H3ABioNet Monthly Newsletter Issue 7 : November 2013

- Foreword
- Education and Training
- Infrastructure
- Research
- User Support
- Node Accreditation
- Next Generation Sequencing workshop KTP / CPGR, 2013
- Announcements
- Important Dates

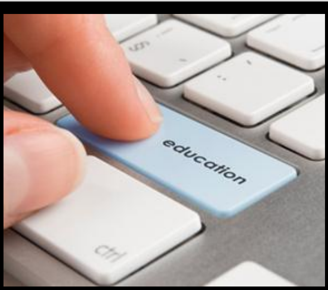
Foreword

November is a busy time of the year for us as it is the end of our academic year and thus involves exam marking, completion of purchase orders etc., ready for the festive season break. We finalised any outstanding year 2 contracts and paying of advances, as well as collected carryover request forms, and we submitted our 18 month report to the NIH. The Central group has also been busy with continued participation in H3Africa and H3ABioNet working groups, as well as newly formed task forces, such as the curriculum development taskforce, NetMap, and a team working on development of an African chip. The curriculum development task force is making good headway on planning of a Masters course and a curriculum development workshop for the new year.

Central has also been involved with several internal meetings coming out of a talk on H3ABioNet given at an eResearch conference in Cape Town. We have met with members of TENET, CSIR and UCT libraries and ICT to resolve issues such as poor bandwidth and data storage. We hope these fruitful discussions will continue next year.

As the year winds down, I would like to wish everyone well over the festive season, if you are taking leave I wish you happy holidays, and I hope everyone returns in January revived and ready to work full steam ahead on our next set of deliverables!

Prof. Nicky Mulder.



&

Education and Training

The Education and training working group will be preparing for two upcoming events in the following year:

1) Preparation for a curriculum development workshop
2) Preparation for the upcoming Masters training course to be held

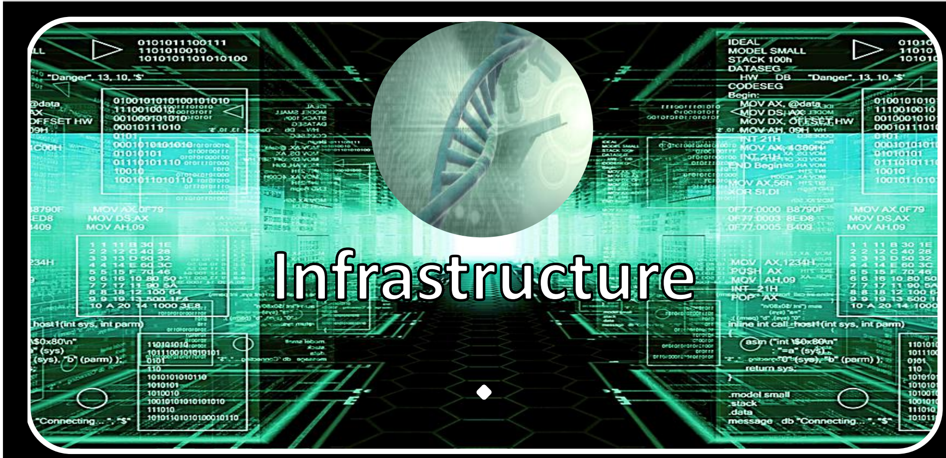
There has been a lot of feedback to H3ABioNet from various groups that are internal and external to the H3africa projects and within H3ABioNet of the possibility of devising a Masters curriculum for bioinformatics within Africa. Existing curriculums for Masters degrees within Africa have been collected but it is by no means complete there is no information / curriculums obtained on degree programmes from many regions such as Ghana and Nigeria. The E&TWG has formed a task force to help develop this and also make recommendation for the topics for the upcoming Masters course. A motivation template will be completed to hold a two / three day curriculum development workshop in Botswana possibly around the March / April in 2014. It is hoped that some concrete recommendations can be made as to what a Masters bioinformatics curriculum should be and how it can be implemented. The Masters course to be run by H3ABioNet will be aimed at students studying for a degree which has a significant component of the project in bioinformatics and are funded by H3ABioNet. At present, the number of students being funded and the regions they are from are being collected via Year 2 staff forms in order for the E&TWG to determine where the highest concentration of students are located by region with a view to hosting the H3ABioNet Masters course in that region. The curriculum task force with the assistance of the E&TWG has so far narrowed the broad range of topics suggested by members to the following:

Topic	Duration
Introduction to linux and programing	1 week and throughout if possible
Introduction to biostatistics	1 week
Sequence alignment, proteomics and bioinformatics resources	1 week
Databases, mining and analysis workflows / pipelines	1 week
NGS / GWAS	1 week (core)
Total number of weeks : 5 weeks	

A major limiting factor for the duration and number of topics covered by the Masters course would be the cost involved to provide accommodation and catering for an extended period of time as well as the cost of flying in external lecturers. The next step would be to obtain a list of potential lecturers based in Africa that would be willing to teach some of the topics, so if you do have any suggestions please do email them onto the working group.

Dr. Nash Oyekanmi.

- Infrastructure



The infrastructure working group Chairs have devised a list of specific milestones that will be the focus of this working group’s activities for the upcoming period. The milestones include the Iperf tests, the determination of Galaxy needs, updates on the server and eBioKit purchase status by the various Nodes, data storage and management. Three task forces have been created to work on the different milestones which include:

1) Data management task force
2) NetMap project task force
3) Systems administration task force

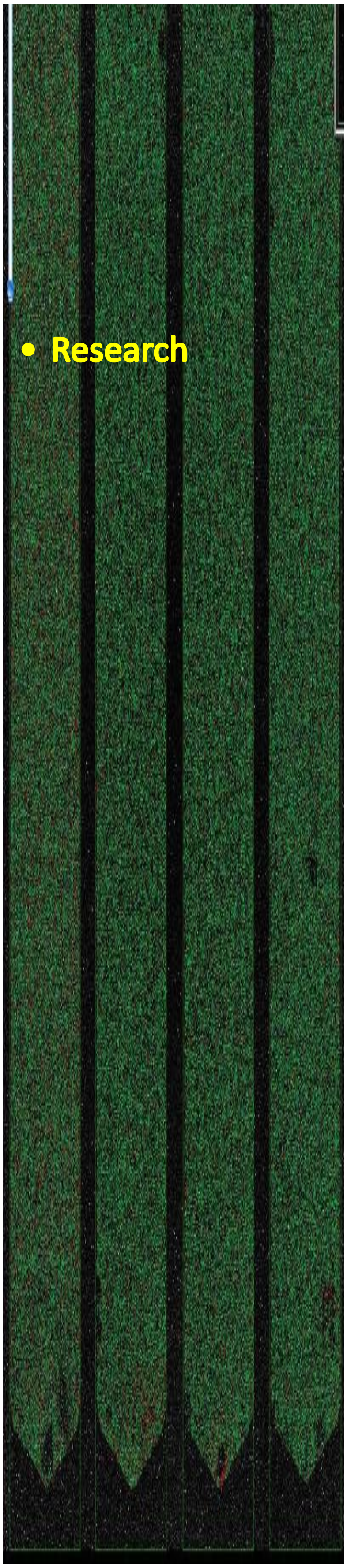
The data management task force has had 2 meetings to examine the data management policy devised by Central and discuss whether the storage system to be created for H3Africa should be centralised or decentralised. A centralised option for data storage and management was settled upon due to the concerns about security and management of the data that would need to be done at the various potential sites. In terms of location for the storage facility, South Africa was deemed to be the best option due to a constant supply of power, good internet connectivity which is important for data acquisition and submission, technical expertise with a lot of the full Nodes being located within South Africa, ability to leverage Institutional resources.

The NetMap task force has regular weekly meetings with the various members being assigned different Python coding tasks to develop a system for testing the various H3ABioNet servers on a regular basis and to obtain the results and statistics and push them to a spreadsheet application for further analysis. Subsequent meetings resulted in a GitHub project being created with the various members (Scott, Peter, Ayton and Gerrit) volunteering to write various parts of the Iperf testing module in python code and checking on each other’s progress.

The systems administration task force is dealing with the installation of various Linux operating systems (OS) on the servers received by the different Nodes and compiling documentation in the process. So far, Scientific Linux (CentOS), Ubuntu and Debian are the flavours of Linux that are currently being trialled with a view to provide support for. Most of the bioinformatics software available for running large jobs use a flavour of one of the above mentioned Linux OS. In terms of servers that have a Windows OS, virtual machines for one of the Linux OS will be recommended to be created and used.

Dr. Alia Benkahla. **Prof. Scott Hazelhurst.**

Research



- Research

The Research working group has been in discussions about obtaining and devising various bioinformatics workflows and will work with the User Support working group on this task. The idea is to create a graphically represented set of commonly used workflows for various bioinformatics analysis such as genotyping to GWAS or variant calling which can be placed on the H3ABioNet website.

Each component of the workflow process will be represented by a figure and hyperlinked so when that image is clicked, the user is provided information as to the aim of that workflow step, the required inputs and outputs from that step and a suggestion of possible software that can be used. Some members have provided pre-existing workflows from their Institutions and more are being collected with the Node Accreditation Working group giving permission for the SOPs they are developing to be used.

The workflows that have been prioritised to work on initially are the genotyping and GWAS workflows and the NGS variant calling workflows. A task force has been put together which will develop a common library of images for use and will also work on the text for the various workflow processes devised.

An online form will be created for the various Nodes to indicate if they would have any datasets they would like to analyse or if a Node has any particular expertise that it can provide to another Node to facilitate collaboration between the various Nodes. This will help the transfer of various skills across the Network and hopefully increase the research output of the various Nodes and the Network as a whole.

Dr. Julie Makani. **Prof. Ezekiel Adebiyi.**

- User Support

User Support

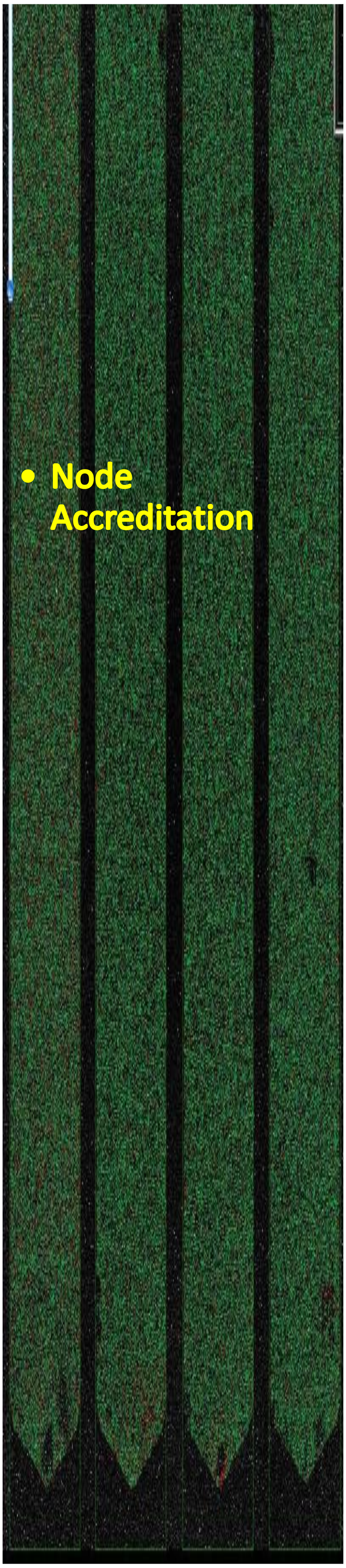
In conjunction with the Research working group, the User support working group is working on collecting various bioinformatics workflows to place on the H3ABioNet website as described within the Research working group section of this newsletter.

The H3ABioNet helpdesk usage has been looked at for the first period with various metric being collected. The H3ABioNet Bioinformatics helpdesk has been moved to the public webpages for unregistered and non H3Africa personnel to make use of upon the suggestion of the SAB. So far there are 10 categories (genotyping arrays, NGS data analysis, NetCapDB, software development / programming, software license request, systems administration, website / mailing list, general project administration and an "other" category). To date there have been 46 tickets submitted of varying complexity with a total of 947 minutes being spent working on these requests. The systems administration and general administration categories have the most tickets so far (15 and 12 respectively). We expect the usage of the H3ABioNet and categories chosen will change once the H3Africa projects start generating data and doing analysis and having more data driven questions.

The next focus for the H3ABioNet helpdesk will be to have an induction for members who volunteered to be part of the helpdesk and create a mailing list for this group to communicate to each other easily.

In conjunction with the Education and Training working group, the USWG will be looking at ways of facilitating a monthly seminar series and what the logistics for having a series will entail. It is envisioned that the seminars will be available online for a task force to decide on a content schedule and enable participants to view for a month and open a forum for discussion at a particular time although the actual logistics is to be planned.

Dr. Judit Kumuthini. Dr. Jonathan Kayondo.



- Node Accreditation

Node Accreditation

In November, the NAWG has been working on the following issues:

1. The finalization of the Standard Operating Procedures and test datasets for the variant calling and genotyping assessments. Some concerns have been raised that the variant calling workflows had a low sensitivity for the detection of synthetic variants in the test datasets and this is currently being addressed. The NAWG also needs to make sure that the SOPs are informative enough to serve as guides to the candidate nodes, but not so prescriptive as to make the exercise meaningless which has been challenging.

2. The implementation of the first assessments in early 2014. Several South African nodes have expressed the desire to be assessed as soon as possible. This will be an excellent opportunity to ensure that the procedures and datasets are suitable, and that the evaluation procedures are transparent and fair. We will work with the nodes in question to iron out any remaining kinks. The evaluations are expected to being in around February - March 2014.

3. The preparation of new assessment exercises. Priority will be given to the analysis of 16S rDNA amplicons and the derivation of the taxonomic complexity of microbial communities, as this is a component of several H3Africa research projects. Analysis of RNA-Seq data, bacterial genome assembly and annotation, and analysis of microarray-based gene expression data will be next. The Central and CPGR nodes will contribute their existing procedures, and the SOPs will be developed jointly by them and Illinois.

4. The distribution to the H3ABioNet community of all the SOPs and datasets prepared by the NAWG. We agreed that all these should be made publically available, with the exception of the datasets used for the actual evaluations. They will be put on the H3ABioNet Web site starting in late January.

Dr. Victor Jongeneel.

- Next Generation Sequencing workshop KTP / CPGR, 2013

Next Generation Sequencing Workshop-KTP / CPGR 2013

The Center for Proteomics and Genomics Research recently hosted a NGS workshop from the 14th to the 15th of November as part of the Knowledge Transfer Programme (KTP). KTP is an initiative organized by CPGR where instead of sending individuals to learn new skills related to specific techniques, technologies, methodologies, or applications at another institute, an expert is brought in to work on a specific project or to train individuals through a workshop in order to maximize the transfer of skills to local scientists and create capacity within hosting institutes, companies or organizations. To provide African scientists the opportunity for hands-on experience in analysing specific local data from various current “omics” platforms, particularly Next-Generation Sequencing (NGS) data for training in novel applicable techniques in an applied commercial or academic setting, a two-day workshop was organized through KTP at the Centre for High Performance Computing (CHPC) in Cape Town. Topics included two of the most common NGS applications: variant detection and de-novo assembly. The workshop consisted of lectures and computer exercises with hands on data analysis. No previous experience with NGS data was required, however, basic knowledge of Linux was a pre-requisite and participants were expected to come with various backgrounds from biology and computer science. The course was designed at an introductory level, while at the same time covering the topics in sufficient detail so that the methods are useful in practice. Computer exercises were designed so that, students - with instructions printed in hardcopy fashion - could perform analyses of their own after completion of the course. All software tools used in the course were non-commercial and publicly available to facilitate use and application in the participants’ home or work environment.

The workshop attracted 78 applicants from different backgrounds and nationalities and from various Universities and research institutes across South Africa, of which only 31 were selected to participate due to limited funding and space. The course consisted of two NGS topics, namely variant detection and de-novo assembly. One day was allocated for each topic. Each day started with a lecture explaining the concepts, logic, goals, algorithms and work-flow behind the software tools; this was followed by computer exercises incorporating practical data analysis. During the computer exercises, each participant had a chance to explore the training data and analyse it with specific software tools. Two different software tools were selected for each application. This was aimed at exposing participants to critical thinking about available methodologies and to enhance interpretation of results.

Owing to the diversity of the participating group, different levels of skills and computer proficiency were expected. Computer exercises were designed so that more experienced users could experiment and try different ways to run the programs. At the same time, less experienced users were provided with printed hand-outs detailing dedicated commands to carry out analyses in a more structured manner. The workshop was presented by Dr. Panu Somervuo (University of Helsinki, Finland).

- Next Generation Sequencing workshop KTP / CPGR, 2013

Next Generation Sequencing Workshop- KTP / CPGR 2013

Dr. Somervuo received his PhD from Helsinki University of Technology, Finland, in 2000. His background includes signal processing, pattern recognition, and machine learning. Between 1994 and 2005 Dr. Somervuo worked with automatic speech recognition and artificial neural networks at Neural Networks Research Centre, Finland. From 2002 to 2003 he did a 12-month research visit to International Computer Science Institute (ICSI), Berkeley, USA. In 2005 Panu moved into bioinformatics and has since then been working in several laboratories at Viikki campus in University of Helsinki. Between 2010 and 2013 he was a member of University of Helsinki's Metapopulation Research Group and participated in a butterfly genome sequencing project which included genome assembly, transcriptome assembly, gene expression and variant detection using data from Illumina, 454, SOLiD, and PacBio sequencers. Before coming to CPGR, Panu taught various microarray and NGS courses in Finland, Sweden, Italy, and Mauritius.

We requested feedback from participants to determine the impact of the workshop with regard to organization, quality of the training, knowledge and skills acquired, and the general impression of the course. This information will be used to improve design and organization of future training courses, and to determine the success of the training provided. Feedback was collected using an online questionnaire. Out of 31 participants, 26 responded, representing an 83.87% sample. The feedback is presented in a structured fashion in sections that follow the quality of workshop preparation and organization

In order to assess the quality of workshop preparation, organization and to assess the quality of instruction, online feedback from all the participants were received for whether trainers met the training objectives; if adequate time was provided for questions and discussion, and whether the trainers were knowledgeable enough. The results indicate that participants generally agreed that the workshop was well organized and of high quality.

KTP /CPGR NGS workshop participants. The group photo was taken at the CHPC on the 15th of Nov by Mr. Dane Kennedy.

Acknowledgements to CHPC for subsidizing for meals and providing infrastructure for hosting the workshop.

Dr. Judit Kumuthini.

- Announcements
- Important dates

Announcements

- All H3ABioNet working group meetings apart from the task forces will be put on hold from December 6th till February 2013
- The Egyptian node has organized a training course about "Introduction to Bioinformatics" at biotechnology lab in faculty of agriculture, Fayoum university from 6/11/2013 to 8/11/2013 with the participants shown below:

Important Dates

- 25th November 2013 – all Nodes to submit their carryover requests
- 6th January 2014 – all Nodes to submit their updated staff forms and biographical sketches
- 30th January 2014 – 5 year developmental and annual reports to be submitted by all Nodes