H3ABioNet Monthly Newsletter Issue 16 : July 2016

Foreword

This newsletter describes the H3ABioNet activities from January to June 2016. The first half of this year has seen the successful completion of a number of training workshops and the Education and Training working is busy working on the organization of several new workshops which will take place over the second half of the year. This issue includes an article describing the experiences of a trainer from one of the workshops which H3ABioNet provided support for. In other exciting training news, our first Introduction to Bioinformatics Course has started and is being broadcast to 19 classrooms around Africa. We will keep you updated on its progress!

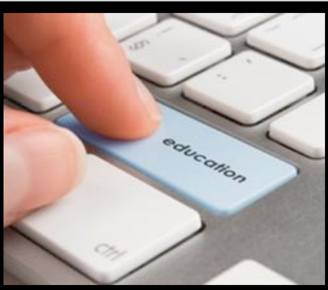
The other working groups are also working hard towards achieving their milestones and some of their activities are described below in their working group reports. The research working group established a Research proposals sub-committee to provide support for grant proposals through peer review. I hope the students and researchers will take advantage of this initiative to improve funding and other proposals before they get submitted. The working group has also successfully run a monthly webinar series which has seen some excellent research talks presented by both junior and senior scientists.

At H3ABioNet Central we submitted the H3ABioNet annual report to the NIH, which include requested budgets for year 5, our final year of the grant! We are now preparing for the next phase, and templates have been sent out to existing and potential new nodes to complete if they are interested in participating in the new proposal for an H3Africa Informatics Network.

I invite you now to continue reading to find out more about our activities over the last few months.

Prof. Nicky Mulder

- **Education and Training**



&

A number of training workshops have been planned and run during the first half of 2016 with various levels of involvement from H3ABioNet. The first was a Whole Genome and RNA Sequencing workshop hosted by the University of Khartoum, Sudan in March 2016. The training was conducted as a pre-conference workshop during the University of Khartoum's Postgraduate Conference. The workshop was held over five days and covered the practical analysis of whole-exome and RNA sequencing data. Although H3ABioNet did not directly fund the workshop, several H3ABioNet students were given travel fellowships to attend. You may read more about this workshop in a feature article included in this newsletter.

A workshop, focused on the Bioinformatics Analysis and Annotation of Variants in NGS data, was held as a satellite workshop to the Pharmacogenomics and Precision Medicine Conference, in April 2016 at University of Cape Town (UCT). The workshop was aimed at genetics, genomics and bioinformatics researchers interested in the use of genomic data for personalizing treatment. It included a combination of lectures and hands-on practical exercises covering topics such as NGS theory (an overview) and annotation of variants, including functional analysis and implications. H3ABioNet provided travel fellowships to 6 African students to attend the workshop and contributed to teaching and setting up of infrastructure for the workshop.

The H3Africa Nicol project, in collaboration with the J. Craig Venter Institute, USA, ran a practical workshop on microbiome analysis at UCT in April 2016. H3ABioNet provided technical and administrative support and contributed to the teaching. The workshop provided a background to microbiome studies and covered topics such as sample preparation, sequencing, run interpretation, pipeline processing for taxonomic classification and post-processing analysis. The workshop was designed for participants who already have basic bioinformatics skills and who may already have been involved in microbiome work.

The H3ABioNet Education and Training working group (E&T WG), in conjunction with the H3ABioNet Infrastructure working group, held a data support workshop prior to the 8th H3Africa consortium meeting in Senegal in May 2016. H3ABioNet flew in 17 H3Africa Data Support working group members to attend. This interactive workshop provided an overview of data management, analysis and submission for H3Africa projects. It also provided an opportunity for Data Support Working Group members who represent the H3Africa projects, to engage with H3ABioNet and discuss their needs.

PTO

- **Education and Training**

&

To complement our formal training endeavours, a number of interns have also been placed in order to gain some one-on-one training at our partner nodes in H3ABioNet. Mr Shaun Aron from the University of the Witwatersrand H3ABioNet node was awarded an internship at Harvard, where he spent 3 months working with various experts in population genetics. The latest internships include Ms Azza Ahmed, who has been awarded an internship at our University of Illinois H3ABioNet node, where she will learn about high performance computing for whole exome and RNA sequence data. Trust Odia and Falola Oluwadamilare from Covenant University in Nigeria have been placed at the Institute Pasteur node in Tunisia to work on modelling of host-pathogen interactions.

The E&T WG is also excited to announce that the online Introduction to Bioinformatics course was successfully launched on the 13th July 2016 and is currently underway. An immense amount of planning and preparation has gone into developing the course and we are excited to have over 350 participants registered for the course across 19 classrooms in 11 African countries! We have just completed the first two-week module for the course (Databases and Resources) and have received some great positive feedback from the classrooms thus far. Be sure to check out the course webpage that contains information on all the classrooms as well as the recorded course material (lecture recordings, slide decks and practical assignments) as they become available. http://training.h3abionet.org/IBT_2016/?page_id=130.

The IBT_2016 course is an introduction to the field of Bioinformatics, with a focus on important Bioinformatic tools, algorithms and resources. Over a period of 3 months of intensive biweekly contact sessions, the course will combine theoretical and practical sessions to allow participants to gain practical experience in using various tools and resources.

The course is organised and run by H3ABioNet.

IBT website homepage. Lecture videos, slides and assignments are open access

Shaun Aron and Nicky Mulder

- **Infrastructure**

The Infrastructure working group (ISWG) has been busy in the first six months of 2016.

- We helped run a very successful Data Support Workshop, in conjunction with the E&T WG, at the 8th H3Africa Consortium Meeting in Dakar, Senegal. This brought together research groups, representatives of the biorepositories, and other key stakeholders to discuss and plan how data, generated by the research groups, will be processed and moved. After hearing from participants on their goals, we had sessions on: laboratory information management systems; data harmonisation; overview of key pipelines and analysis; data processing, storage, transfer and security options; and drawing up a data management plan.

- The H3Africa Participants Recruitment Database went into operation. At the request of the funders we developed a database system to allow H3Africa projects to report on progress in recruitment of participants and collection of data. Implemented using REDCap, the system was developed in a short period of time and has received good feedback from all. Many people were involved in this, but particular credit goes to Mamana Mbiyavanga.

- The Cloud Computing Task force held several on-line meetings, and started a collaboration with Victor Jongeneel's group at the University of Illinois to work on projects of common interest. We are planning a Cloud Hackathon for August 2016 to develop portable computational pipelines of relevance for H3Africa and to develop skills on the continent.

- Many members of the ISWG have been very busy with the H3Africa Custom Chip design project. Using 300 genomes generated by H3Africa projects as well as other data sets, H3ABioNet is working with partners to develop a custom chip for the GWAS projects being done in H3Africa. This has required huge cross-continental data transfers (over 100TB) and lots of computation.

- We are still busy with a number of long term projects, including: rollout of Netmap and Globus Online endpoints for managing and monitoring data transfers; deploying and upgrading eBiokits; and updating systems administration documentation.

PTO

- **Infrastructure**

After a period of relatively little work, the Data Management Task force is starting to gear up for the expected deluge of data from the H3Africa Projects. This will be a major area of work for the next 6 months. Roll out of the Globus Online and netmap endpoints will continue, too.

Data Support workshop at 8th H3Africa consortium meeting in Dakar, Senegal

Suresh Maslamoney **Scott Hazelhurst**

Research

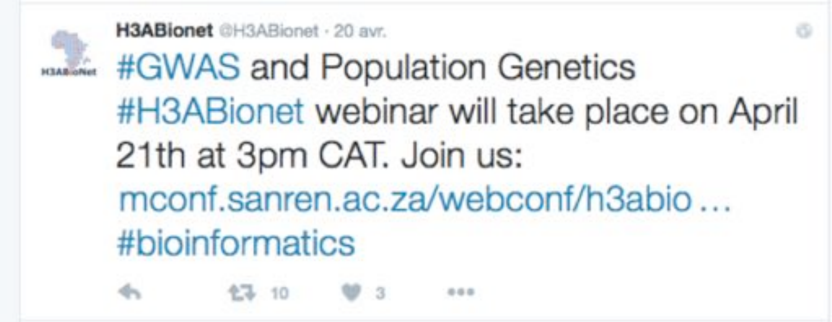
Since January 2016, a bioinformatics theme has been chosen for each research seminar run by the Research Working Group (RSWG). For each session, two speakers working in the field are invited to give a 20-minute presentation, followed by 10-15 minutes for interactions with participants, questions and answers. Topics covered so far are: GWAS and Population Genetics; Metagenomics; Big Data; NGS; and planning ones career within the field of bioinformatics.

The H3ABioNet Research Proposals Sub-committee was established. The chair of the RSWG complied a draft for general proposal guidelines, which was circulated among the members of the working group as well as other interested members of the H3ABioNet network. Some members of this sub-committee expressed their willingness to review nodes' proposals as well as graduate students' proposals.

A second Mozilla Study Group was launched in Khartoum, Sudan in March 2016 as a follow up for the NGS training workshop that took place there. Azza Ahmed is leading the study group, which meets every Monday. During each meeting, two members of the group, on a rotation basis, present and summarize a chapter within a selected bioinformatics book or article. This is followed by discussion of the presented material. We are aiming to launch other study groups at the different H3ABioNet nodes.

H3ABioNet is organizing a joint hackathon with the IBM Research Africa team. The main objective of this hackathon is to analyse transcriptional data sets from a range of Malaria parasite isolates exposed to artemisinin in culture, which have been generated by University Notre Dame (US). This data needs to be pre-processed and prepared for an open challenge that will be launched under the DREAM Challenges initiative within the next 2 years. The hackathon will be held in Johannesburg, South Africa from 5th to 10th September 2016. For further information, kindly refer to the hackathon call available at this link: http://h3abionet.org/h3abionet-ibm-research-africa-hackathon-malaria-drug-resistance-2016

Social media channels like twitter and Facebook have also been used to advertise for the H3ABioNet Research working group seminars to promote the seminars and attract more attendees.



H3ABionet @H3ABionet · 20 avr.

#GWAS and Population Genetics #H3ABionet webinar will take place on April 21th at 3pm CAT. Join us: mconf.sanren.ac.za/webconf/h3abio… #bioinformatics

Amel Ghouila **Faisal Fadlelmola**

- User Support

User Support

Overview

This period (January – July 2016) came right on the heels of the fourth consortium Annual general meeting in November 2015, and so began with lots of promise. Strategy for the Year 4 activities was presented at the general meeting and was to be quickly adopted and implemented. High on the list were: establishment of the long-overdue node liaison scheme, the Help desk management committee, new tools for outreach among these slide decks, flyers, brochures, the group article, and an improved website - all geared towards engaging a wider set of audiences.

Progress so far

Of the listed activities above, we have made some progress on the following:

- Member terms of reference (TOR) - these have been updated and revised.
- Work group reconstitution - we have registered dwindling meeting participation from members. We are currently revisiting working group composition, calling for new volunteers, re-examining compiled deliverables and laying out new strategys for future operations.
- Helpdesk management committee - advanced work plans are in place for this activity.
- Outreach - we are gearing to start on this in earnest.
- Helpdesk article - has been started on but is yet to be finalized, so there is still room for additional contribution and authorship. Please get in touch if interested.

Challenges and way forward

- The dwindling activity participation mentioned above has greatly slowed down the pace of our group activities. In view of this we are re-branding, expanding membership to include various categories, and are focusing more on the outreach aspects of our compiled deliverables. For this we shall need a diverse set of skill sets and we see lots of opportunities for contribution. We are also contemplating the launch of a monthly group journal club to draw more people. We are calling for renewed commitments from all volunteers, as success depends on collective participation in our coming group activities.

Jonathan Kayondo **Pandam Salifu**

- Progress in the H3ABioNet node assessment exercise

Progress in the H3ABioNet node assessment exercise

During the initial H3ABioNet General Assembly in 2012, the Node Managers decided that the network should develop and implement a Node Assessment Exercise (NAE). The primary goal for the NAE was to give individual Nodes the opportunity to prove, in an objective fashion, that they had reached the level of technical competence that could be expected from groups that would provide data analysis services to H3Africa research projects. It is one of the major goals of the H3ABioNet network to develop technical skills and infrastructure on the African Continent to a level where African scientists no longer depend on colleagues in Europe, Asia or North America to analyse the genomic data generated by their research projects.

A Node Assessment Task Force (NATF) was assembled after the General Assembly, and tasked with developing a framework for the organisation of the NAE as well as the materials for two specific exercises driven by the needs of the H3Africa research projects: the calling of genomic variants from high-throughput genome sequence data, and the processing of chip-based genotyping data followed by genome-wide association studies (GWAS). The NATF produced an Assessment Protocol outlining the procedures used to administer the NAE, including an independent evaluation by experts external to H3ABioNet. It also produced the materials for the first two NAE: standard operating procedures (SOP) for the analysis, training datasets to help nodes prepare for taking the NAE, and personalized test datasets used during the actual assessment. The University of Cape Town Node took the leadership for the GWAS NAE, while the University of Illinois worked on the variant calling NAE. All the necessary materials were ready by end 2013.

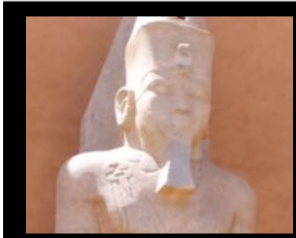
During the 2013 General Assembly, several nodes expressed the desire to have additional NAE available to them, to test analysis skills that may be more widely shared. As a result, the NATF produced two more NAE, for the analysis of microbiomes from the high-throughput sequencing of 16S rDNA amplicons, and for gene expression analysis from RNA-seq datasets. The necessary materials were prepared in a collaboration between the UCT and Illinois nodes, and have been available from the fall of 2014.

The first H3ABioNet node to actually take the plunge was the Wits node in Johannesburg, who submitted to the NAE in the summer of 2014. The Wits node took advantage of an H3ABioNet GWAS course and of the presence of an intern from the Botswana Node to assemble an analysis team and a shadow team, who independently analysed the test data. The report they submitted was reviewed very positively, and they were accredited with flying colours. It took over a year before the next two nodes declared themselves ready in the fall of 2015: the UCT node for the variant calling NAE, and the CUBRE node in Nigeria for the GWAS NAE. Again, both nodes did very well, and were accredited as a result. It is very encouraging to see that all nodes that took the NAE performed extremely well in the view of our external experts.

It should be noted, however, that H3ABioNet includes 32 different nodes, and that so far only three have taken one of the NAE, and none have taken more than one. The H3ABioNet leadership has made it a priority for all participating nodes to concentrate on training their staff, through courses and internships, to get ready and take the NAE before the end of the current grant in 2017. We sincerely hope that more nodes will prove their commitment to actually develop their technical capacity and to participate in the NAE within the next year.

Victor Jogeneel

- **Whole exome and RNA seq analysis course, March 2016, Khartoum**



Whole exome and RNA seq analysis course, Khartoum

In March 2016 I travelled to Sudan to help run a weeklong whole exome and RNA sequencing data analysis course at the University of Khartoum. The course was attended by 25 students, mostly from the University of Khartoum with a wide range of backgrounds, including researchers from medicine, veterinary and chemistry departments.

The group in front of the University Library

Course Design

I was tasked with covering the whole exome sequencing (WES) process. I started with an introduction to sequencing technologies, different sequencing approaches, and the challenges posed by each approach. Over the next few days I covered alignment, variant calling and annotation, the bioinformatic processes required to turn the raw sequencing data into a list of variants and how to analyse and interpret these variants. The mornings consisted of talks providing a background to the concepts and tools, which were then applied in the practical session after lunch.

The first practical session took students from the raw data (BAM file) through to a list of variants (vcf file), covering quality control and applying different filters. In the second session students took a real diagnostic case, using the patient's phenotype to design a gene panel, annotate variants and apply filters and bioinformatic tools to identify the causative variants.

The course ran very smoothly

The course and practical sessions ran very smoothly with students very interactive and open, asking lots of questions and demonstrating understanding of the material presented. Under the leadership of Prof. Faisal Fadlelmola the University has developed a very good computer infrastructure, ideally suited to running such a course, with 20 workstations running BioLinux and a cluster, all of which are in very safe hands with Azza Ahmed, the group’s system administrator.

I would also like to thank Rehab Ahmed and Azza (again!) for their help answering questions and troubleshooting throughout the practicals.

- **Whole exome and RNA seq analysis course, March 2016, Khartoum**

The computer suite is well set up for running such a course

Course photo in classroom

Experience of Sudan

I didn't know what to expect from Khartoum, however I was very pleasantly surprised. Everyone I met was very friendly and hospitable, a point demonstrated by nearly half the course acting as expert guides for us on a tour of Khartoum the day after the course had finished. One of the students kindly presented us with a painting of an Alzeer, a traditional pot to store water in the Alhoush (the communal courtyard of a family home) she had painted herself.

I felt very safe travelling around the city and despite the city not being a tourist hotspot the hotel was very comfortable and we ate very well at a range of restaurants! One restaurant even treated us to a performance from one of finalists (?) from Sudanese Pop Idol!

Some things were as I expected - it was hot! Most days were above 40^{O}C (apparently reaching 47^{O}C one day!) with the odd sand storm. Luckily the PC room had fans and AC! It was interesting to see the PC's being wrapped up each evening and how despite this they were still covered by sand and dust the following morning!

- **Whole exome and RNA seq analysis course, March 2016, Khartoum**

The (?) finalist from pop idol performing at a restuarant

The University library at night

Thanks

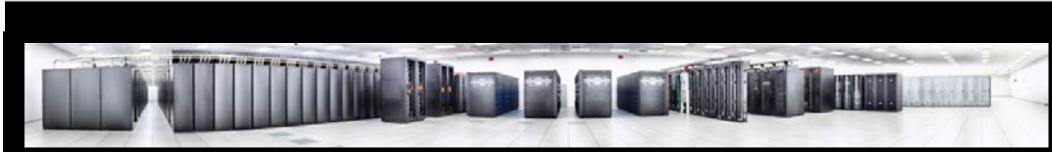
I would like to thank H3ABioNet and the Mozilla Foundation for funding without which the course would not have been possible and Jelena and Risha for the introductions and encouragement!

I'd also like to thank Faisal for making us feel so welcome and hosting us throughout the week and the University of Khartoum who were very helpful, making the Visa and immigration processes very smooth.

This article comprises excerpts from an article by Aled Jones

Original article available at http://trendinafrica.org/wp-content/uploads/2016/04/Sudan-school.pdf

- **Super computer changing Genetic medicine in Africa**



Super computer changing Genetic medicine in Africa

Recently, Susan Szuch wrote a newspiece for the National Center for Supercomputing Applications (NCSA) newspage, in which she described the H3Africa African chip design project and the invovlement of the NCSA Blue Waters team in making it possible “to discover genomic variants in over 300 deeply sequenced human samples to help construct a genotyping chip specific for African populations”. As many of you are aware, the H3Africa Genome Analysis Working Group partnered with H3ABioNet and the Wellcome Trust Sanger Institute to design and construct this chip, which will be used as a tool for rapid and inexpensive genotyping of individuals in order to identify genomic bases of disease, specifically in African populations. Due to the large amounts of data, this project required a very fast computer - where better to go than NCSA’s Blue Water; the fastest super computer on a university campus.

Szuch notes the effort of the Blue Waters team as being instrumental to the success of the project, in which Professor Nicola Mulder’s team at the University of Cape Town (UCT), in partnership with Dr Manj Sandhu’s team at the Wellcome Trust Sanger Institute, developed the computational workflow for extracting genomic variants, and in which the High Performance Biological Computing node (HPCBio) helped initiate the workflow on Blue Waters.

In the article, Szuch quoted Dr. Liudmila Mainzer, senior research scientist with NSCA and the HPCBio group, in saying that "without their [Blue Waters’] involvement and professionalism, this work would have been very difficult to complete. For our project, the Blue Waters team was as important as the compute, storage and networking resources."

The article also describes how the project was a mutually beneficial one and that the NSCA Blue Waters team discovered new tools for data transfer through troubleshooting problems in conjunction with the H3ABioNet team of bioinformaticians. The importance of this project in terms of social impact and reach was also noted.

Kim Gurwitz

The original full article, written by Susan Szuch, is available at **http://www.ncsa.illinois.edu/news/story/supercomputer_changing_genetic_medicine_in_africa**

- Announcements

Announcements

Congratulations to Gerrit Botha and Liudmila Sergeevna Mainzer who recently provided oral presentations on the bioinformatics data processing of the H3Africa Genotyping Chip and Variant calling approaches on the Blue Waters Supercomputer at the 4th Annual NCSA Blue Waters Symposium (https://bluewaters.ncsa.illinois.edu/bw-sym-2016)

Gerrit Botha also received the prize for the most data transfers conducted on Blue Waters.

Congratulations to two Consortium students (Olivier Sheik Amamuddy and Phillip Magambo), who completed one-year Bioinformatics MSc programmes in the Research Unit in Bioinformatics at Rhodes University in 2015 and started their PhD degrees at RUBi this year. Olivier is from Mauritius and a member of Prof Yasmina J. Fakim's group as part of H3ABioNet, and Phillip is from Uganda and a member of Prof Enock Matovu's group as part of TrypanoGEN.

Congratulations to RUBi members Phillip Magambo and David Brown who received best poster and best oral presentation, respectively, at the national SASBMB conference in July 2016 within the Bioinformatics and Omics track. Phillip Magambo presented his MSc work on a drug discovery project against Human African Trypanosomiasis (sleeping sickness). David Brown presented his work on "Human Genome Mutation Analysis Portal – HUMA" which is being designed as part of the H3ABioNet project to host human mutation information and analyse nonsynonymous mutations at the protein level.

Congratulations to Jaquiline Mugo from central node, who was awarded a scholarship to visit our SAB member Mark McCarthy's lab in Oxford and to Emile Chimusa, who has taken up a Senior Lectureship in the Division of Human Genetics at UCT. Congratulations to Nicola Mulder, who co-authored a Science paper with the Global Alliance for Genomics and Health: http://www.ncbi.nlm.nih.gov/pubmed/27284183

RUBi members, Rowan Hatherley, David Brown, Michael Glenister and Ozlem Tastan Bishop, have completed the first phase of the homology modelling pipeline, PRIMO (PRotein Interactive Modelling pipeline), and it is available for the use of the Consortium members at https://primo.rubi.ru.ac.za/. PRIMO is for protein monomer modelling, and allows users to model ligands and ions in complex with their protein targets. Its user-friendly interface makes it easy to use and if necessary to change the modelling parameters.

Any announcements for the next edition of the newsletter may be send to kim.gurwitz@uct.ac.za AND sumir.panji@uct.ac.za

- Upcoming events

Upcoming events

August 22nd to August 26th 2016; H3ABioNet Cloud computing hackathon, Unviersity of Pretoria, South Africa (http://www.h3abionet.org/17-h3abionet-courses/h3abionet-courses-upcoming/266-h3abionet-cloud-computing-hackathon)

September 5th to September 10th 2016; H3ABioNet / IBM Research Africa Hackathon on Malaria Drug Resistance for a DREAM challenge, Johannesburg, South Africa (http://www.h3abionet.org/h3abionet-ibm-research-africa-hackathon-malaria-drug-resistance-2016)

September 20th to September 23rd 2016; Joint 2016 conference of the South African Societies for Bioinformatics and South African Genetics Society, Durban South Africa. ****Early registration deadline 15th of August 2016** (http://www.sasbi-sags.co.za/)

October 3rd to October 14th 2016: INDA Hands-on NGS-Statistics course, Dakar, Senegal. ****Applications close on August the 15th 2016** (http://www.h3abionet.org/component/content/article/17-h3abionet-courses/h3abionet-courses-upcoming/269-inda-hands-on-ngs-statistics-course)

October 27th to October 31st 2016: Ninth H3Africa consortium meeting, Mauritius (http://h3africa.org/events/9-news/284-9th-meeting-of-the-h3africa-consortium-2)

November 6th to November 10th 2016; Ninth Annual RECOMB/ISCB Conference on Regulatory & Systems Genomics, with DREAM Challenges & Cytoscape Workshop, Phoenix, Arizona, USA (https://www.iscb.org/recomb-regsysgen2016)

November 21st to November 23rd 2016; The fourth International Society for Computational Biology Latin America Bioinformatics Conference (ISCB-LA), jointly organized with A2B2C, will take place in Buenos Aires, Argentina (https://www.iscb.org/iscb-latinamerica2016)

Important dates for PIs:

November 1-4th : H3ABioNet SAB (1-2nd) and Annual General Meeting (3-4th), Cape Town, South Africa.

31st August is the deadline for the sumission of your final claims for year 4.

Any upcoming events for the next edition of the newsletter may be send to kim.gurwitz@uct.ac.za AND sumir.panji@uct.ac.za

- **Upcoming H3ABioNet Working group meeting schedule**

H3ABioNet Working Group Meeting Schedule until end August 2016

Summary of H3ABioNet upcoming working group meetings:

Working Group	Time	Date	Month (2016)
Cloud computing task force meeting	UTC = 14:00	Thursday, 11	August
RWSG	UTC = 11:00	Friday, 12	August
H3ABioNet seminar series	UTC = 13:00	Thursday, 18	August
ISWG	UTC = 13:00	Friday, 19	August
USWG	UTC = 9:00	Friday, 26	August
E&T WG	UTC = 11:00	Tuesday, 30	August

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	CDT	Chicago, USA
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

This edition of the newsletter was compiled and edited by Kim Gurwitz. For any corrections, please contact Kim at kim.gurwitz@uct.ac.za