H3ABioNet

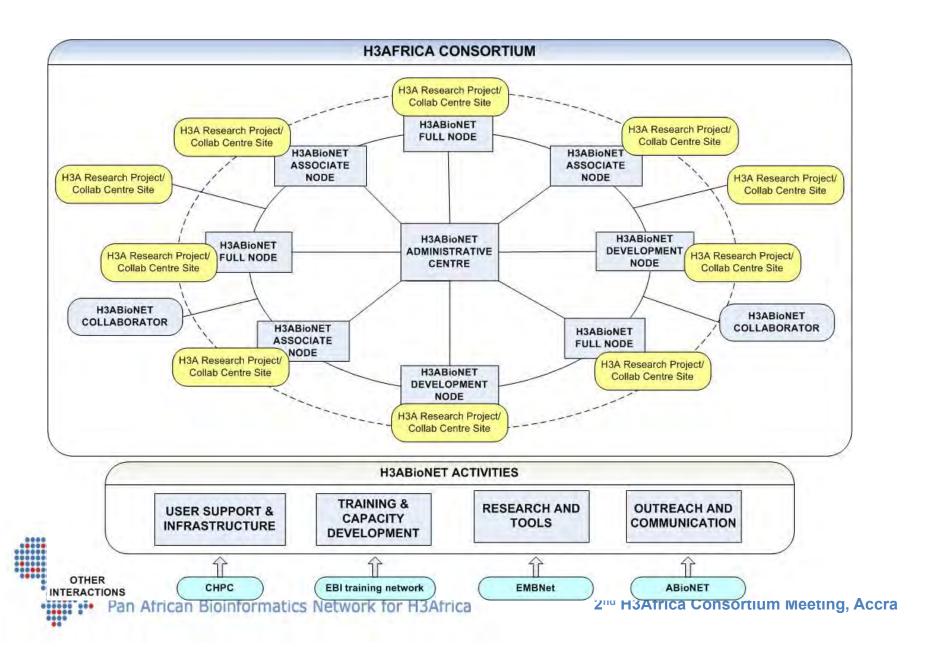
Pan African Bioinformatics Network for H3Africa



H3ABioNet Project Goal

To build H3ABioNet -- a sustainable
 African Bioinformatics Network -- to
 provide bioinformatics infrastructure and
 support for the H3Africa consortium.

Network structure



Partner institutions

Administrative hub at UCT

34 partner institutions, 32 in 15 African countries, 2 in USA





Partner institutions -projects



PI sites, other sites around Africa too



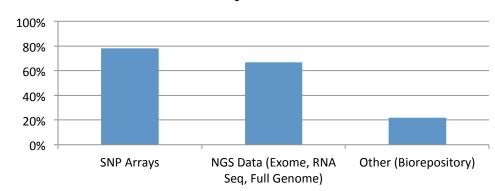
H3Africa's needs

- Large-scale data analysis for:
 - Genotyping by arrays
 - Next generation sequencing
 - GWAS
- Support for analysis:
 - General questions
 - Access to computing resources
 - Technical computing support
- Data access and visualization (public and new)
- Data storage, backup and transfer
- Data submission
- Training



H3Africa data expected (1)

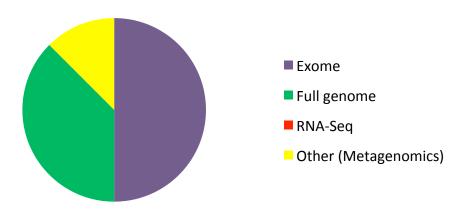
Types of Data To Be Generated H3Africa Projects



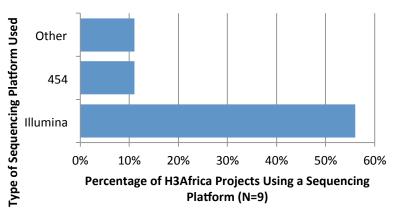
Types of Data Generated

Percentage of H3Afrca Projects (N=9)

Types of Sequence Data Generated by H3Africa Projects



Sequencing Platform Used



2nd H3Africa Consortium Meeting, Accra

H3Africa data expected (2)

GWAS arrays:

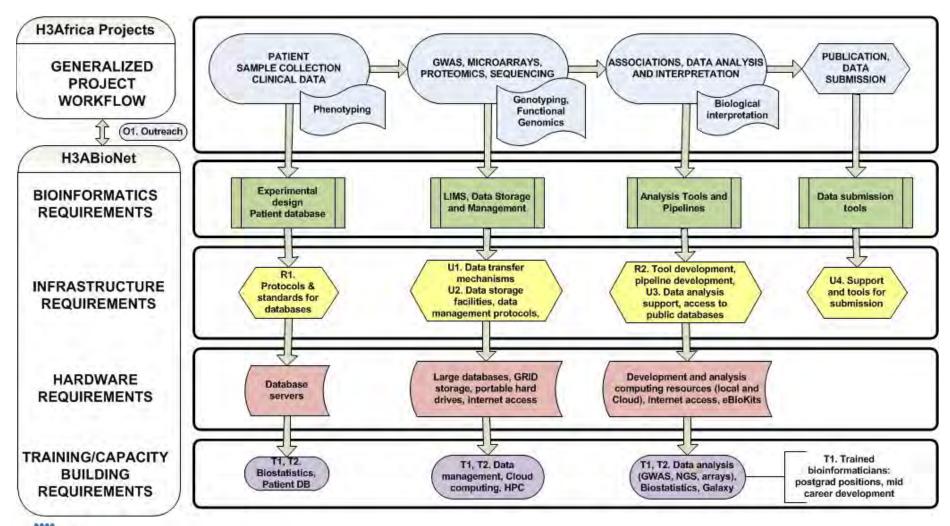
- 30Mb/cell file (per sample), 150Gb per 96 well plate processed.
- Samples planned ~33,380
- Storage required for processed data: 52TB

Sequencing NGS:

- Exome = illumina, ~45 depth, 90bp paired end, 17GB raw fastq,
 Full = Illumina, ~33 depth, 101 bp pared end, 250GB raw fastq
 (per sample)
- Samples planned in total 12, 580 (exome 11,330, full genome 1,250)
- Total storage for raw reads: 505TB, processed 2.5PB



Network activities





Major activities

Infrastructure Development

 Develop the hardware and software infrastructure capacity of African Bioinformatics laboratories in preparation for genomic scale studies.

User Support

 Provide user support to H3Africa researchers and Bioinformatics capacity development through a pool of specialists and supported tools & resources.

Node Assessment and Accreditation

 Provide an accreditation programme to help prepare and develop African Bioinformatics' laboratories in line with current international standards.

Research and Tool Development

 Provide research and development of Bioinformatics and genomics tools, standardised workflows and pipelines for H3Africa research projects.

Education and Training

 Develop Bioinformatics capacity of African scientists through co-ordinated education and training programmes.



Infrastructure development

- Technical support –programming, pipeline development, sys admin
- Access to HPC, Cloud, eBioKits
- Access to public data
- Standard operating procedures and guidelines for data analysis
- Tools/support for data submission
- Storage, movement & management of data

Infrastructure activities and plans (1)

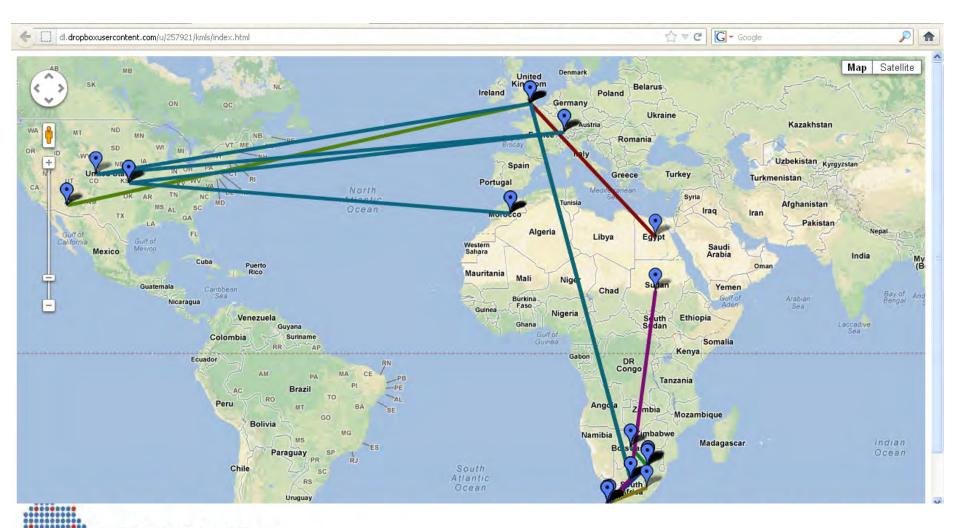
- Working group met 6 times
- Developed specifications of types of machines for different types of Bioinformatics workflows
- Negotiations for purchasing of ~ 20 high end servers capable of doing NGS analysis for various nodes
- One-on-one consultations with development and associate nodes for advice on hardware
- Purchasing of e-BioKits for training and processing at nodes

Infrastructure activities and plans (2)

- Conducting Iperf (internet bandwidth tests) between servers from all H3ABioNet nodes to determine:
 - True internet bandwidth across Africa
 - Best place to host mirror sites
 - How to move data around
- Researching data storage and processing options:
 - Getting Galaxy running at CHPC



Infrastructure activities and plans (2)



Infrastructure activities and plans (2)

- Conducting Iperf (internet bandwidth tests) between servers from all H3ABioNet nodes to determine:
 - True internet bandwidth across Africa
 - Best place to host mirror sites
 - How to move data around
- Researching data storage and processing options:
 - Getting Galaxy running at CHPC

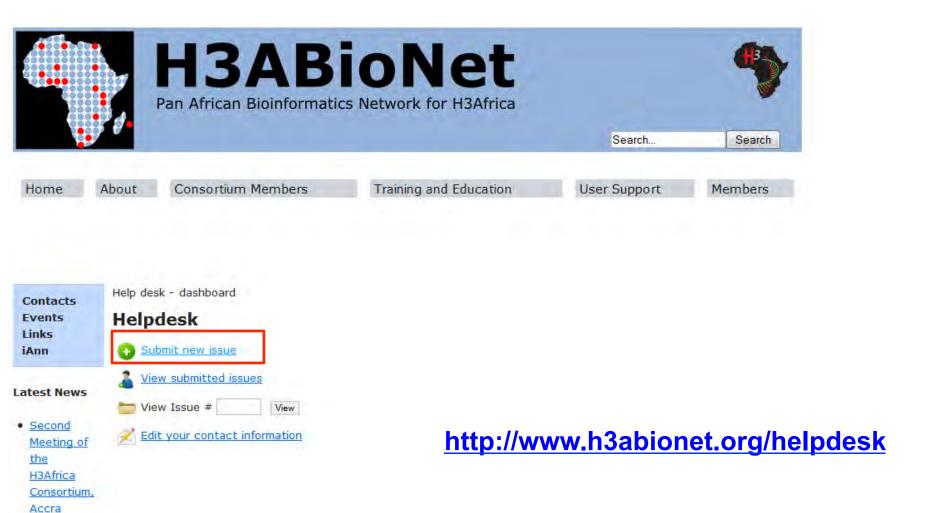


User support

- H3Africa research projects provide support on:
 - Small bioinformatics queries (scripts etc)
 - Data management
 - Large-scale analysis
 - Access to computing infrastructure (data & hardware)
 - Submission to EGA
- Support provided through:
 - Helpdesk
 - Training

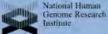


H3ABioNet help desk

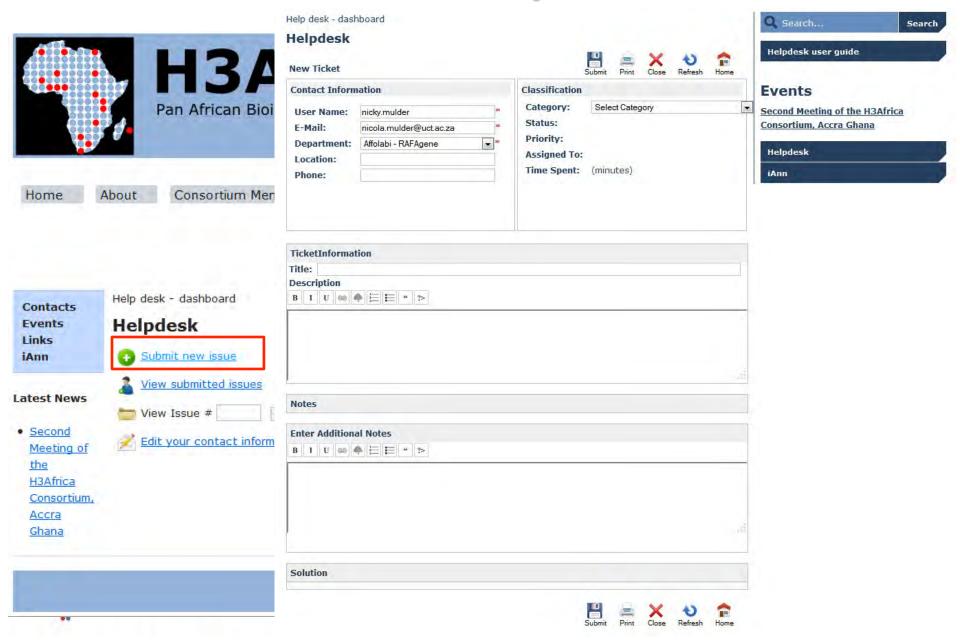




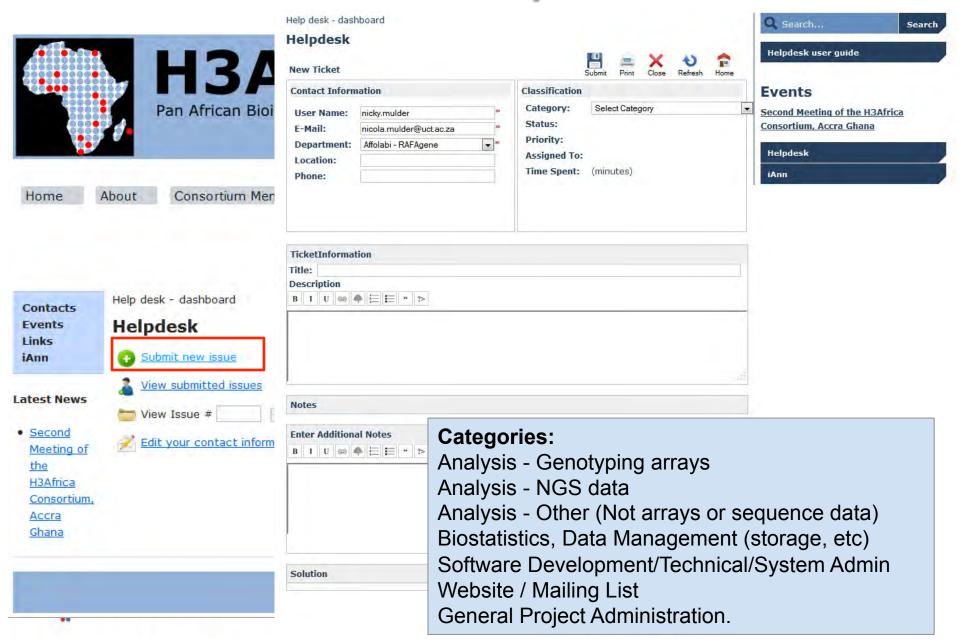
Ghana



H3ABioNet help desk



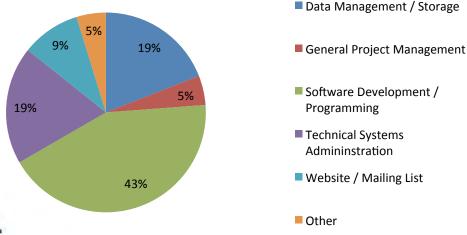
H3ABioNet help desk



Help desk continued

- Help desk system has been tested
- Created documentation & technical specs
- Created pool of 20 people to answer queries

Personnel Contributing to H3ABioNet Helpdesk Categories





Node assessment & accreditation

- Assessing internet capacity at each node
- Node assessment exercises
- Set of workflows with simulated data covering:
 - Analysis of NGS data (exome and full genome)
 - Variant calling
 - Genotyping array data analysis
 - GWAS analysis



Node assessment & accreditation

- Accreditation board established with international experts:
 - Brad Chapman, Harvard University School of Public Health, USA
 - Noah Zaitlen, University of California San Francisco, USA
 - Fran Lewitter and George Bell, Whitehead Institute, USA
- Getting datasets together (simulated)
- Nodes can choose when they are ready for accreditation

Research and tools

- Working group met 5 times
- Aim to facilitate collaborative projects
- Working group has established project and collaboration documents
- Nine collaborative projects set up so far
- Some overlap with the User Support and Infrastructure working groups

Research and tools

- Data management and storage
 - Patient databasing protocols
 - BioMart for genotyping data
 - Grid-based tool for data storage
- Data analysis tools
 - Galaxy for NGS and Genotyping analysis
 - Genome assembly pipeline
 - Functional SNP calling pipeline
 - Admixture mapping and network tools
 - Data visualization tools
 - Recombination tools
 - Structural SNP analysis tool
- Joint research projects



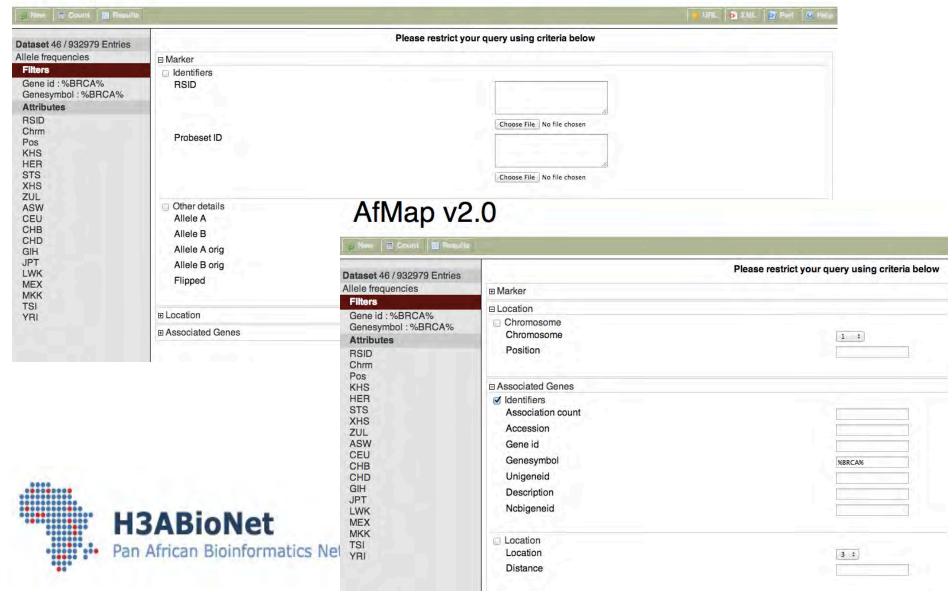
BioMart for genotyping data

- BioMart –suite of tools for integration of data in RDBMS
- Widely use for public databases, e.g. Ensembl, InterPro
- Has web front-end, BioMart library for R or import into Galaxy
- Have developed BioMart for local genotype data and interface for searching
- Can develop BioMart for public data
- Can query across BioMarts
- Can protect data in BioMart



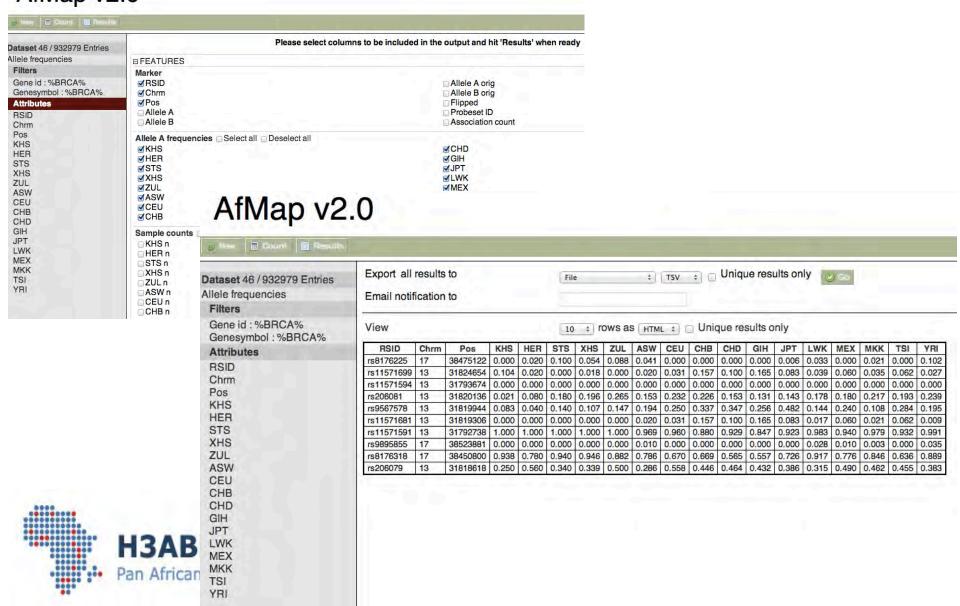
Example of BioMart interface 1

AfMap v2.0



Example of BioMart interface 2

AfMap v2.0



Example of BioMart interface 3



YRI

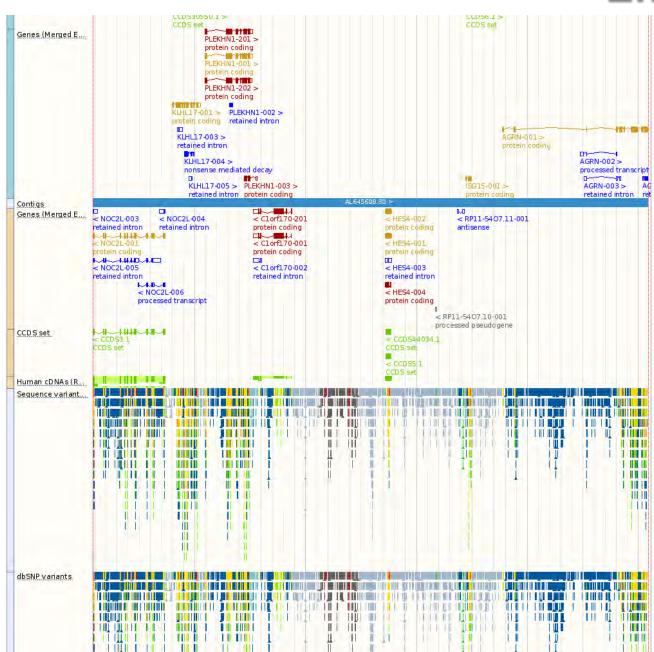
Galaxy for NGS and arrays

- Galaxy –web-based framework for building workflows for data analysis
- Contains modules for searching and extraction of human data and many NGS analyses
- Can plugin in own modules
- Workflows and histories can be customized, saved and shared
- Can export data to visualization tools
- Different levels of protection
- Can integrate with HPC solutions

Galaxy example

Galaxy		Ana	lyze Data V	Vorkflow Share	ed Da	ta -	Visu	alization Help User	Using 451,8 Mb
Tools ©	chr1	881626	881627	rs2272757	0	+	A/G	NOC2L	History •
PopGen	chr1	909325	909326	rs61573829	0	+	T/C	PLEKHN1	E .
Get Data	chr1	949653	949654	rs8997	0	+	A/G	ISG15	□ □ □
Lift-Over	chr1	981930	981931	rs2465128	0	+	A/G	AGRN	Unnamed history 0 bytes
Text Manipulation	chr1	982993	982994	rs10267	0	+	T/C	AGRN	1: AFR-NON-E5M2O10- ◎ Ø %
Filter and Sort	chr1	1118274	1118275	rs61733845	0	+	T/C	TTLL10	1: AFR-NON-E5M2O10- © 0 X unknown.bed
Join, Subtract and Group	chr1	1254840	1254841	rs10907179	0	+	C/G	CPSF3L	2,077 regions
CONTRACTOR	chr1	1290407	1290408	rs147996767	0	+	G/A	MXRA8	format: interval, database: hg19
Convert Formats	chr1	1354593	1354594	rs3766165	0	+	G/A	ANKRD65	Info: uploaded interval file
Extract Features	chr1	1635021	1635022	rs1136978	0	+	G/A	CDK11A	
Fetch Sequences	chr1	1684471	1684472	rs7407	0	+	T/C	NADK	view in <u>GeneTrack</u>
Operate on Genomic Intervals	chr1	1896332	1896333	rs61734005	0	+	A/G	KIAA1751	display at Ensemb Current
Statistics	chr1	2704111	2704112	rs56328393	0	+	C/T	TTC34	display at RViewer main
Graph/Display Data	chr1	3301720	3301721	rs2282198	0	+	T/C	PRDM16	1.Chrom 2.Start 3.End 4 5
Regional Variation	chr1	3394440	3394441	rs17854723	0	+	T/C	ARHGEF16	chr1 881626 881627 rs2272757 0
Multiple regression	chr1	3669200	3669201	rs10797410	0	+	A/G	CCDC27	chr1 909325 909326 rs61573829 0
Multivariate Analysis	chr1	3697699	3697700	rs6694319	0	+	A/G	LRRC47	chr1 949653 949654 rs8997 0
Evolution	chr1	6271124	6271125	rs1052598	0	+	G/A	RNF207	
NGS: RNA Analysis	chr1	6508711	6508712	rs77656032	0	+	C/T	ESPN	chr1 981930 981931 rs2465128 0 4
NGS: SAM Tools	chr1	6529442	6529443	rs730600	0	+	G/A	PLEKHG5	chr1 982993 982994 rs10267 0 4
SNP/WGA: Data; Filters	chr1	9005974	9005975	rs34649225	0	+	A/G	CA6	chr1 1118274 1118275 rs61733845 0 4
SNP/WGA: QC; LD; Plots	chr1	10218438	10218439	rs2273299	0	+	C/T	UBE4B	4 III
SNP/WGA: Statistical Models	chr1	10511543	10511544	rs628462	0	+	T/C	APITD1-CORT	
Human Genome Variation	chr1	10511543	10511544	rs628462	0	+	T/C	CORT	
VCF Tools	chr1	11087523	11087524	rs1782455	0	+	G/A	MASP2	
VCI 10013	chr1	11103407	11103408	rs7536030	0	+	A/G	MASP2	
Workflows	chr1	11132216	11132217	rs1130683	0	+	G/A	EXOSC10	
All workflows	chr1	11205057	11205058	rs1057079	0	+	C/T	MTOR	
	chr1	11301713	11301714	rs1135172	0	+	A/G	MTOR	
	chr1	11561936	11561937	rs12144924	0	+	C/T	PTCHD2	
	chr1	11579818	11579819	rs61739687	0	+	T/C	PTCHD2	
	chr1	11854456	11854457	rs4846051	0	+	G/A	MTHER	

Ensembl view

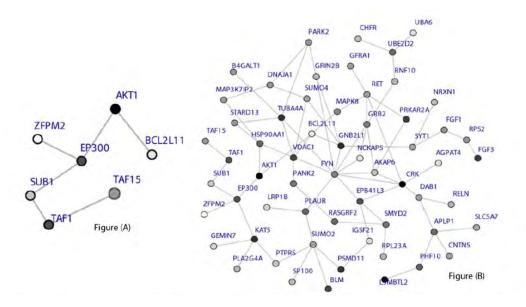


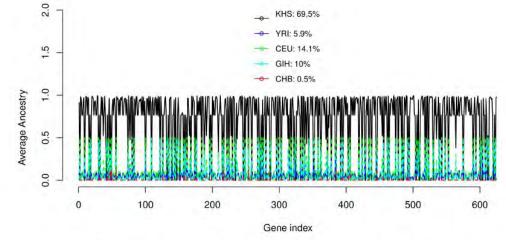
11 11

Admixture mapping tools

- Most tools were developed for 2-way admixture, although some can handle 4-way
- Some local populations are at least 5-way admixed
- Developing new tools for multi-way admixture mapping
- Developing tools for network analysis of admixed populations

GWAS network tools





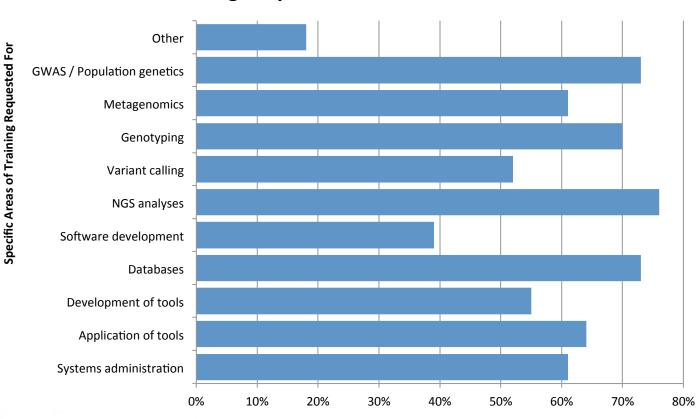


Education and training

- Working group has met 7 times
- Researching current programs:
 - Collected Bioinformatics curriculums from different Universities in Africa –very few!
- Developing application and selection procedures for courses and internships
- Developing training courses

Training –H3ABioNet

Training Requests of H3ABioNet Nodes



Percentage of Respondant Requesting for Training (N=33)



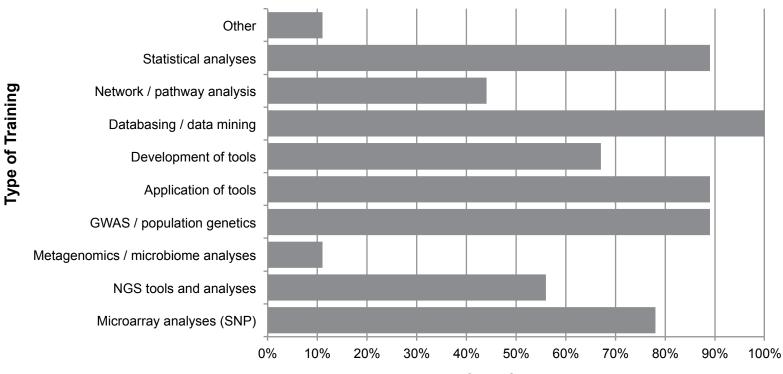
H3ABioNet training plan

- Train-the trainer
- Graduate training MSc/PhD/Postdocs
 - Shared course curriculum
 - Co-supervision across nodes
 - Project approval by advisory board
 - Specialised courses
- Technical staff
 - Computing/sys admin courses
 - HPC, Cloud, data management
- Internships with external partners



Training –H3Africa

Types of Training Required from H3ABioNet



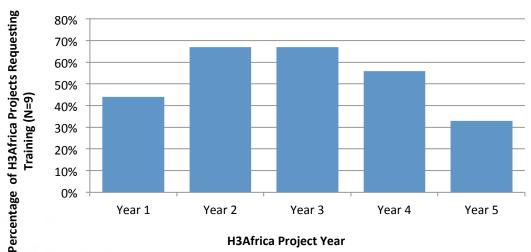
Percentage of H3Africa Projects (N=9)



H3Africa training

- Specialised courses (NGS, genotyping data analysis, pop genetics, databasing)
- Internships for H3-researchers to visit bioinformatics nodes

Estimated Year H3Africa Projects Require Bioinformatics Training





First training course

- NIH grants management workshop 27th 28th May, UCT
- Anticipated number 25 30
- Topics covered:
 - NIH and recipient roles and responsibilities, roles and responsibilities of the grantee and consortium sites, understanding the Notice of Award, award budgets, costs and allowable expenditures, award Delays, Progress Reports/Completing the PHS 2590, Consortium/ Subcontract Policy requirements, Payment Management System training, Federal Financial Report reporting, Post Award Actions

Second training course

- Pretoria for systems administrators, developers
 3rd 14th June
- Anticipated number 25
- Topics covered:
 - HPC, Grid Computing, Linux admin and set up of purchased servers, backups, Cloud computing, Data management and visualisation and Data Security
- Students will learn to install and manage machines they purchase

Third training course

- Bioinformatics training of H3ABioNet trainers to be held in Central location –ICIPE, Nairobi, Kenya
- 25 participants from H3ABioNet nodes
- Criteria must be staff, have some bioinformatics experience, involved in teaching and provide support and training for their nodes when they return
- Course from 8th to 26th July and covers:
 - Unix / Linux, biostatistics, programming, data management,
 NGS pipelines, GWAS and genotyping.

Fourth training course

- eBioKit workshop ICIPE 29th July 1st August
- Anticipated number 25
- Some will be from bioinformatics course
- Topics covered:
 - UNIX, Biological databases/ENSEMBL NGS (Galaxy mainly), phylogenetics or wEMBOSS

Training courses

- Technical course will be webcast to some sites using USTREAM
- Train-the trainer course will have live Vidyo streaming to 2 sites Tunisia and Nigeria
- In total >150 people all over Africa anticipated to have access to training
- All lectures will be recorded
- This will enable us to reach a wider audience

Knowledge transfer program -CPGR

- KTP brings together experts from around the world to Africa to train upcoming researchers on specialized projects
- Transfer of knowledge from higher concentration to lower concentration for longer period of time e.g. 3-6 months
- Faster turnaround time of effective transfer of knowledge
- The programme cuts running costs as more people are trained within their local environment
- KTP allows training within the resource deprived environment and provides a wider experience for visiting experts
- You can take part in the programme as an expert, associate or host of a project.
- Online registration will be available at <u>www.ktp.cpgr.org.za</u>



Future training plan

- Work with H3Africa projects to set up suitable training plan for years 2 and 3
- Ensure courses reach wide audience
- Coordinate with other training programs –Wellcome Trust, EMBO, etc.
- Work with GOBLET on training objectives, quality etc.
- Set up internship program
- Set up co-supervision program
- Work towards degree programs at African institutions



Communication

- Meetings
 - Kick-off meeting Nov 2012
 - Meeting in Casablanca March 2013
 - WG meetings fortnightly, MC monthly
- Web portal:
 - Members info and expertise
 - Documents
 - Tools and resources, etc.
- Mailing lists, biweekly bulletin
- Internal database –info on nodes



Kick-off meeting









Africa

H3ABioNet website -public



H3ABioNet website -private



« < May 2013 > »						> »	A comprehensive, informative and interactive map of	Ultimatel III Server Server
s	М	T	w	T	F	S	members accross Africa and internationally	National Human Genome
8	29	30	1	2	3	4		Research Institute
5	6	7	8	9	10	11	0	NIH National Institutes Of Health
2	13	14	15	16	17	18	The state of the s	Of Health
9	20	21	22	23	24	25		This project is funded by NIH grant
6	27	28	29	30	31	1		number U41HG006941





Working Groups





Consortium Node status

Training & Education

Meetings

User Support Research & Tools **Education and Training**

Links Calendar

Contacts

Forum

Log out

Members

Terms of Reference

Members

Name Institute Nash Oyekanmi NABDA Shaun Aron Wits James Brandful **NMIMR** Segun Fatumo NABDA Judit Kumuthini **CPGR** Ahmed Moussa AEU Shakuntala Baichoo **UoM** Nicky Mulder UCT Jonathan Kayondo UVRI Faisal M. Fadlelmola FUS Victor Osamor CU Jean-Baka Domelevo Entfellner SANBI Sylvester Leonard Lyantagaye UDSM Odile Ouwe Missi Oukem CERMES Owusu-Dabo Ellis KCCR Sampson Salifu Pandam KCCR PostDoc C / O Dean Everett MLWCTR Winston Hide **HSPH** Shannan Ho Sui **HSPH** Ahmed Mansour Mohamed Mansour Elzohairy ZU Samar Kassim ZU Anne Fischer ICIPE Sumir Panji UCT

Minutes

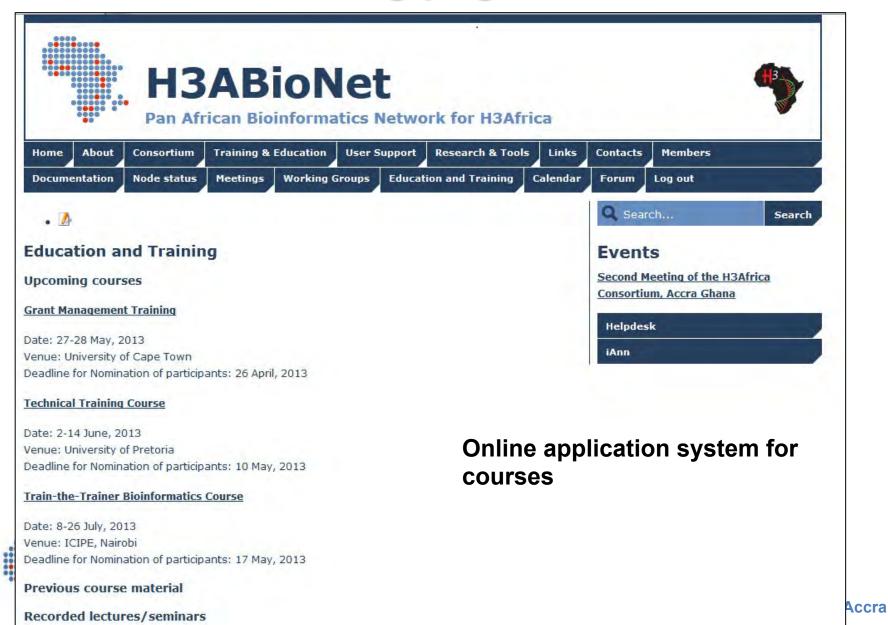
- Meeting #1 (2013-01-29)
- Meeting #2 (2013-02-12)
- Meeting #3 (2013-02-26)
- Meeting #4 (2013-03-13)
- Meeting #5 (2013-04-09) Meeting #6 (2013-04-23)

Search... Search **Events** Second Meeting of the H3Africa Consortium, Accra Ghana Helpdesk iAnn

Working group page

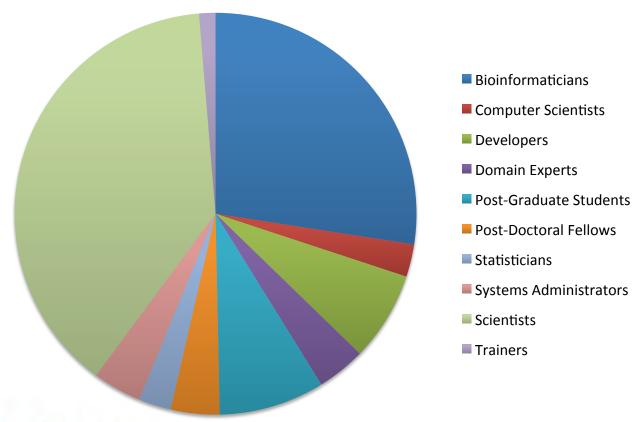
nsortium Meeting, Accra

Training page



H3ABioNet funded personnel

Composition of H3ABioNet Members (N=153)



Internal database











A

Logged in as H3A_UCT (<u>Logout</u>)
Help Change Password

You are here: NODE MEMBERS :: Dashboard personnel

NEWS & UPDATES

NODE MEMBER DASHBOARD - tables and action shortcuts

Click on the heading to go straight to the corresponding table or else choose an action shortcut. (Table visible to all nodes shown by: 🚳)

NODES

NODE MEMBERS

H3AFRICA INVOLVEMENT

FUNDING, COLLABORATIONS & AFFILIATIONS

TEACHING & TRAINING

PUBLICATIONS & CONFERENCES

REPORTING PERIODS

NODE REPORTS

DETAILS:

- # ADD A NEW PERSON
- **Ø EDIT** Select... ▼ S DETAILS
- **LIST PERSONNEL RECORDS**
- Q SEARCH PERSONNEL DETAILS

QUALIFICATIONS:

- ADD A NEW QUALIFICATION
- **LIST QUALIFICATION RECORDS**
- Q SEARCH QUALIFICATION RECORDS

STATUS:

(Employment/Educational)

- # ADD A NEW RECORD
- **LIST STATUS RECORDS**
- Q SEARCH STATUS RECORDS
- To EDIT select records from LIST or SEARCH

RESOURCES:

- + ADD A NEW PERSONNEL RESOURCE
- Q SEARCH PERSONNEL RESOURCES
- To EDIT select records from LIST or SEARCH

PROJECT INVOLVEMENT:

- ADD A NEW RECORD
- **I** LIST PROJECT INVOLVEMENT
- Q SEARCH PROJECT

PROJECTS REFERENCE TABLE:

(Sub-table of PROJECT INVOLVEMENT)

- + ADD A NEW PROJECT
- **LIST PROJECTS**
- Q SEARCH PROJECTS

TRAVEL:

- ADD A NEW JOURNEY
- **∷** LIST JOURNEYS TAKEN
- Q SEARCH JOURNEYS TAKEN

SKILLS:



- E LIST SKILLS
- Q SEARCH SKILLS

Internal database



or SEARCH







Logged in as H3A_UCT (Logout)
Help Change Password



You are here: NODE MEMBERS :: Dashboard personnel **NEWS & UPDATES** NODE MEMBER DASHBOARD DATA COLLECTION Click on the heading to go straight to th NODES **DETAILS:** You are here: NODE MEMBER QUALIFICATIONS :: Add New Record NODE MEMBERS NODE MEMBER QUALIFICATIONS # ADD A NEW PERSON **NEWS & UPDATES H3AFRICA** INVOLVEMENT Click on table heading to return to action shortcuts **Ø EDIT** Select... ▼ S DETAILS LIST SEARCH NODES FUNDING, ADD NEW RECORD SHOW CURRENT RECORD DELETE CURRENT RECORD SHOW ALL **LIST PERSONNEL RECORDS** COLLABORATIONS & AFFILIATIONS NODE MEMBERS Add a new qualification for someone from the UCT node: Q SEARCH PERSONNEL DETAIL **TEACHING & H3AFRICA** TRAINING INVOLVEMENT Please Select ... ▼ Other... Node member PUBLICATIONS & FUNDING, Qualification • Please Select ... - Other ... CONFERENCES COLLABORATIONS & AFFILIATIONS → Other... Awarding Institution • Please Select. **PROJECT** REPORTING TEACHING & PERIODS Discipline(s) . Other... INVOLVEMENT: Please Select. TRAINING Molecular Toxicology **PUBLICATIONS &** + ADD A NEW RECORD CONFERENCES NODE REPORTS Start date (YYYY-MM-DD) LIST PROJECT INVOLVEMENT REPORTING **PERIODS** Date awarded/completed (YYYY-**Q** SEARCH PROJECT MM-DD) INVOLVEMENT NODE REPORTS To EDIT - select records from LI Save

Outreach

- Casablanca conference exhibition
- Represented at SASA meeting
- Coordination with other training programs
- Work with GOBLET –attended their TGAC workshop
- External mailing lists –interest groups
- Work more closely with H3Africa projects

Acknowledgements

Funding: NIH/NGHRI

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





Thanks: Sumir Panji –Project Manager



Questions to H3Africa

- How will you get your data from the generator to your lab?
- Are there any specific journal ToCs you want access to?
- What else do you want from us?
- How best can we engage with you?