H3ABioNet

Pan African Bioinformatics Network for H3Africa

Progress relevant to consortium May 2015

Nicky Mulder: nicola.mulder@uct.ac.za

H3ABioNet
Pan African Bioinformatics Network for H3Africa

Summary of initial goals

- Short/medium term:
  - Store H3Africa data & enable submission to EGA
  - Build human and computing infrastructure
  - Train in bioinformatics theory and techniques
  - Apply techniques to data, with skills transfer
- Long term:
  - Train bioinformatics academics and support staff in H3ABioNet and H3Africa projects
  - Build network of bioinformaticians and data analysers

H3ABioNet
Pan African Bioinformatics Network for H3Africa

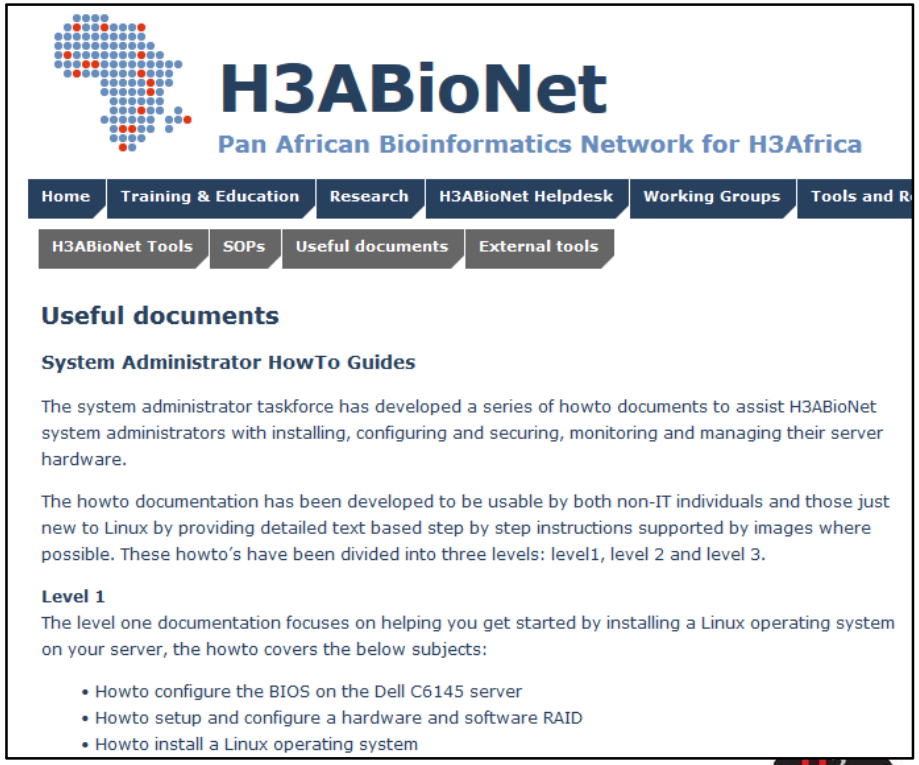
What's new?

- User support and infrastructure update
  - Developing computing infrastructure
  - Helpdesk
- H3Africa archive progress
- Training activities
- Research activities
- H3Africa consortium projects

H3ABioNet
Pan African Bioinformatics Network for H3Africa

Developing computing infrastructure

- Several nodes now have computing facilities
  - **Use H3ABioNet helpdesk to get info and access**
- Developed documentation for sys admins
  - **Useful for anyone investing in computing equipment**
- Doing survey on HPC and Galaxy
- Globus for data transfers
- NetMap for bandwidth testing



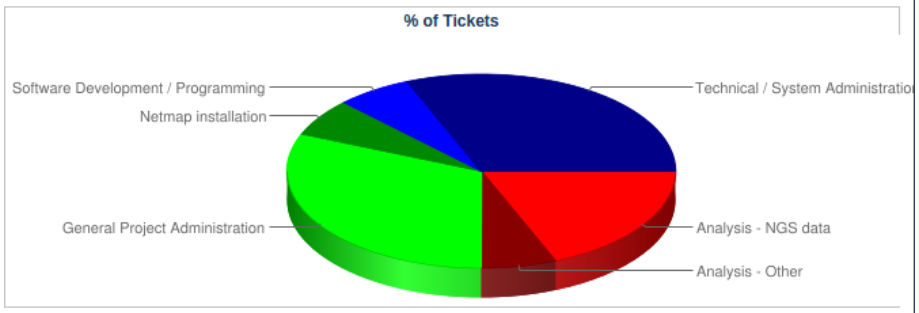
H3ABioNet
Pan African Bioinformatics Network for H3Africa

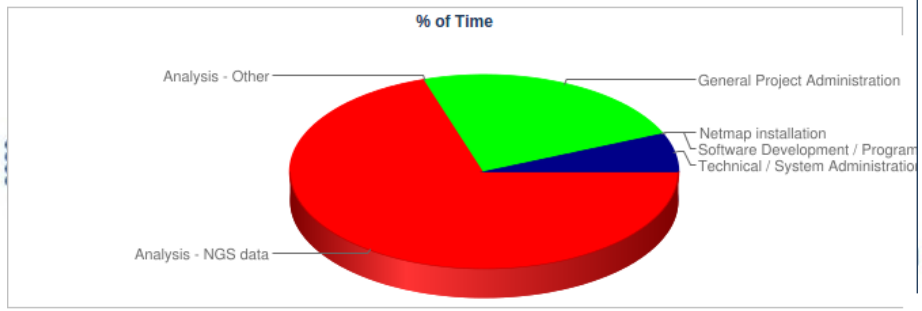
H3ABioNet help desk

- http://www.h3abionet.org/support/help-desk
- Managed by 15 volunteers with experience and expertise in Bioinformatics and Genomics.

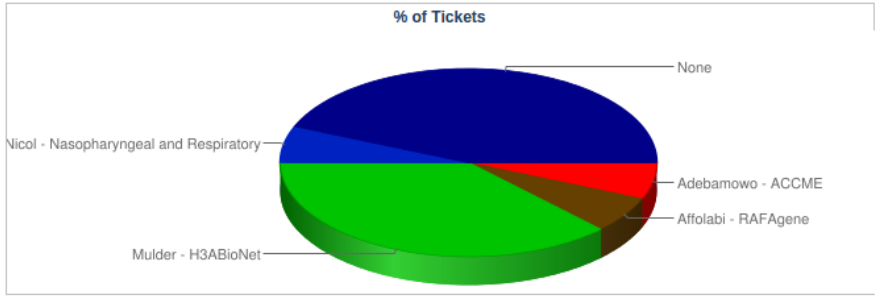
Stats 11/11/14-05/05/15

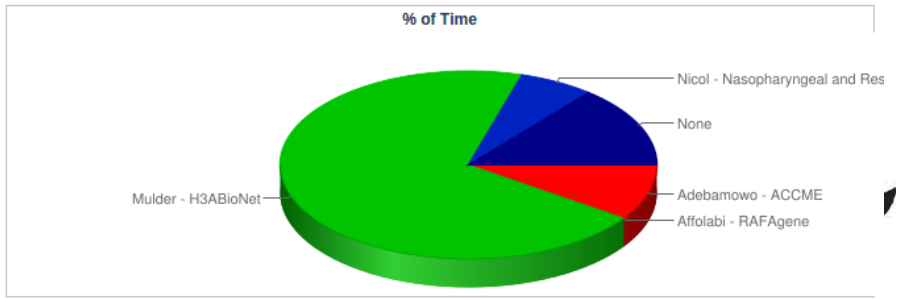
Category	Tickets	Time (min)	Average Time (min)	% of Tickets	% of Time
Analysis - NGS data	3	110	37	18.8	70.1
Analysis - Other	1	0	0	6.3	0.0
General Project Administration	5	37	7	31.3	23.6
Netmap installation	1	0	0	6.3	0.0
Software Development / Programming	1	0	0	6.3	0.0
Technical / System Administration	5	10	2	31.3	6.4
Total	**16**	**157**	**10**		





Department	Tickets	Time (min)	Average Time (min)	% of Tickets	% of Time
Adebamowo - ACCME	1	15	15	6.3	9.6
Affolabi - RAFAgene	1	0	0	6.3	0.0
Mulder - H3ABioNet	6	110	18	37.5	70.1
Nicol - Nasopharyngeal and Respiratory	1	10	10	6.3	6.4
None	7	22	3	43.8	14.0
Total	**16**	**157**	**10**		





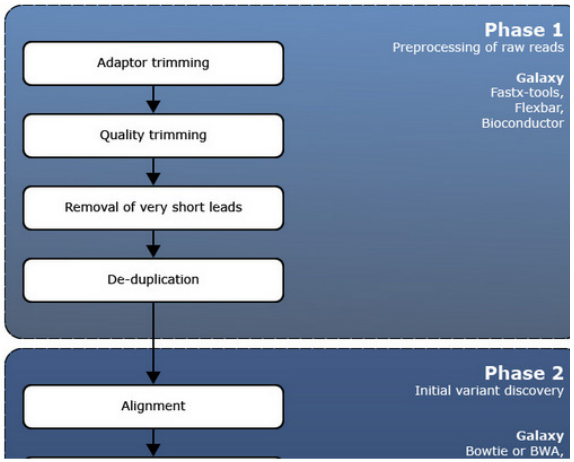
SOPs

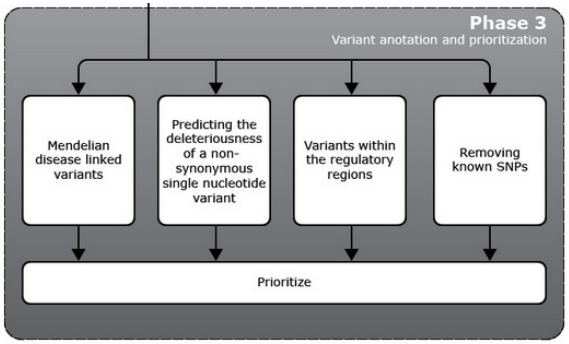
http://h3abionet.org/tools-and-resources/sops

SOP for NGS variant calling in whole exomes of Man

This Standard Operating Procedure describes the technical steps for Node Accreditation process for variant calling in exome data. This SOP was created by Dr. Liudmila Sergeevna Mainzer and Prof. Jongeneel C. Victor from the HPCBio facility at the University of Illinois at Urbana–Champaign and all copyright resides with them.

NGS Exome Variant Calling Workflow Schematic





GWAS

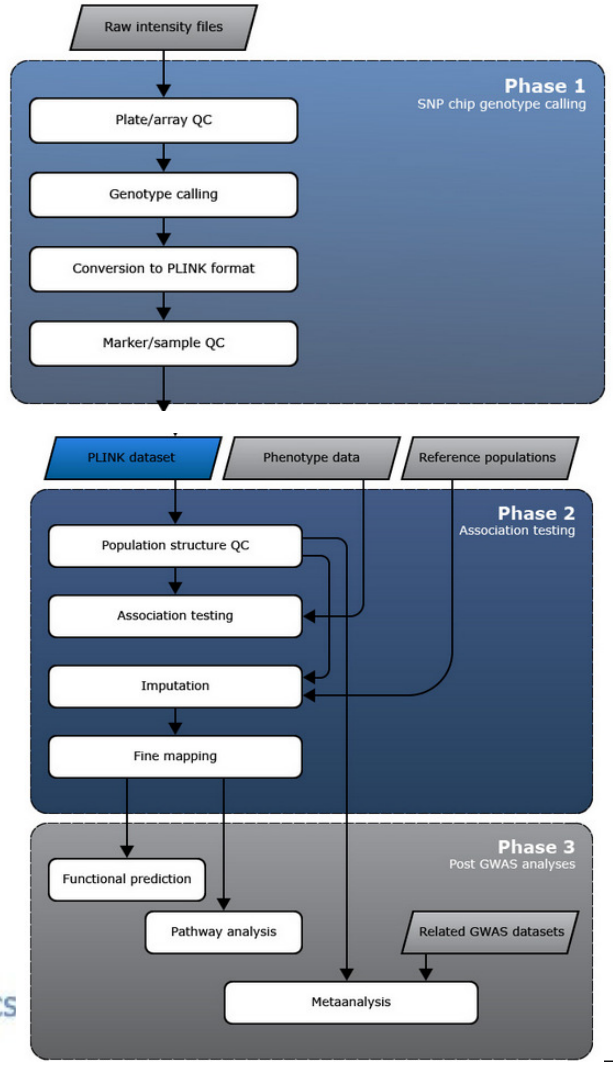
SNP chip genotype calling

- SOP
- Practice dataset

Association testing

- SOP
- Practice dataset

GWAS Workflow Schematic



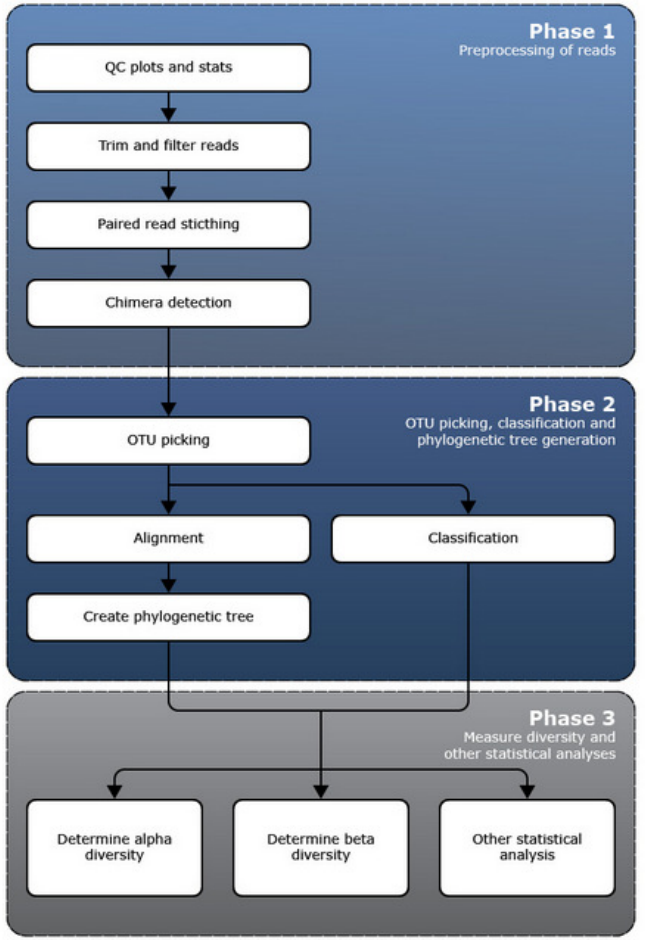
Standard operating procedure for 16S rRNA diversity analysis

Introduction

The genes encoding the RNA component of the small subunit of ribosomes, commonly known as the 16S rRNA in bacteria and archaea, are among the most conserved across all kingdoms of life. Nevertheless, they contain regions that are less evolutionarily constrained and whose sequences are indicative of their phylogeny. Amplification of these genomic regions by PCR from an environmental sample and subsequent sequencing of a sufficiently large number of individual amplicons enables the analysis of the diversity of clades in the sample and a rough estimate of their relative abundance. The analytical process is known as "16S rDNA diversity analysis", and is the focus of the present SOP.

The procedure and tools are only recommendations and it is up to the user to evaluate what works best for their needs.

Schematic workflow of the analysis



H3Africa Archive

- Submission of H3Africa data is a funder requirement
- Estimated overall storage capacity of 500TB
- Architecture modeled on EGA system, components:
- Landing area
  - datasets are encrypted by submitter
- Vault area
  - Focus on security, data only ever decrypted in Vault
  - All access and operations are logged
  - Analyses are limited to
    - QC validation
    - Checking EGA file format requirements
    - H3A metadata validation
  - Can assist with additional analyses, but only by agreement with PI
- Archive area
  - Purely for storage, no processing
  - encrypted files are mirrored to a separate physical location

H3ABioNet

Pan African Bioinformatics Network for H3Africa

H3Africa archive submission process

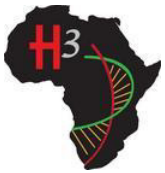
- **Register project with the H3A Archive 2 months prior to submission and provide:**
  - **Submission timeline**
  - **Blank consent form, ethics clearance numbers**
  - **Estimated sample count**
  - **What phenotype data**
    - **Will be collected by the project**
    - **Will be submitted for H3A storage**
    - **Submitted to EGA**
- EGA accessions will be provided
  - Study accession on H3A registration
  - Sample accessions on EGA submission

H3ABioNet
Pan African Bioinformatics Network for H3Africa

First submission: AWIGEN project

- Pilot data
- 973 samples
- Illumina MetaboChip (~200k SNPS after QC)
- Phenotypes captured:
  - Sex
  - Age
  - Height
  - Weight
  - Hip circumference
  - Waist circumference
  - Ethnicity

H3ABioNet
Pan African Bioinformatics Network for H3Africa



H3ABioNet training activities (1)

- Co-organised metagenomics workshop in Mauritius in Dec 2014
  - Participants from 11 African countries
  - Trainers from H3ABioNet, Greece, Sweden, Mauritius
- Provided travel fellowships to CafGEN GWAS course in Botswana
- Ran Advanced Systems Administration workshop in Pretoria in Feb 2015
  - Course in 2 parts, beginner and advanced
  - 24 participants from H3Africa projects and H3ABioNet
  - Trainers local and/or H3ABioNet

H3ABioNet
Pan African Bioinformatics Network for H3Africa

H3ABioNet training activities (2)

- Organised 2 workshops at end of ISCB Africa ASBCB Bioinformatics conference in Tanzania –Exploring variation data (EBI) and Data visualization (Amel Ghouila)
- Ran Introduction to biostatistics workshop in Tunis in March 2015
  - ~20 participants from H3Africa projects and H3ABioNet
  - Local and H3ABioNet trainers
- Ran Medical population genetics and GWAS for complex diseases workshop and symposium in Cape Town in April 2015
  - 32 participants from H3Africa, AIMS and other locals
  - Divided into 2 streams –data analysis, advanced modelling
  - 3 trainers from abroad, 2 H3ABioNet
  - Followed by symposium and discussions

H3ABioNet
Pan African Bioinformatics Network for H3Africa

H3ABioNet training activities (3)

- **Trained >450 people to date**
- Placed 8 interns –**internships available***
- Ongoing evaluation of online courses
- Wordpress sites available for all courses –many have recorded lectures. Links can be found at: http://training.h3abionet.org
- Progress on African Bioinformatics Education Committee activities:
  - University of Bamako, Mali started their first Masters course based on our curriculum with several H3ABioNet trainers

H3ABioNet
Pan African Bioinformatics Network for H3Africa

Plans for future training

- UVRI NGS course in July
- Introduction to bioinformatics course –live broadcasting to multiple sites
- Regional GWAS workshop for Tanzania

- **Awaiting your needs!!**

H3ABioNet
Pan African Bioinformatics Network for H3Africa

Research and outreach

- Several projects within the network to build capacity and skills, as well as tools
- Started special interest groups in e.g. GWAS, NGS, etc. –all welcome!
- Organised webinar series, testing platforms
- Work presented at ISCB Africa ASBCB Bioinformatics conference –March, Tanzania
- Increased presence on Facebook and Twitter

H3ABioNet

Pan African Bioinformatics Network for H3Africa

ISCB Africa ASBCB conference

H3ABioNet and Wellcome Trust provided ~30 travel fellowships

H3ABioNet
Pan African Bioinformatics Network for H3Africa

Facebook page

H3Africa consortium projects

- Recruitment ‘database’
- Ontologies for metadata
- *Biobank data integration*
- Trainer/trainee database
- Custom chip design

H3ABioNet

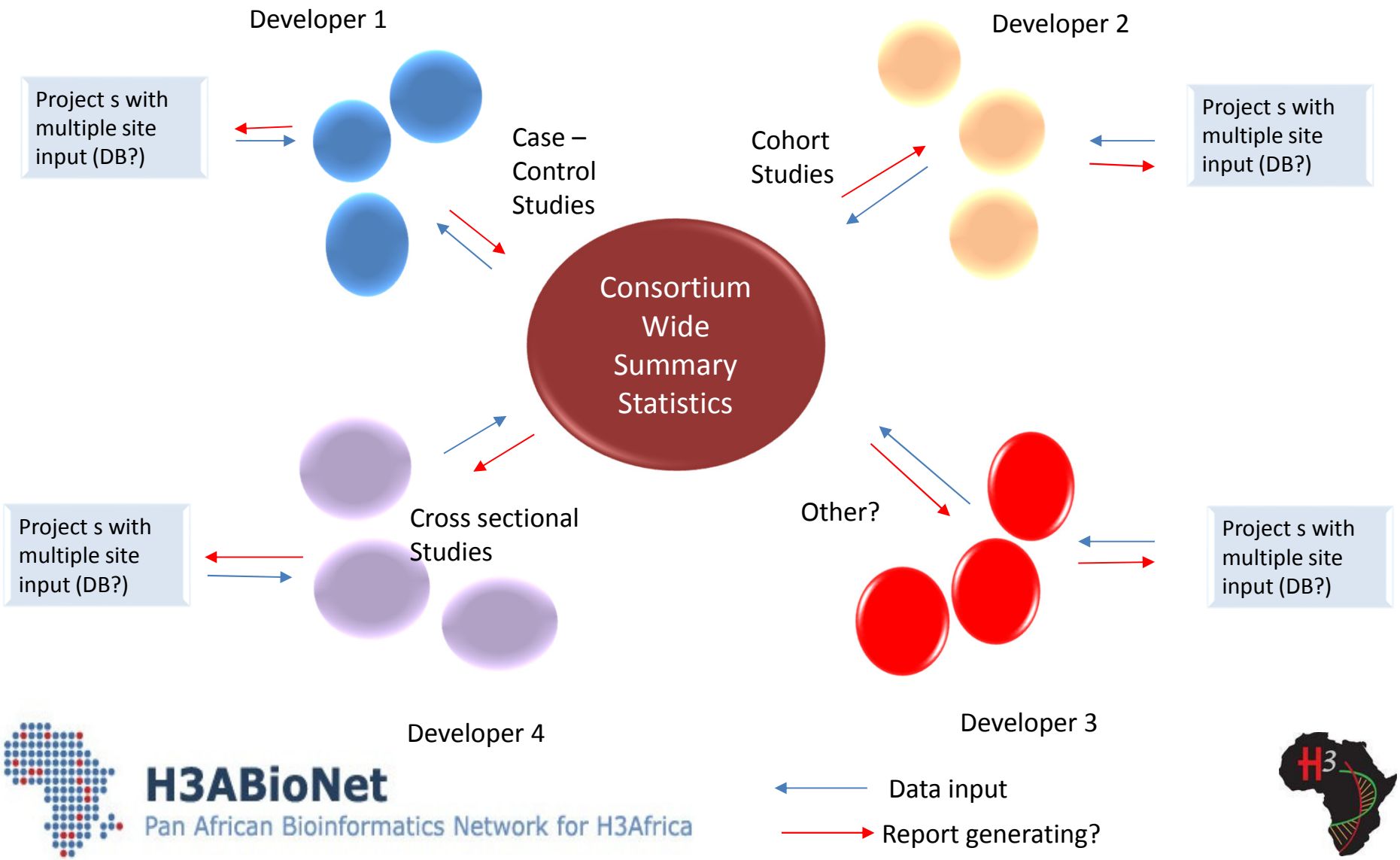
Pan African Bioinformatics Network for H3Africa

Recruitment database

- To track participant recruitment and data/sample workflow
- Aim to do as much pre-population of data as possible
- Need to customise it for each project

H3ABioNet
Pan African Bioinformatics Network for H3Africa

Recruitment database



H3ABioNet
Pan African Bioinformatics Network for H3Africa

Recruitment database mock-up

h3ardb

Search this site

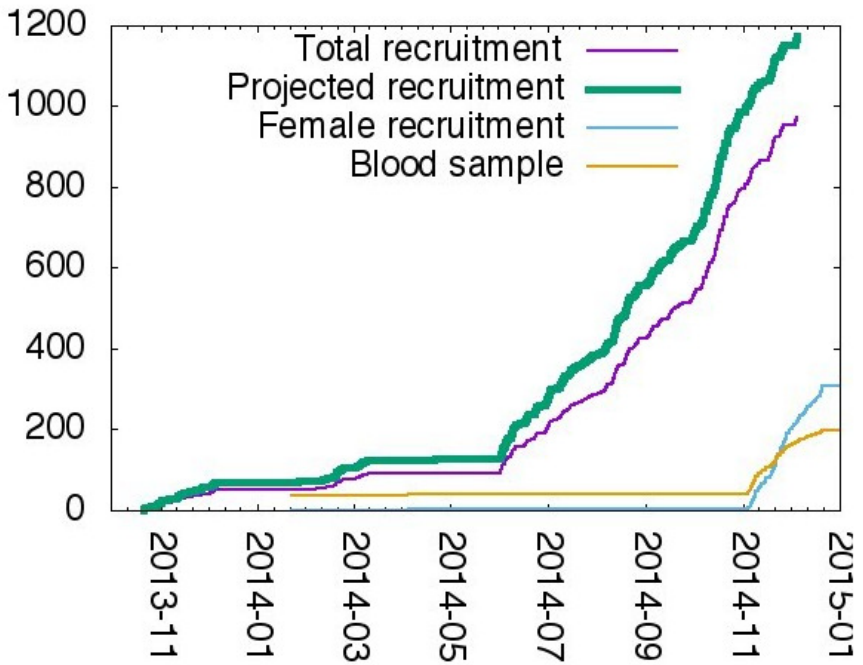
View Data Report

H3Africa Participant Recruitment Database >

Research Project Landing Page: H3A WomPom Diversity Project

- View report
- Enter new recruitment data
  - Online Spreadsheets
    - Use online spreadsheet
    - Progress of the participant recruitment per country
    - Progress by a reporting period
  - Set automated
  - Upload spreadsheet
- Upload recruitment projections

- View Data Recruitment report
- View Recruitment Progress per Country
- View Actual recruitment vs Projected recruitment per Country
- View Actual recruitment vs Projected recruitment per Site

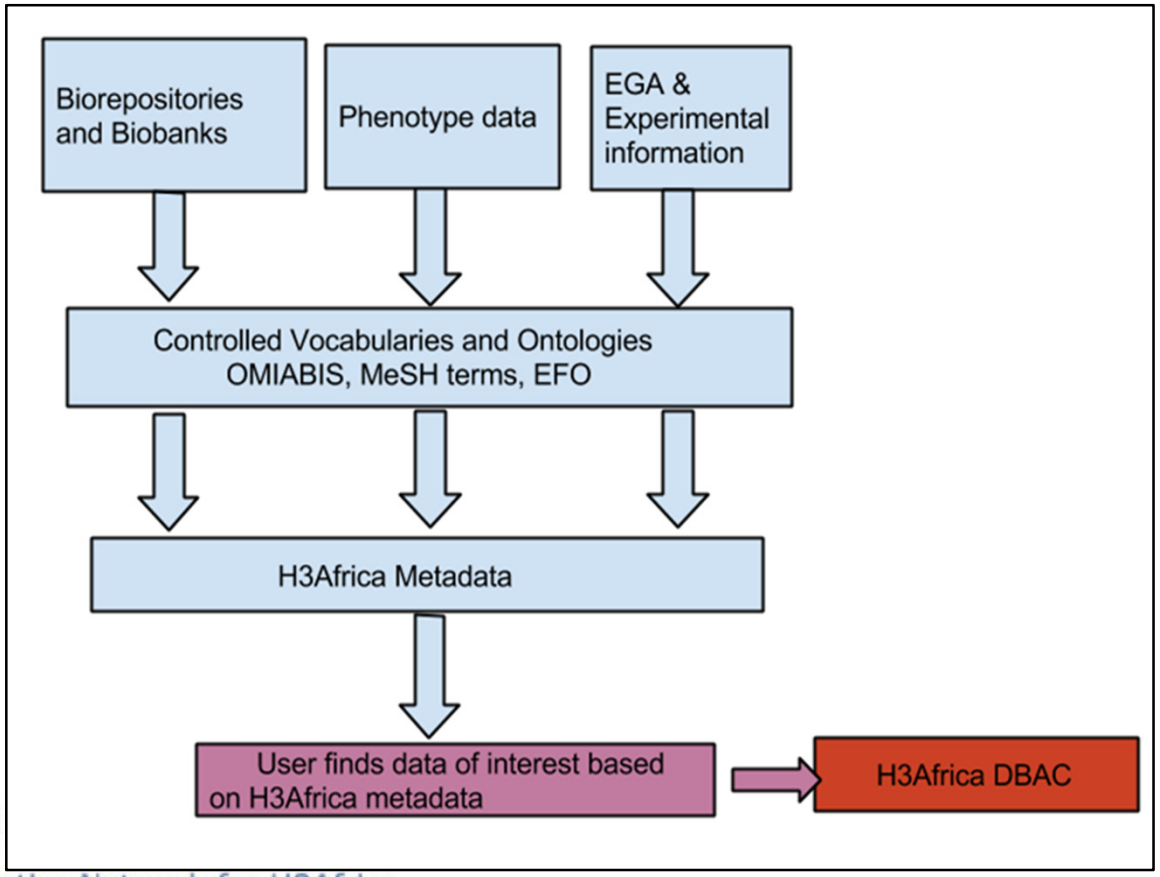


H3ABioNet

Pan African Bioinformatics Network for H3Africa

Ontologies & controlled vocabularies TF

- The groups aims to harmonize H3Africa data using standardized vocabularies and ontologies
- This harmonization will make it easier to query H3Africa metadata



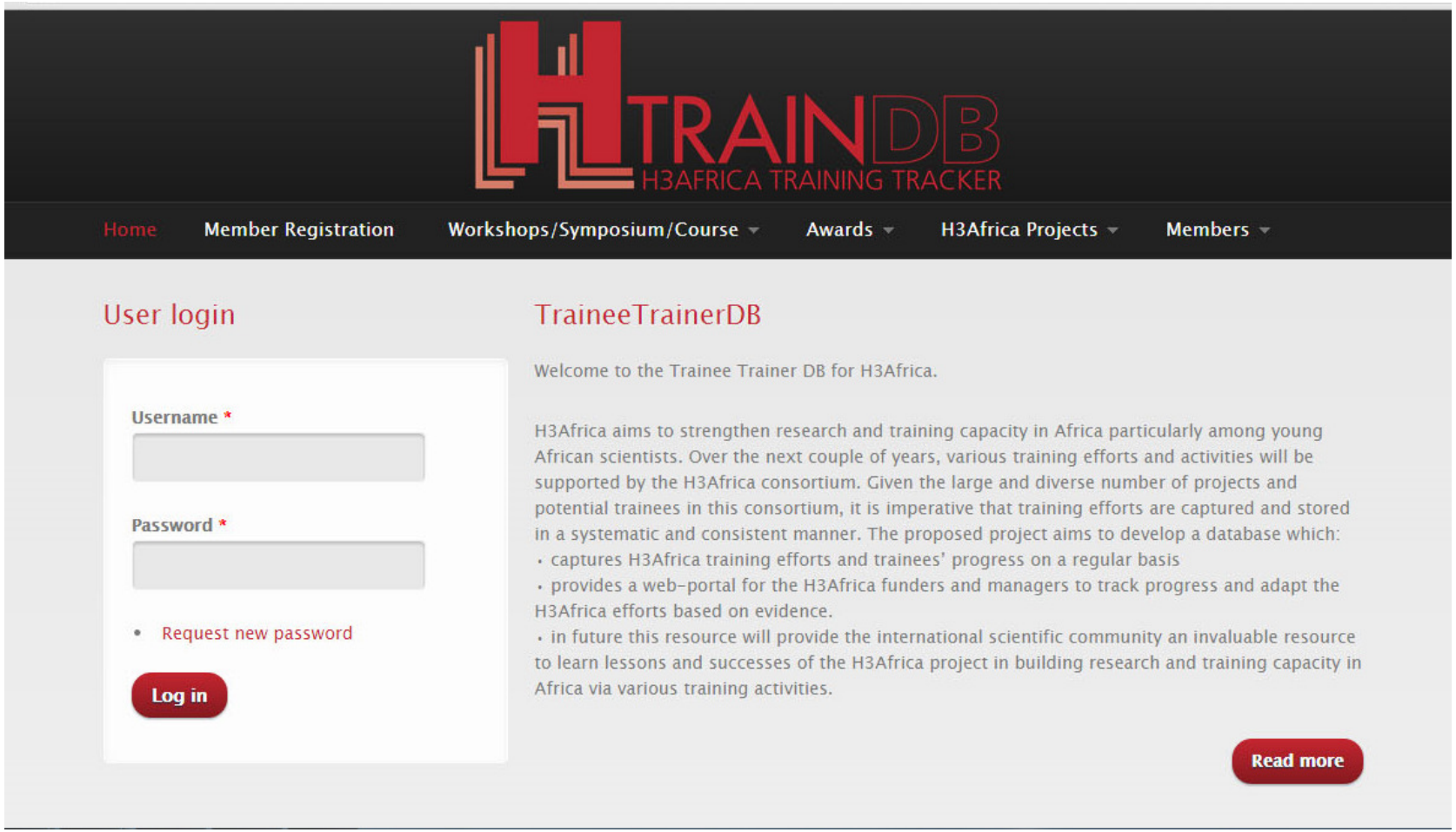
H3ABioNet
Pan African Bioinformatics Network for H3Africa

Progress and Challenges (OCVT)

- Access to CRFs only - phenotype data would be better.
  - We recommend standardization of CRF design (use PhenX)
  - Ambiguous questions in CRFs difficult to harmonize
- Previously no access to experimental information
- Need to work with Biobanks/Biorepository for their metadata

H3ABioNet
Pan African Bioinformatics Network for H3Africa

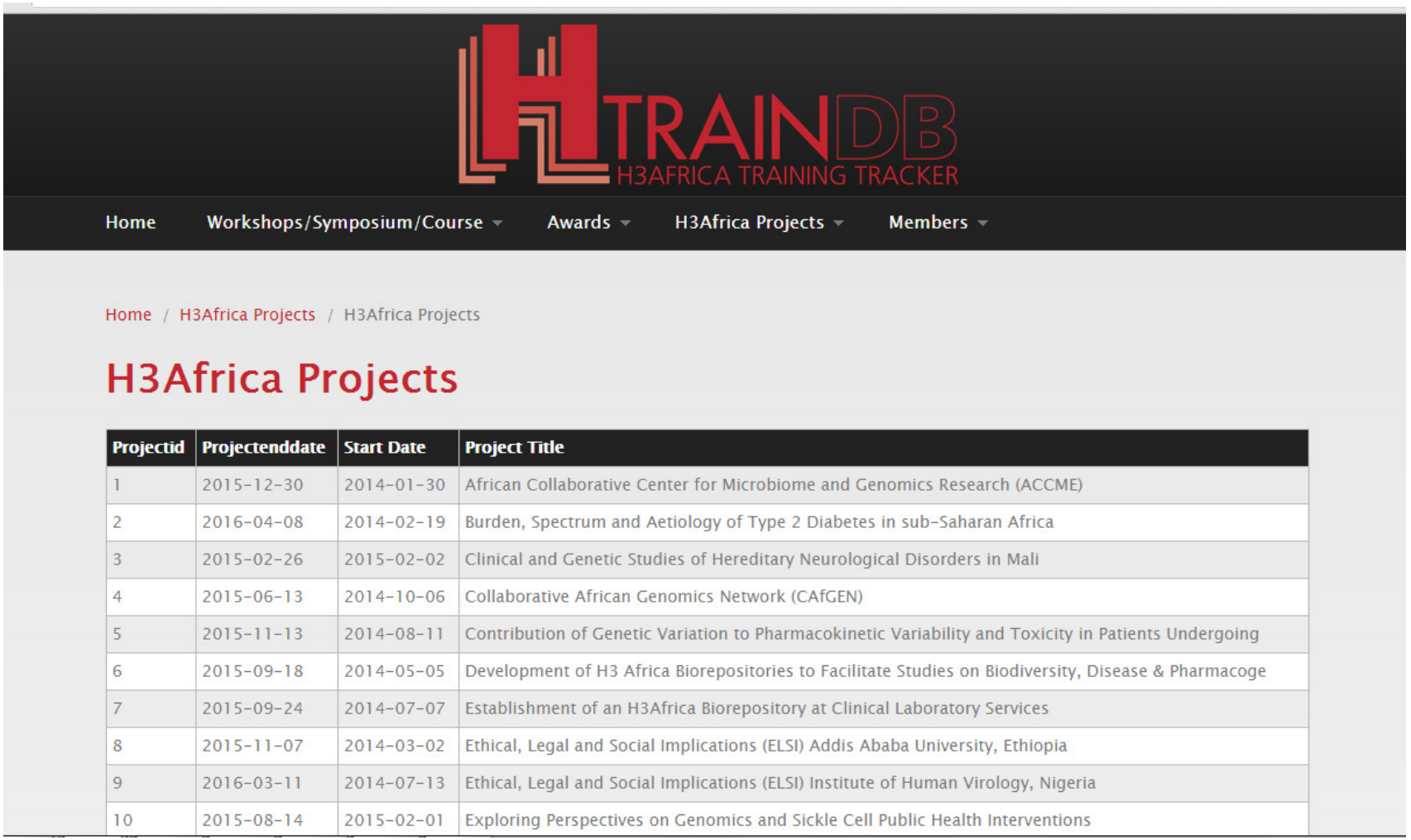
Trainer/trainee DB –HTrainDB (1)



Zahra Mungloo- Dilmohamud
Shakuntala Baichoo –University of Mauritius

H3ABioNet
Pan African Bioinformatics Network for H3Africa

Trainer/trainee DB –HTrainDB (2)



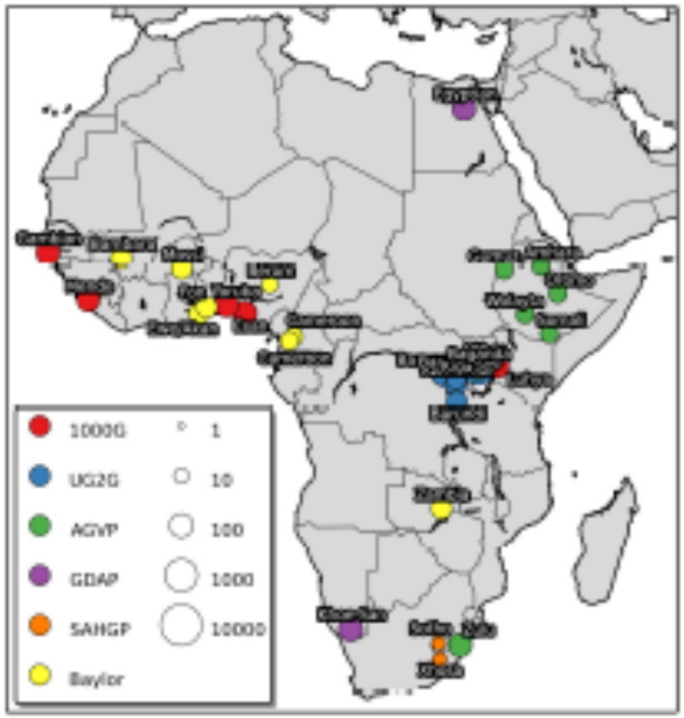
Projectid	Projectenddate	Start Date	Project Title
1	2015-12-30	2014-01-30	African Collaborative Center for Microbiome and Genomics Research (ACCME)
2	2016-04-08	2014-02-19	Burden, Spectrum and Aetiology of Type 2 Diabetes in sub-Saharan Africa
3	2015-02-26	2015-02-02	Clinical and Genetic Studies of Hereditary Neurological Disorders in Mali
4	2015-06-13	2014-10-06	Collaborative African Genomics Network (CAfGEN)
5	2015-11-13	2014-08-11	Contribution of Genetic Variation to Pharmacokinetic Variability and Toxicity in Patients Undergoing
6	2015-09-18	2014-05-05	Development of H3 Africa Biorepositories to Facilitate Studies on Biodiversity, Disease & Pharmacoge
7	2015-09-24	2014-07-07	Establishment of an H3Africa Biorepository at Clinical Laboratory Services
8	2015-11-07	2014-03-02	Ethical, Legal and Social Implications (ELSI) Addis Ababa University, Ethiopia
9	2016-03-11	2014-07-13	Ethical, Legal and Social Implications (ELSI) Institute of Human Virology, Nigeria
10	2015-08-14	2015-02-01	Exploring Perspectives on Genomics and Sickle Cell Public Health Interventions

H3ABioNet
Pan African Bioinformatics Network for H3Africa

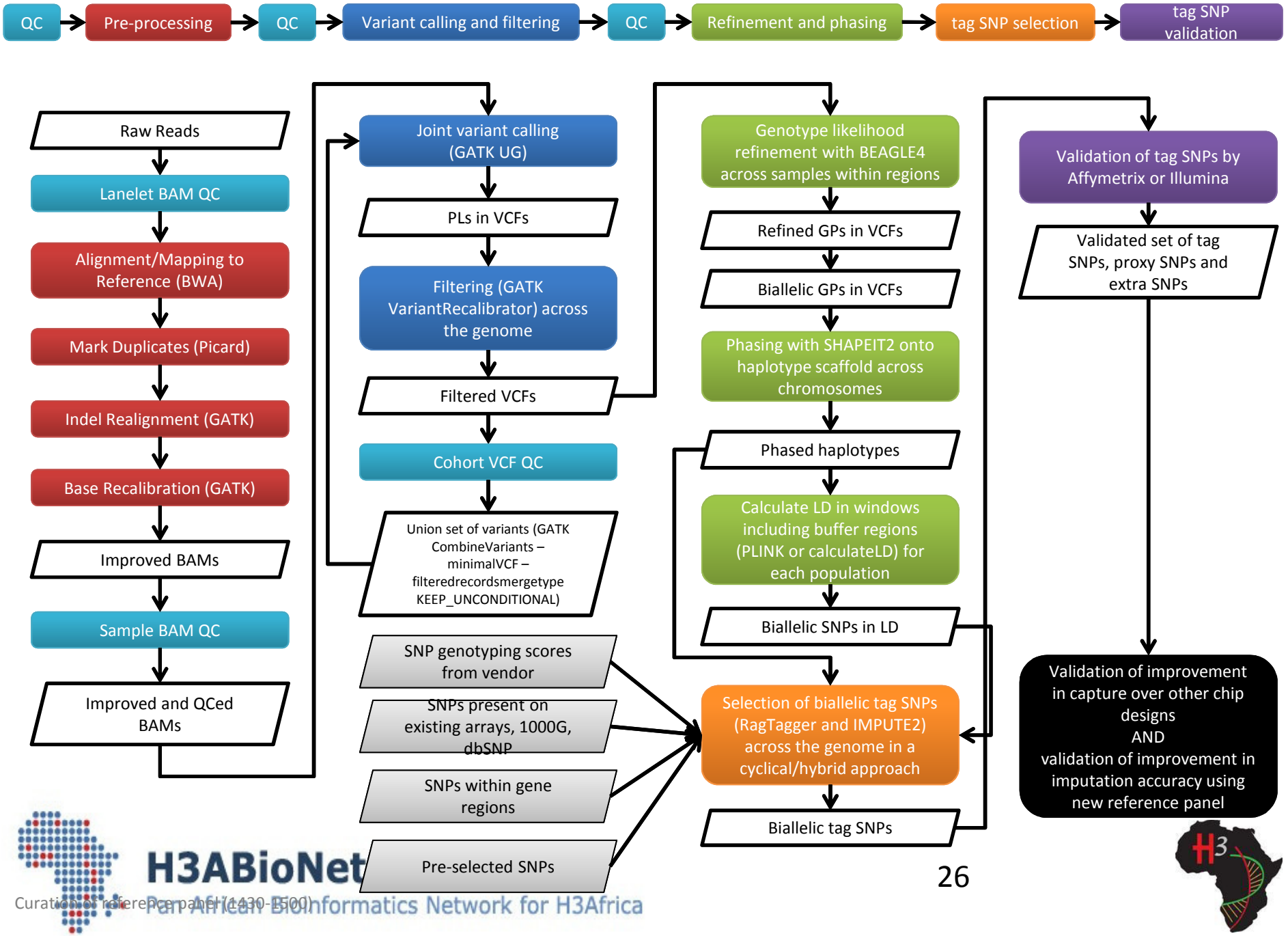
Zahra Mungloo- Dilmohamud
Shakuntala Baichoo –University of Mauritius

H3Africa custom chip design

- Chip design working group through Genome Analysis WG, in collaboration with Wits, Sanger and many other experts
- Computing: CHPC, NCSA
- Data ~ 3400 samples
- Timelines:
  - Depends on Baylor sequencing (when samples arrive)
  - Aim for list of SNPs in Aug/Sept

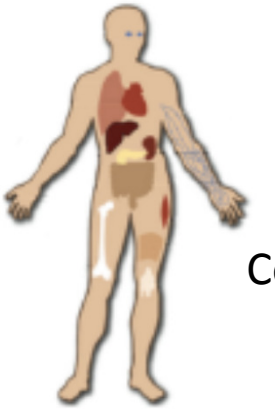


H3ABioNet
Pan African Bioinformatics Network for H3Africa


QC
Pre-processing
QC
Variant calling and filtering
QC
Refinement and phasing
tag SNP selection
tag SNP validation
Raw Reads
Lanelet BAM QC
Alignment/Mapping to Reference (BWA)
Mark Duplicates (Picard)
Indel Realignment (GATK)
Base Recalibration (GATK)
Improved BAMs
Sample BAM QC
Improved and QCed BAMs
Joint variant calling (GATK UG)
PLs in VCFs
Filtering (GATK VariantRecalibrator) across the genome
Filtered VCFs
Cohort VCF QC
Union set of variants (GATK CombineVariants – minimalVCF – filteredrecordsmergetype KEEP_UNCONDITIONAL)
SNP genotyping scores from vendor
SNPs present on existing arrays, 1000G, dbSNP
SNPs within gene regions
Pre-selected SNPs
Genotype likelihood refinement with BEAGLE4 across samples within regions
Refined GPs in VCFs
Biallelic GPs in VCFs
Phasing with SHAPEIT2 onto haplotype scaffold across chromosomes
Phased haplotypes
Calculate LD in windows including buffer regions (PLINK or calculateLD) for each population
Biallelic SNPs in LD
Selection of biallelic tag SNPs (RagTagger and IMPUTE2) across the genome in a cyclical/hybrid approach
Biallelic tag SNPs
Validation of tag SNPs by Affymetrix or Illumina
Validated set of tag SNPs, proxy SNPs and extra SNPs
Validation of improvement in capture over other chip designs
AND
validation of improvement in imputation accuracy using new reference panel

H3ABioNet service offerings

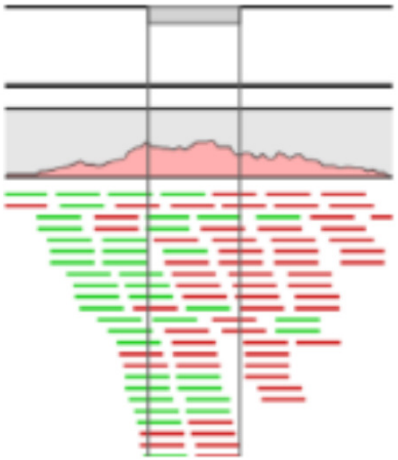
Participant cohorts



Collect samples

Extract DNA

Biorepository



Generate data

Data analysis and storage

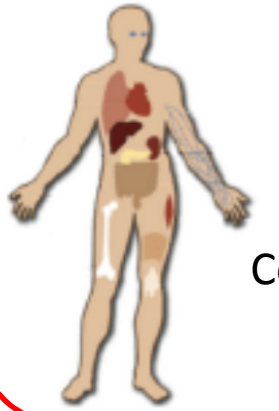
Data repository

Computing Infrastructure	Research/tool development	User support	Training

H3ABioNet
Pan African Bioinformatics Network for H3Africa

H3ABioNet service offerings

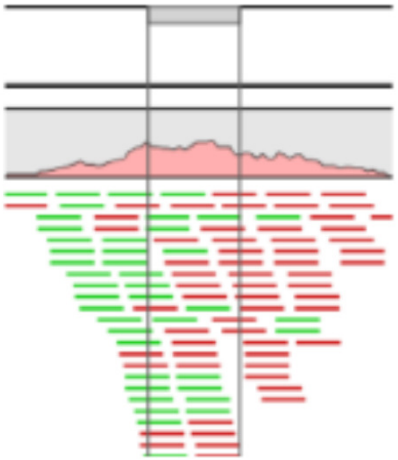
Participant cohorts



Collect samples

Extract DNA

Biorepository



Generate data

Data analysis and storage

Data repository

Computing Infrastructure **Research/tool development** **User support** **Training**

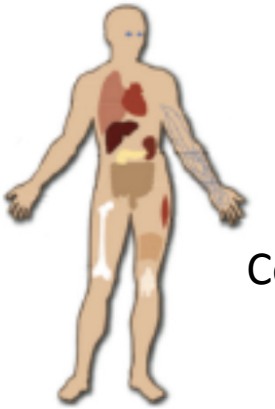
Participant databases, LIMS, data integration

Participant recruitment database

H3ABioNet
Pan African Bioinformatics Network for H3Africa

H3ABioNet service offerings

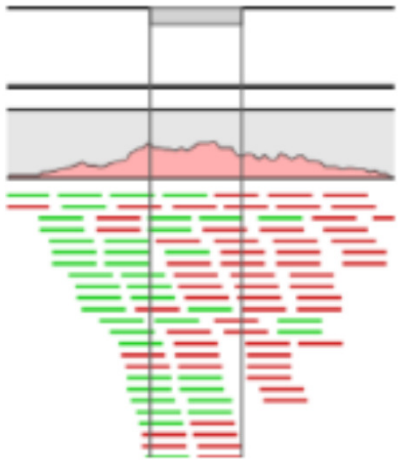
Participant cohorts



Collect samples

Extract DNA

Biorepository



Generate data

Data analysis and storage

Data repository

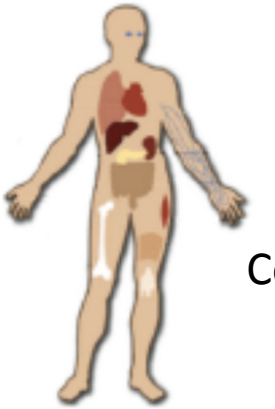
Computing Infrastructure **Research/tool development** **User support** **Training**

Chip design
Data management
Data transfer
Data analysis
Pipelines and SOPs
Galaxy

H3ABioNet
Pan African Bioinformatics Network for H3Africa

H3ABioNet service offerings

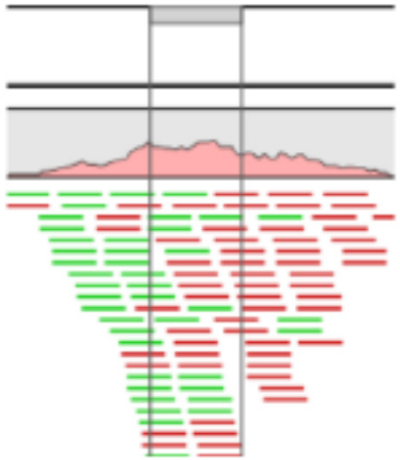
Participant cohorts



Collect samples

Extract DNA

Biorepository



Generate data

Data analysis and storage

Data repository

Computing Infrastructure | **Research/tool development** | **User support** | **Training**

Data archive
EGA submission
Ontologies
Data integration

H3ABioNet
Pan African Bioinformatics Network for H3Africa

Acknowledgements

Name	Institution	Country
Simani Gaseitsiwe	Botswana Harvard AIDS Institute Partnership	Botswana
Ahmed Mansour Alzohairy	Zagazig University	Egypt
James Brandful	NMIMR	Ghana
Ellis Owusu-Dabo	KNUST	Ghana
Daniel Masiga	ICIPE	Kenya
Dean Everett	Malawi-Liverpool Wellcome Trust Clinical research Programme	Malawi
Seydou Doumbia	University of Bamako	Mali
Yasmina Jaufeerally Fakim	SANBio	Mauritius
Hassan Ghazal	University Mohammed First	Morocco
Azedine Ibrahimi	Faculte de Medecine et de Pharmacie de Rabat	Morocco
Ahmed Moussa	Abdelamlek Essaadi University, Tangier	Morocco
Fouzia Radouani	Pasteur Institute Casablanca	Morocco
Fouad Seghrouchni	Institut National d'Hygiène, Rabat	Morocco
Fatima Gaboun	National Institute of Agronomic Research, Rabat	Morocco
Khalid Sadki	Mohammed V University, Rabat	Morocco
Alami Raouf	Centre National de Transfusion Sanguine, Rabat	Morocco
Odile Ouwe Missi	CERMES	Niger
Ezekiel Adebiyi	Covenant University Bioinformatics Research	Nigeria
Oyekanmi Nash	NADBA	Nigeria
Nicky Mulder	University of Cape Town	South Africa
Judit Kumuthini	CPGR	South Africa
Nicki Tiffin	SANBI, University of the Western Cape	South Africa
Ozlem Tastan Bishop	Rhodes University	South Africa
Scott Hazelhurst	Wits University	South Africa
Fourie Joubert	University of Pretoria	South Africa
Hugh Patterton	University of the Free State	South Africa
Faisal Fadlelmola	Future University	Sudan
Sylvester Lyantagaye	University of Dar es Salaam (UDSM)	Tanzania
Nzovu Ulenga	MDH	Tanzania
Julie Makani	MUHAS	Tanzania
Alia Benkahla	Institute Pasteur of Tunis	Tunisia
Jonathan Kayondo	UVRI	Uganda
Victor Jongeneel	NCSA	USA
Win Hide	Harvard School of Public Health	USA

Funding: NIH Common Fund, NGHRI grant: U41HG006941

Project Manager: Sumir Panji


Chubb
ARMED RESPONSE
TEL
0861 021 911
AIMS
SOUTH AFRICA