



# Foreword



- **Foreword**
- **Education and Training**
- **Infrastructure**
- **Research**
- **User Support**
- **Computational Metagenomics Workshop**
- **H3ABioNet Advanced Systems Administration Workshop 2015**
- **Announcements**
- **Upcoming H3ABioNet Working group meeting schedule**

Welcome to the second newsletter for 2015. February and March have been busy with Custom Chip design meetings, running of the Advanced Systems Administration Workshop at the University of Pretoria from the 2nd to 11th Feb, and organization of the Introduction to Biostatistics for Genome Wide Association testing (Tunis, March), and Medical population genetics and GWAS for complex diseases (Cape Town, April). This newsletters reports on the Advanced Systems Administration Workshop and the Metagenomics workshop held at the end of last year at the University of Mauritius.

Another important event which took place in March was the ISCB Africa ASBCB Conference on Bioinformatics. The conference took place at the White Sands resort in Dar es Salaam, and was followed by two workshops at the University. H3ABioNet and Wellcome Trust provided travel fellowships to enable 31 students from around Africa to attend the conference. A report on the conference will be provided in the next newsletter.

H3ABioNet central has continued to work on a number of projects, including custom chip design, developing documentation for systems administrators, and getting the H3Africa archive set up and ready to accept the first submission. On the management side, we had a General Assembly conference call and have been accepting the first annual reports generated by NetCapDB, demonstrating the use of the data in the database. I hope this is encouraging nodes to ensure their data is complete and up to date.

I invite you now to continue reading to find out more about our working group activities and reports on training courses provided in this newsletter.

Prof. Nicky Mulder.



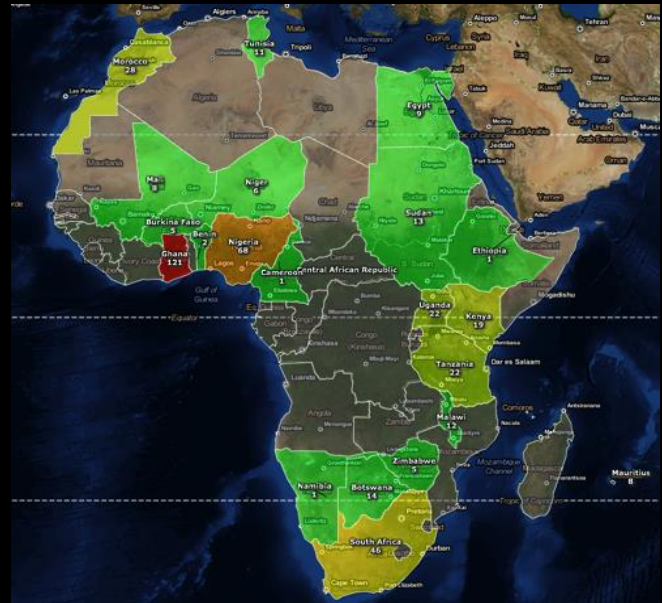
&amp;



## • Education and Training

The Education and Training working group is yet to formally meet for the new year, as the year kicked off with a number of planned workshops. The Introduction to Biostatistics for Genome Wide Association Testing workshop at the Institut Pasteur de Tunis in Tunisia has just been completed and we will report on the course feedback in the next issue. We have updated the terms of reference for the working group as well as the milestones for the upcoming period. Apart from the upcoming GWAS workshop and symposium at the African Institute Mathematics in South Africa, we currently have not planned any workshops for the next few months as we have decided to gauge the exact training needs of the H3Africa projects at the upcoming H3Africa consortium meeting in May 2015. In light of this we thought that we would take this opportunity to provide a brief review of the training that has been conducted by H3ABioNet since its inception.

Figure 1: Geographical distribution of the number of participants who have received training via the H3ABioNet workshops from May 2013 to March 2015



From May 2013 to March 2015 H3ABioNet has hosted or part funded 12 workshops ranging from two 4 week bioinformatics masters training programs, financial grants management, introductory and advanced systems administration, biostatistics, population genetics and GWAS which has provided tailored training to 305 H3ABioNet and H3Africa consortium members who are at various stages of their career development.

In addition to formal workshops, an internship programme has been established for H3Africa and H3ABioNet consortium members to spend a dedicated period of time at a host laboratory to acquire bioinformatics skills relevant to their research. To date there has been 8 H3ABioNet internship placements with students acquiring skills ranging from GWAS, metabolic networks, structural modeling and metagenomics to computational system administration and big data transfer mechanisms.



- **Infrastructure**

Computer network infrastructure is a very important part of achieving the goals of H3ABioNet and the H3A projects. The H3A projects will be generating large volumes of data and the ability to transmit data reliably and securely is vital if we want to ensure that African bioinformatics groups can participate in the analysis of these large data sets. The ISWG has two inter-related projects, Netmap and Globus Online to help monitor and improve our infrastructure, and we currently have a campaign to get these projects achieved. All nodes will have had communication from the ISWG. All PIs are asked to check the status of their Nodes. They should be running Netmap clients and have Globus Online end points.

The Netmap project has as its goal the objectives of mapping over time, the capacity of the network links between the H3ABioNet Nodes. Results can be used to lobby for additional resources, identifying bottlenecks, providing training and then recording how infrastructure changes. The Netmap Task Force has developed software that will be run at each node and at Central for monitoring. Each H3ABioNet Node will participate by running tests to a number of test servers across the continent. A server at Central will collate and publish the data so we'll be able to have a real-time map of network links between us to estimate network capacity, as well as have historical views. This will be particularly important for planning large data transfers. We have a number of servers operating and about 10 of our nodes are running the client software and reporting and hope to publish a consortium paper on this work.

The goal of the Globus Online Project is to provide capacity for all nodes to transfer large sets of data between themselves and partners outside of Africa. Globus is a system that developed out of a collaboration between the University of Chicago and Argonne National Labs to provide a user-friendly way of exploiting the gridftp protocol. A site or individual wanting to use Globus installs an "end point" — a software package that facilitates access to the Globus system. Users who have credentials at an end point can transfer data to other end points reliably and exploit the available bandwidth as much as possible. Last year Mohammed Alibi spent time as an H3ABioNet funded intern at the University of Illinois node where he received training and developed very thorough documentation and SOPs which are being rolled out across the network by the Systems Administration Task Force. We currently have 8 nodes up and running and a number of other nodes are busy with the installation. We have developed a monitoring facility using Nagios to monitor the status of the Netmap servers and clients as well as the Globus Online endpoints. If Nagios detects a problem, an email is sent to the relevant system administrator providing quick feedback of problems as they occur. This will allow H3ABioNet to demonstrate sustained capacity in terms of rolling out and maintaining infrastructure for the consortium.



- **Research**

Members were asked to assign themselves to the activities and tasks of the working group as soon as possible on a Google document prepared for that purpose, otherwise members will be nominated by the Chairs and the Network Manager. On March 6<sup>th</sup> 2015, Google hangouts was tested as a platform for conducting the seminar series and a mock seminar was delivered for the purpose of testing the platform. Most of the test participants were satisfied with the excellent features of hangouts and its user-friendly style. Early this year the central office sent out an H3ABioNet consortium survey form for all the H3ABioNet nodes to fill in their updated information including the month that each member of the node can provide a 10 to 15 minute seminar. As soon as all the survey forms have been submitted, an email will be sent asking them for a seminar title and abstract, then the seminar series will start using the Google hangouts platform. As the seminar series is one of the main and important milestones for RSWG for this year, we thought it would be crucial to establish a seminar series coordination task force of small group to track the seminar activities and report back to the chairs and members of RSWG on a regular basis.

An H3ABioNet Twitter Party was proposed. Twitter parties typically last 1-2 hours and are a wonderful way for people to connect and discuss a topic of choice. Most twitter parties have an expert panelist and party host to keep the party on topic. For H3ABioNet Twitter Parties, we could make use of events like while students attending workshops to get their feedback about the course or if there is a need to get a quick response from members with regard to urgent issues concerning the network or discussing current bioinformatics themes.

A Facebook page was created and members of the social media task force requested the H3ABioNet members to provide them with event photos and news to be posted at the page:

<https://www.facebook.com/pages/H3ABioNet/1388050288174486>

The issue of how to control unwanted postings on the Facebook page was raised and it was agreed that the moderator of the page has to find a means of stopping such activity and report back to the members of the WG.

It was agreed that at each RSWG meeting, 10 minutes will be allocated for progress reports on the student collaborative projects presented by the research project leader or group representative (5 minute for each project). There was a discussion that the Sorghum project is not progressing compared to the fibroid project. It was agreed that the RSWG chairs, Network Manager and the former RSWG Chair will hold a side meeting with the key personnel of the Sorghum project and identify the reasons for the delay in executing and pushing forward the project objectives as explained in the draft concept paper.



## User Support

- **User Support**

The period's highlight is the inaugural monthly working group meeting for 2015. The newly constituted team met on Friday the 27<sup>th</sup> February, during which draft terms of reference (TORs), milestones, membership and work strategies for the group were reviewed. The attendance was low, but is expected to pick up after finalization of the on-going recruitment of node representatives from the various consortium partners.

The USWG's main focus remains the helpdesk, whose profile is set to be further promoted during the course of the year. Towards this, some propositions to spur this milestone have been lined up, among which are:

- Revamped help desk outlook (-where available and non available expertise is clearly listed; Subject categories are more specific, etc)
- Creation of a Task force to manage the Help desk including the reviewing of web analytics—invitations for volunteers coming out as soon as the WG composition been finalized.
- Increased Help desk promotion through traditional and new outreach methods (e.g. social media) so members with ideas welcome.

The proposed terms of reference are not much different from those used by the out-going team, though it is clear that active node representation is going to be critical for the successful realization of set group deliverables. So in view of this, we are aiming for the recruitment of at least two representatives from each participating node where possible, and appeal upfront to the prospective volunteers to give it their very best.

We are in the final stages of our workgroup's re-launch and a list of planned activities, TORs and group membership will be out in due course.

Dr. Pandam Salifu.

Dr. Jonathan Kayando.



## • Computational Metagenomics Workshop

# Computational Metagenomics Workshop University of Mauritius

The Computational Meta-genomics Workshop held from 1<sup>st</sup> to 5<sup>th</sup> Dec 2014 at the University of Mauritius provided a weeklong series of training seminars, practical demonstrations, hands-on tutorials and research talks on the use of software tools to analyze and interpret meta-genomics data. The scope of the workshop was quite broad and offered an overview of cutting-edge fundamental theory and practical applications of meta-genomics.

Conventional genomic research on microorganisms determines the DNA sequences of individual microbes by examining cultivated strains. In meta-genomics, DNA sequence information is extracted from entire microbial communities *in situ*. Meta-genomic approaches use this bulk data to infer underlying properties of both individual microbes and microbial communities as a whole.

The participants for the workshop came from 11 African countries including Mauritius, Botswana, Egypt, Ethiopia, Kenya, Malawi, Morocco, South Africa, Sudan, Uganda and Zimbabwe. The workshop was supported by the University of Mauritius, H3ABioNet, Linnaeus University (Sweden), CERTH (Thessalonica, Greece) and the University of Cyprus.

The instructors were:

- Prof. Christos Ouzounis from CERTH Thessalonica, Greece
- Dr. Daniel Lundin from the Linnaeus University, Sweden.
- Dr. Gerrit Botha from the University of Cape Town (H3ABioNet), South Africa
- Dr. Shakuntala Baichoo from the University of Mauritius, Mauritius
- Dr. Vasilis Promponas from the University of Cyprus Nicosia, Cyprus

The contents covered were:

### Day 1: Fundamentals for computational metagenomics

- Fundamentals of Biology: genes, genomes, habitats
- Fundamentals of Computing: data, databases, algorithms
- Fundamentals of Bioinformatics: computation, biology, computational biology
- Introduction to Linux

### Day 2: Structure, function & evolution of biological macromolecules

- Principles of molecular structure, DNA/RNA/proteins, -omics
- Principles of sequence alignment: sequence comparison, database searches
- Principles of multiple sequence alignment: phylogenetics, phylogenomics
- Sequence comparison, tricks of the trade

### Day 3: Concepts and motivations for meta-omics

- Metagenomics: genomes, pangenomes, paleogenomes
- Environmental genomics, metagenomics
- Environmental genomics, metatranscriptomics (case studies)
- LAST+MEGAN: A quick way of getting a first look



- **Computational Metagenomics Workshop**

## Computational Metagenomics Workshop University of Mauritius

### Day 4: Tools and resources for metagenomics

- Metagenomics Platforms I: EBI's metagenomics portal, web resources
- Metagenomics Platforms II: JGI's IMG/M, Microbial Earth Project (MEP)
- Metagenomics: Genomic/Functional Encyclopedia of Bacteria & Archaea
- Data analysis of 16S rRNA amplicons with QIIME, uparse & phyloseq

### Day 5: Thematic areas & ongoing work

- Thematic Areas: microbial ecology, bioprospecting, 16S rRNA
- Thematic Areas: human microbiome, biosecurity, highlights/milestones
- Thematic Areas: extreme environments, exobiology
- Practical: Analysis of reduced or simulated datasets, student challenges



Computational Metagenomics Workshop organizers and trainers (left-right) Zahra Mungloo-Dilmohamud, Danial, Shakuntala, Prof. Yasmina, Vasilis, Christos and Gerrit.



Group picture of the Computational Metagenomics Workshop participants, organizers and trainers.

Zahra Mungloo- Dilmohamud. Dr. Shakuntala Baichoo.



- **H3ABioNet  
Advanced  
Systems  
Administration  
Workshop 2015**

## **H3ABioNet Advanced Systems Administration Workshop - Bioinformatics and Computational Biology Unit, University of Pretoria**

The second H3ABioNet Advanced System Administrators Workshop 2015 held at the University of Pretoria in the training room of the Bioinformatics and Computational Biology Unit, covered the basic knowledge needed by system administrators of the nodes to help them maintain their bioinformatics computer infrastructure that should be up and running. During the course, the lectures were divided into two sections and were taught by Prof. Fourie Joubert, Prof. Scott Hazelhurst, Peter van Heusden and Mohamed Alibi.

In the beginner section, courses included the basic system and package installation, configuration, usage and management. Then, trainees were exposed to a variety of basic network and security knowledge to help maintain their equipment in their institution and ensure their network is running adequately and with the capacity to access them from anywhere at the site. In another phase, as part of their server management, trainees were exposed to the best practices for storage management (setup and backup). Finally, Trainees finished their first section with training on other ways to install programs and how to schedule and treat jobs.

After finishing the beginner section trainees were ready to start the next phase of the workshop, the advanced system administration course. This course started with an introduction to bash (Shell Script) programming to ensure trainees have a full knowledge on how Linux command management works. They moved onto virtualization techniques to teach them how to establish and manage their own proper virtualization system. After that, trainees moved to advanced storage management to be trained on how to setup a network shared storage space (NFS, Samba) and how to setup the RAID technology configuration (Hardware, software). After having a rest day, trainees moved to the advanced system management to gain knowledge on automated management (Puppet), and then to version control (GIT). After that, they were trained on cluster setup, management and monitoring, with some advanced cluster software version usage technology (Environment modules). They finished the clustering section by running some parallel programs using MPI and had an introduction to Gluster for cluster storage management. Finally, Trainees finished the training with an overview of grid computing, certification management and a solution for network security for Intrusion detection and prevention (Snort).

During the training, trainees got two courses to learn how to install Globus Connect and the NetMap project client on their machines to replicate the installation at their home institution, as these are two milestones provided by the H3ABioNet Infrastructure Working Group.

Trainees and trainers stayed at the same Guesthouse during the whole course to create an environment of comradeship, to enable trainers and trainees to interact and to establish a relationship between each other for future work needed for the project.





- **H3ABioNet  
Advanced  
Systems  
Administration  
Workshop 2015**

## **H3ABioNet Advanced Systems Administration Workshop - Bioinformatics and Computational Biology Unit, University of Pretoria**

Most of the trainees have been using MS-Windows OS for their entire careers, so most of the trainers tried to reference things from the MS-Windows system to help introduce technologies and concepts. This made it more of a migration focus than an introduction to new technologies.

The training aim was not only to create system administrators capable of managing their infrastructure, but also those capable of evolving in this field, resolving their problems, and even helping resolve other people's problems inside their home institution. It was also to enable them to help other nearby H3ABioNet nodes if needed, to make them a part of the ISWG Task forces. The Trainers' second task was to identify motivated and interested system administrators for the ISWG Task forces.

Many nodes' system administrators were happy to have access to such training that will give them the capacity to evolve in their field and to experience new technologies that can simplify and secure their work as administrators for their system and network.



Group picture of H3ABioNet Advanced Systems Administration workshop participants at the Bioinformatics and Computational Biology training laboratory run by Prof. Fourie Joubert.



- H3ABioNet Advanced Systems Administration Workshop 2015

## H3ABioNet Advanced Systems Administration Workshop - Bioinformatics and Computational Biology Unit, University of Pretoria



Suresh Maslamoney and Sibiri Sawadogo subtly indicating to the trainers that it is time to take a coffee break to allow the knowledge gained to permeate.



Group picture of participants and trainers at the guesthouse where they were residing and provided an opportunity for the workshop participants to freely interact with the trainers after the workshop.

Mohamed Alibi.



- **Announcements**

## Announcements

- H3ABioNet Central will be scheduling one on one meetings with H3ABioNet Node PIs for the month of April 2015.
- NetCapDB and Staff forms for 2015 for the H3ABioNet Nodes are overdue for submission – please contact the Network Manager to determine what documents are outstanding from your Node.
- Congratulations to the MUHAS H3ABioNet Node for their manuscript acceptance for publication in PLoS One titled “Negative Epistasis between Sickle and Foetal Haemoglobin Suggests Reduction in Protection against Malaria”, PMID: coming soon.
- The “Medical population genetics and GWAS for complex diseases” symposium and workshops organized by the H3ABioNet CBIO and CPGR Nodes will take place between the 19<sup>th</sup> and the 24<sup>th</sup> of April 2015 at the African Institute for Mathematics in Cape Town, South Africa.
- The sixth H3Africa Consortium meeting will take place between the 9<sup>th</sup> and the 11<sup>th</sup> of May 2015 in Livingstone, Zambia.
- The Wellcome Trust will be giving a “Genomic Epidemiology in Africa” workshop between the 21<sup>st</sup> and 26<sup>th</sup> of June 2015 at the Africa Centre for Health and Population Studies, University of KwaZulu-Natal, Durban, South Africa.



## H3ABioNet Working Group Meeting Schedule until June 2015

Summary of H3ABioNet upcoming working group meetings:

| Working Group | Time       | Date | Month (2015)       |
|---------------|------------|------|--------------------|
| E&TWG         | UTC = 11am | 31   | March              |
| E&TWG         | UTC = 11am | 14   | April              |
| E&TWG         | UTC = 11am | 28   | April              |
| E&TWG         | UTC = 11am | 12   | May                |
| E&TWG         | UTC = 11am | 26   | May                |
| E&TWG         | UTC = 11am | 9    | June               |
| E&TWG         | UTC = 11am | 23   | June               |
| USWG          | UTC = 11am | 2    | April              |
| USWG          | UTC = 9am  | 1    | May (Bank Holiday) |
| USWG          | UTC = 9am  | 22   | May                |
| USWG          | UTC = 1pm  | 5    | June               |
| RSWG          | UTC = 11am | 17   | April              |
| RSWG          | UTC = 11am | 15   | May                |
| RSWG          | UTC = 11am | 12   | June               |
| ISWG          | UTC = 1pm  | 30   | April              |
| ISWG          | UTC = 1pm  | 8    | May                |
| ISWG          | UTC = 1pm  | 19   | June               |
| NATF          | UTC = 1pm  | 27   | March              |
| NATF          | UTC = 1pm  | 24   | April              |
| NATF          | UTC = 1pm  | 29   | May                |
| NATF          | UTC = 1pm  | 26   | June               |

- **Upcoming H3ABioNet Working group meeting schedule**

Acronyms used for the various H3ABioNet working groups:

|        |                                      |
|--------|--------------------------------------|
| E&T WG | Education and training working group |
| RSWG   | Research working group               |
| ISWG   | Infrastructure working group         |
| USWG   | User support working group           |
| NAWG   | Node accreditation working group     |

Time zone conversion table to UTC for all of H3ABioNet working group meetings:

| UTC Time Offset     | Time Zone Name | Regions / Countries in the Timezone offset              |
|---------------------|----------------|---------------------------------------------------------|
| -5 hours (-6 hours) | CDT            | Chicago, USA (2nd November, 2014)                       |
| -4 hours (-5 hours) | EDT            | Boston, USA (2nd November, 2014)                        |
| 0 hours             | GMT            | Burkina Faso, Ghana, Mali, Morocco, Senegal             |
| +1 hour             | WAT            | Cameroon, Chad, Gabon, Namibia, Nigeria, Niger, Tunisia |
| +2 hours            | CAT            | Botswana, Egypt, Malawi, South Africa, Sudan, Zambia    |
| +3 hours            | EAT            | Ethiopia, Kenya, Tanzania, Uganda                       |