Applying core competencies in Africa: the H3ABioNet experience

Nicola Mulder
University of Cape Town

H3ABioNet
Pan African Bioinformatics Network for H3Africa

What is H3Africa?

- **"The Human Heredity and Health in Africa (H3Africa) Initiative** aims to facilitate a contemporary research approach to the study of genomics and environmental determinants of common diseases with the goal of improving the health of African populations. To accomplish this, the H3Africa Initiative aims to contribute to the development of the necessary expertise among African scientists, and to establish networks of African investigators."

www.h3africa.org

H3ABioNet

Pan African Bioinformatics Network for H3Africa

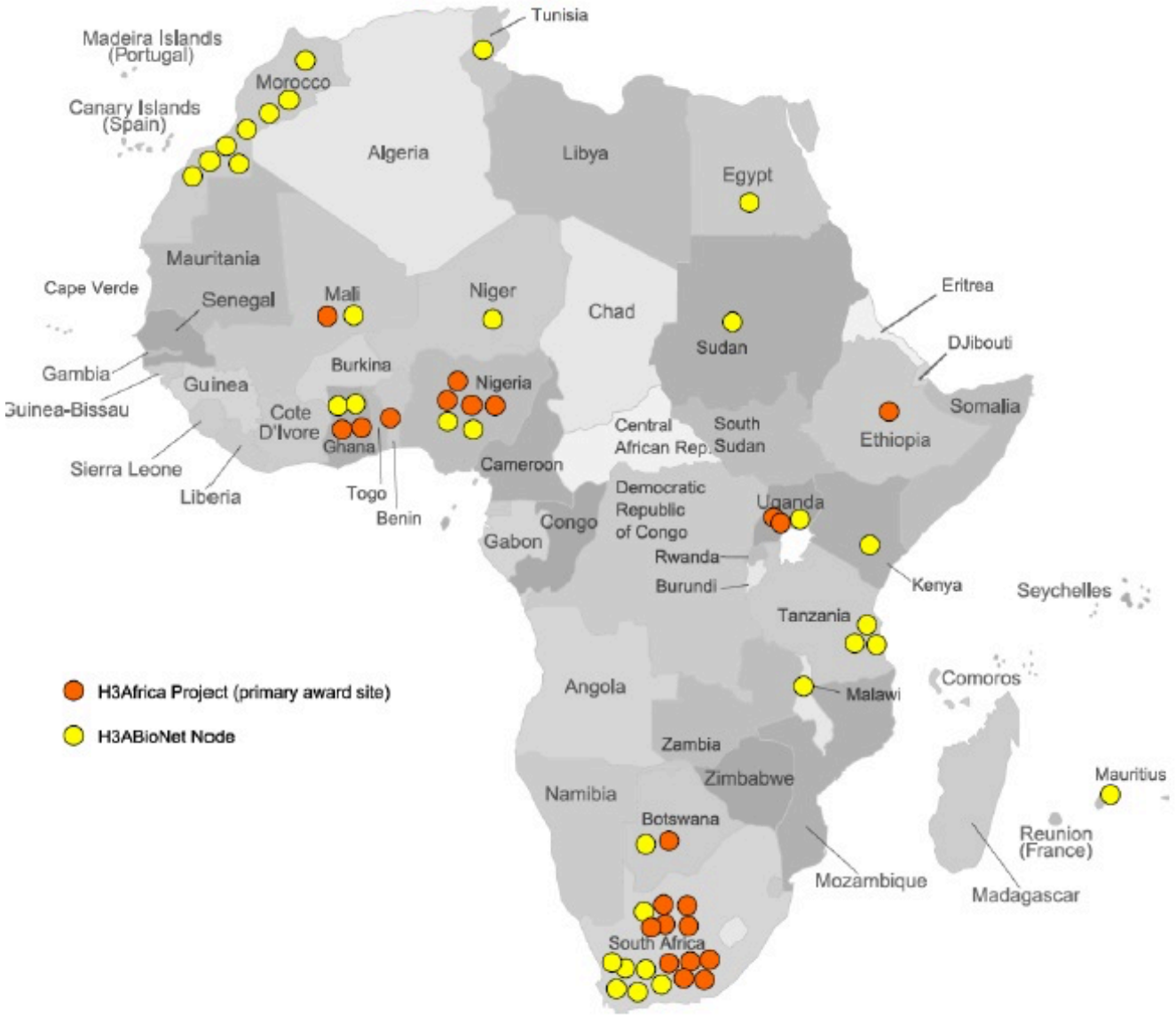
What is H3ABioNet?

- H3ABioNet is a pan-African bioinformatics network that aims to develop sustainable African bioinformatics capacity (intellectual and physical infra-structure) to cope with the on going “omics” revolution and ensure African science is not left behind.
- The “omics” revolution for human populations is coming to Africa via a series of H3Africa funded projects.
- H3ABioNet has 32 institutions in 15 African countries, 1 in USA and 1 in UK, led by Computational Biology Group at the University of Cape Town.

www.h3abionet.org

H3ABioNet
Pan African Bioinformatics Network for H3Africa

H3ABioNet distribution in relation to H3Africa primary award sites



- 32 Nodes within Africa
- Nodes at different levels of bioinformatics expertise
- Geographically placed to help develop regional bioinformatics capacity for the Node and H3Africa project in the region

Need for MScs in Bioinformatics

- Only a handful of MSc in Bix outside SA
- Many institutions wishing to establish programs
- Held Bioinformatics curriculum development workshop in Botswana
- Established African Bioinformatics Education Committee (ABEC)

H3ABioNet

Pan African Bioinformatics Network for H3Africa

ABEC

- http://www.h3abionet.org/training-and-education/african-bioinformatics-education-committee
- Curriculum development task force
- Set up website with guideline documents:
  - Considerations –key steps for starting a program
  - University processes for developing new programs
  - Existing programs and curricula
  - Curriculum development

H3ABioNet Bioinformatics curriculum

- Set of core and elective modules for a bioinformatics program using the ISCB guidelines have been defined
- Based on discussion, needs of individual institutions and existing curricula
- Trainers determined the content and contact hours for these modules
- Have included suggested lecturers from Africa
- http://training.h3abionet.org/curriculum_development_wg/

H3ABioNet
Pan African Bioinformatics Network for H3Africa

H3ABioNet Bioinformatics curriculum

H3ABioNet Curriculum Development Taskforce

H3ABioNet taskforce to develop a curriculum of bioinformatics to be used throughout the continent

H3ABioNet Curriculum Development Taskforce

In the wake of the Gaborone Workshop (H3ABioNet / University of Botswana, March 11th-12th 2014), H3ABioNet decided to build a taskforce to develop a standard curriculum in bioinformatics, at the MSc level.

This site is the repository for the documents produced by this Curriculum Development Task-Force (CDTF), and primarily for the module outlines as proposed by the CDTF members.

Please find here the different modules proposed and the list of volunteers for these. The menu on top of this page provides links to the different outlines.

The meeting minutes are to be found on this website, here (tab "WG Meeting Minutes" on top of this page).

Example module page

Module outline by the H3ABioNet Curriculum Development Taskforce

Biostatistics I

2014-03-06 07:03:32 Jean-Baka Domelevo Entfellner

Prepared by: **Jean-Baka Domelevo Entfellner**
Possible Lecturers: Jean-Baka Domelevo Entfellner, or any bioinformatician with a strong background in mathematics and statistics, ideally from his/her primary education.
Contact hours: For consistency reasons, each contact hour is fixed at 45min.
Theory (23), Practicals (30)

SPECIFIC OUTCOMES ADDRESSED

1. Generally speaking: develop an understanding of stochastic experiments
2. Understand and be able to build the framework of a statistical test

BACKGROUND KNOWLEDGE REQUIRED

Basic general-purpose scientific knowledge, basic arithmetic skills, and some familiarity with basic linear algebra.

BOOKS & OTHER SOURCES USED

1. Fundamentals of Biostatistics, 7th edition, by Bernard Rosner (Cengage Learning, 2011)
2. Biostatistics with R — An introduction to statistics through biological data, by Babak Shahbaba (Springer, 2012)

COURSE CONTENT

(A) Theory lectures

I. Probability theory:
1. Atomic and complex events, probabilities as a measure on sets.

Example module page

Module outline by the H3Al Curriculum Development Ta

Biostatistics I

2014-03-06 07:03:32 Jean-Baka Domelevo Entfellner

Prepared by: **Jean-Baka Domelevo Entfellner**
Possible Lecturers: Jean-Baka Domelevo Entfellner, or any
with a strong background in mathematics and statistics, id
primary education.
Contact hours: For consistency reasons, each contact hou
Theory (23), Practicals (30)

SPECIFIC OUTCOMES ADDRESSED

1. Generally speaking: develop an understanding of stocha
2. Understand and be able to build the framework of a stat

BACKGROUND KNOWLEDGE REQUIRED

Basic general-purpose scientific knowledge, basic arithmet
familiarity with basic linear algebra.

BOOKS & OTHER SOURCES USED

1. Fundamentals of Biostatistics, 7th edition, by Bernard R
Learning, 2011)
2. Biostatistics with R — An introduction to statistics throug
by Babak Shahbaba (Springer, 2012)

COURSE CONTENT

(A) Theory lectures

I. Probability theory:
1. Atomic and complex events, probabilities as a measure

3. Enumerative combinatorics: counting permutations, combinations and partitions. Binomial coefficients.
4. Some common discrete probability distributions: Bernoulli, binomial, Poisson. Behaviour of a binomial when the number of trials tends to infinity. Concepts: probability mass, expectation of a discrete distribution.
5. First continuous probability distributions: uniform, exponential
6. Central limit theorem and the normal distributions

7. Other continuous distributions: Student's t and chi-square distributions.
II. Statistical hypothesis testing
III. Analysis of variance and regression models
IV. Multidimensional dataset analysis: Principal Component Analysis

(B) Practical component

We suggest the use of Rstudio throughout the course, as an integrated development environment to work with R. Being the fundamental statistical software in use across various research areas, it is essential that the students develop mastery over R during this course.
Alternatively, if computing resources are extremely scarce, use an interactive R interpreter to demonstrate the concepts, plus a simple text editor later on, once the students start writing functions.
This section "practical component" follows the same structure as the previous section "Theory lectures": practicals just aim at having the students manipulate the concepts seen in the lectures, right after they were introduced to them.

ASSESSMENT ACTIVITIES AND THEIR WEIGHTS

We would suggest two written exams during the course of the module (total weight = ½) , and a final programming exam (weight = ½). Of course, practicals can also be for marks all along the module, but our advice is not to make each and every practical for marks, not to put too much counter-productive stress on the students. Practicals are the privileged moments when students actually understand the concepts as they put them into play.

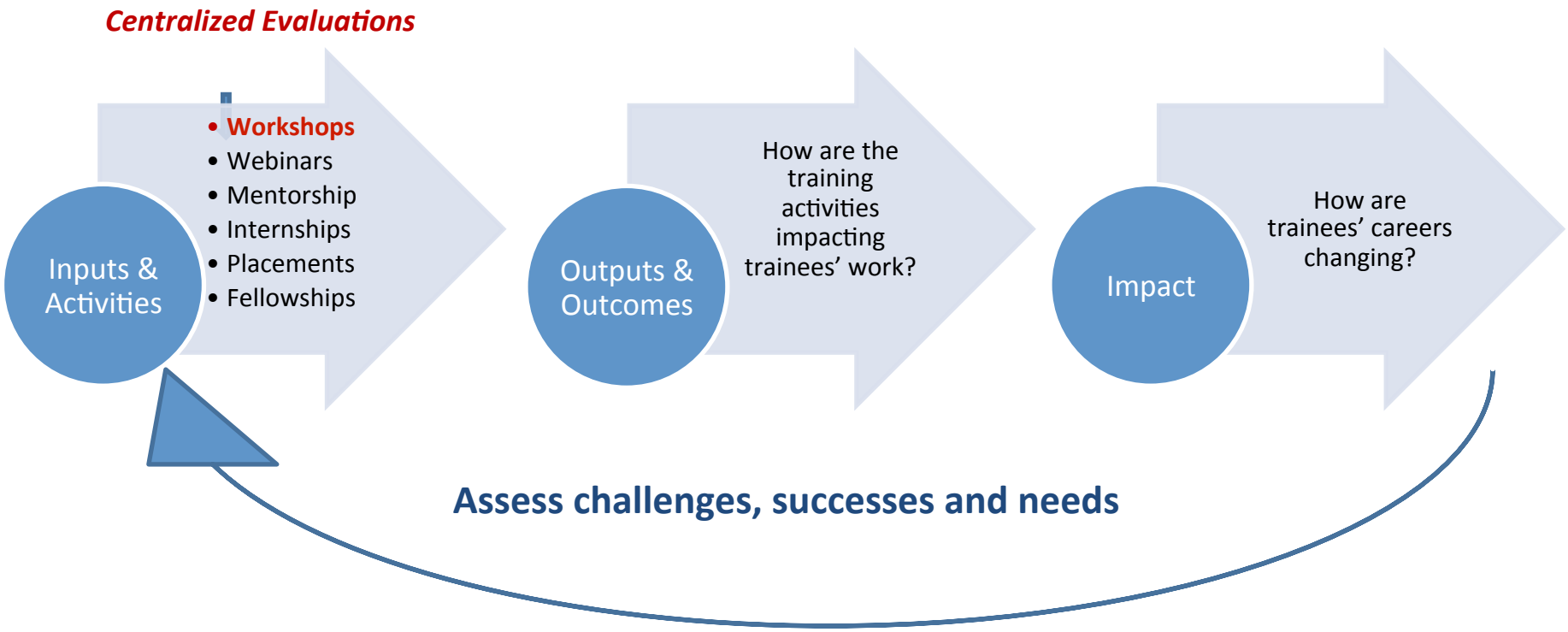
Implementations

- First MSc in Bioinformatics program launched at the University of Bamako this year, lecturers local and distant
- New progam in Malawi about to be submitted for approval
- Other MSc in Bioinformatics programs such as in Kenya, Tunisia are being developed using the defined modules
- Using some as a base for Introduction to Bioinformatics course

H3ABioNet

Pan African Bioinformatics Network for H3Africa

Training evaluation framework

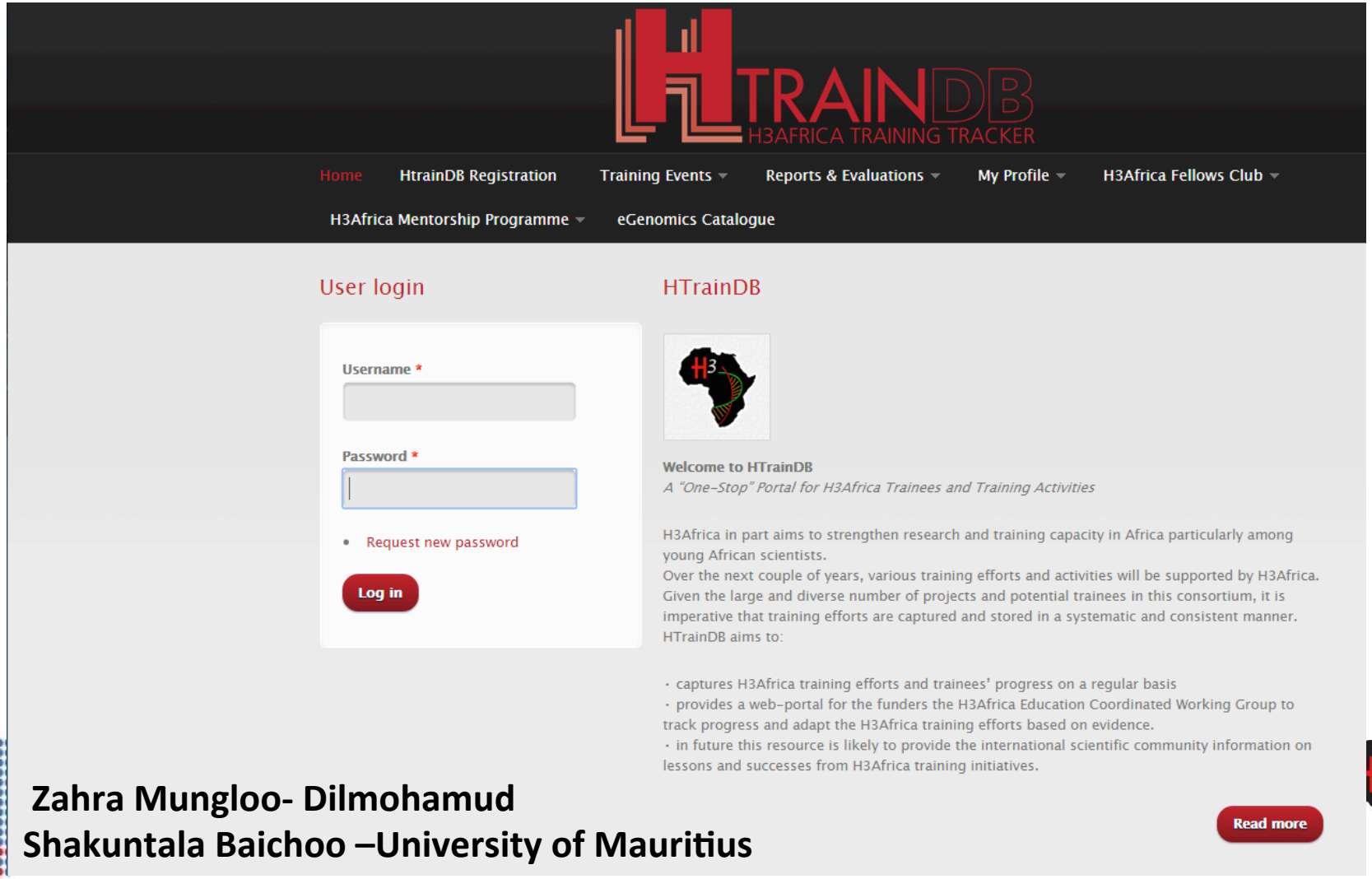


Evaluation results used to adjust training activities

H3ABioNet
Pan African Bioinformatics Network for H3Africa

Tracking Training and Education

- Developing a system to track, monitor and evaluate the training development of students and staff



Zahra Mungloo- Dilmohamud
Shakuntala Baichoo –University of Mauritius

Why HTrainDB?

- A **major goal** of H3Africa Consortium is to **train the next generation of "genomics" scientists**
- Ongoing training is widespread, un-trackable, and therefore not easy to monitor/evaluate
- Need to establish a "one-stop" center for **training tracking, info & opportunities for trainees**
- A tool to **facilitate reflections & planning**: where we are, where we have come from, plan next steps
- See Poster

First survey

- **498** trainees
- **10%** response rate to the survey
  - **76%** participated in H3Africa funded training
  - **94%** of trainees who attended H3Africa funded training shared their knowledge/information with at least 3 people

"The trainings have been very informative and my request is for advanced courses for those who have gone through the introductions and are using the tools"

"Systems admin workshop: i can now confidently setup and administer Linux servers even of other departments"

Other impact:
- Set up new collaborations
- Organised their own workshop

H3ABioNet
Pan African Bioinformatics Network for H3Africa