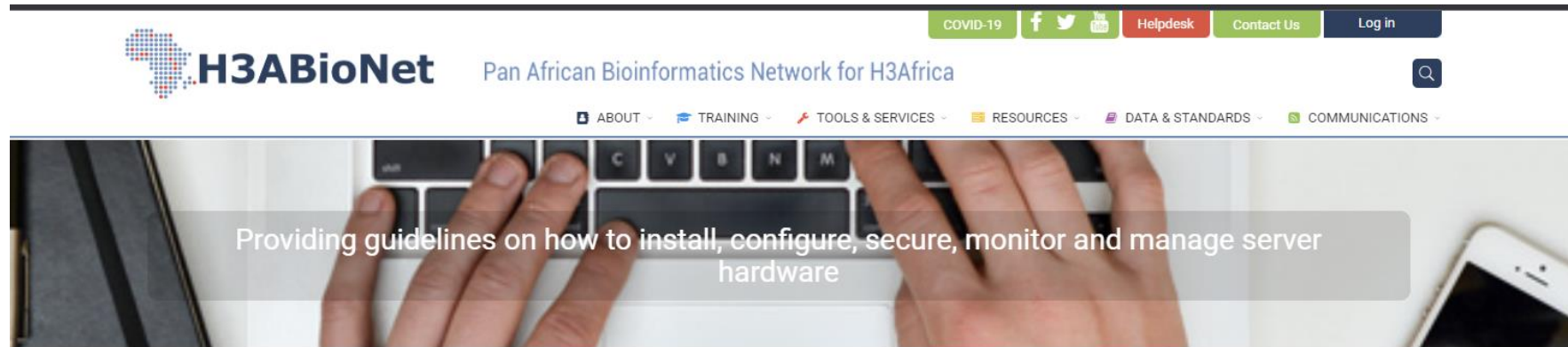
- SOP
- Workflow v1
- Workflow v2








maintain a  
if SOPs



# 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 howto documents to assist H3ABioNet system administrators with installing, configuring and securing, monitoring and managing their server hardware.

The howto 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 howto's have been divided into three levels: level1, 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](#) +

 [Education and Training](#)

 [Tools and Services](#)

 [Resources](#)

 [Communication](#)



**H3ABioNet**

Pan African Bioinformatics Network for H3Africa

# Where to Find More Information?



Pan African Bioinformatics Network for H3Africa

COVID-19



Helpdesk

Contact Us

Log in

ABOUT TRAINING TOOLS & SERVICES RESOURCES DATA & STANDARDS COMMUNICATIONS

H3ABioNet delivers high quality training covering various aspects of bioinformatics

## Training

H3ABioNet Courses and Events

Internships

Bioinformatics Education

Online Training

## 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 specialised ones such 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



H3ABioNet

Pan African Bioinformatics Network for H3Africa

# 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**

Instructor	Country		
Andrew Espira	Kenya		
Winfred Gatua	Kenya		
Ruth Nanjala	Kenya		
Kauthar Omar	Kenya		
Chaimae Samtal	Morocco		
Geoffrey Kimani	Rwanda		
Verena Ras	South Africa		
Nihad Alsayed	Sudan		
Upendo Masamu	Tanzania		
Liberata Mwita	Tanzania		
Melek Chaouch	Tunisia		



**H3ABioNet**

Pan African Bioinformatics Network for H3Africa

# 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**



**H3ABioNet**

Pan African Bioinformatics Network for H3Africa