H3ABioNet Montly Newsletter Issue 2 ; June 2013

H3ABioNet

- Foreword
- Education and Training
- Infrastructure
- Research and Tool Development
- User Support
- Node Accreditation
- H3ABioNet Research Focus
- H3ABioNet Systems Administration Workshop - Pretoria 2013
- H3Africa Consortium Meeting - Accra, Ghana 2013
- Announcements
- Important Dates

Foreword

Welcome to the June 2013 newsletter! This is the second H3ABioNet bulletin and provides a summary of our activities for the month. The main activity in which many of us were involved was the 2 week course for Technical / IT members of nodes. This was designed to equip nodes with skills on data security, Linux installation, Cloud computing and HPC. The course was hosted by Prof. Fourie Joubert from the University of Pretoria node and also provided an excellent opportunity for networking. The participants have a mailing list which continues to be actively used.

H3ABioNet Central has been working hard on negotiations with service providers and one-on-one discussions with nodes to find the best solution for their server/hardware purchases. We are finally at the stage where these can begin, and the technical workshop hopefully equipped the nodes with the skills to manage the equipment.

The H3ABioNet website has been redesigned and content updated (www.h3abionet.org). We have opened applications for internships to enable H3Africa staff and students to spend time in another lab learning bioinformatics or analysing their own data. Our network manager, Sumir Panji designed and presented an H3ABioNet poster at UCT's Africa week, and the network was also presented at the African Society for Human Genetics Conference in Accra, Ghana in May. A number of H3ABioNet PIs attended the H3Africa consortium meeting which preceded the conference and interacted with members of other H3Africa projects.

Another announcement of note is the participation of H3ABioNet members in the Global Alliance, which aims to enable the sharing of genomic and clinical data to accelerate progress in medicine. This is an exciting global initiative and it is important H3ABioNet plays a role in it.

Looking forward, we are completing arrangements for the next workshop, a 3-week train-the-trainer bioinformatics course to be held at International Center for Insect Physiology and Ecology (ICIPE), Kenya, followed by an eBioKit course. For both, we have good representation from different nodes and from H3Africa research projects.

July constitutes the final month of year 1 of the project and promises to be busy!

Prof. Nicky Mulder

H3ABioNet

- **Education and Training**

&

The Education and Training Working Group (E&TWG) has been busy in helping to organise and co-ordinate the logistics for one completed workshop (Systems Administration workshop in Pretoria) and three upcoming workshops which include an intermediate level course of Bioinformatics for Trainers (8th – 26th July, ICIPE, Nairobi), Introduction to Bioinformatics using the eBioKits platform (29th July – 2nd August, ICIPE, Nairobi) and Visual Analytics of Human Genome Variation Datasets (29th – 2nd August, NABDA, Abuja).

Participants from as many H3ABioNet Nodes as possible and some H3Africa projects have been selected to attend the Bioinformatics for Trainers and Introduction to Bioinformatics using the eBioKits platform by a Selection Committee – congratulations! We look forward to welcoming you in Nairobi shortly. The Visual Analytics of Human Genome Variation Datasets workshop is still accepting applications until the 15th of July, more information can be found on the H3ABioNet website under the Training and Education Section.

It has also become clear during the hosting of the H3ABioNet workshops that there are different expectations from participants as to how the logistics and financing of these workshops works. In response the E&T WG at the Management Level has drafted an H3ABioNet Workshop Policy which is currently under review by the H3ABioNet Management Committee (MC) and will be made available for subsequent workshops to avoid any future confusion.

The Systems Administration workshop in Pretoria has successfully been completed and was attended by 18 participants from 11 different African countries (see picture in announcements section) and the main highlight for course participants (apart from the high quality teaching) was the opportunity for getting to know and networking with their peers working in a similar capacity at the various H3ABioNet Nodes. The mailing list constructed for the Systems Administration workshop participants and lecturers has been left active to provide a platform for communication and facilitate continued discussion between workshop participants.

 The E&T WG group would like to extend a big "Thank You" to Prof. Fourie Joubert who provided the use of his computer training laboratory at the University of Pretoria and taught the Galaxy portion of the Systems Administration workshop, Peter van Huesden who taught the Linux installation and administration portion, Dr. Christopher Fields and Daniel LaPine from the University of Illinois who taught the data management and stewardship segment of the workshop and to Prof. Joubert's technical team for providing continual support and ensuing a smooth workshop.

Dr. Nash Oyekanmi.

Infrastructure

- Infrastructure

The Infrastructure Working Group (ISWG) is in negotiations with hardware vendors for the purchasing of servers requested by H3ABioNet Nodes. For the purchasing of servers the following processes were undertaken:

- Individually contact each developmental and associate Node that requested a server to complete a questionnaire which included the intended usage of their server and provide advice and support where appropriate.
- Based on the information gathered, three different server type specifications were derived based on the following scenarios:
  a) Some Nodes required a single server but are unsure what to obtain.
  b) Some Nodes required a server and infrastructure (racks, switches).
  c) Some Nodes required 2 small separate servers.

The brief provided to the vendors was a minimum of 10TB of storage, system should be upgradable, the hardware must carry a 5 year next day business warranty and provide as much RAM and Cores as the budget allowed ($15,000 per server).

For the above scenarios, Dell recommended the C6145 server series with a single server motherboard as it provides the best value for money and resources. The C6145 has the largest internal hard drive capacity, maximum number of cores for the price range and can support future hardware upgrades by accommodating a second server motherboard. In the case where 2 small separate servers were requested, Dell recommended the R715 for running bioinformatics applications and the R240 for running databases.

The quotes for the options have been broken down to "in-country delivery" and "airport only delivery" as Dell operates through various partners and re-sellers within different regions in Africa. Indeed, one of the most challenging issues faced has been communication between the different Dell Africa regional partners each of which have their different mark ups, tax, customs and duty restrictions between different African Countries making it impossible to obtain a single, unified pricing quote for each server option for the whole of Africa. However, we do estimate the bulk negotiation with Dell has resulted in between a 38% and 45% discount. Please consult http://www.h3abionet.org/attachments/article/9/H3ABioNet_server_purchase.pdf for more information. If you are a Node that requested a server, please peruse the documentation sent through the GA mailing list thoroughly.

Dr. Alia Benkahla.   Prof. Scott Hazelhurst.

Research and Tool Development

- **Research and Tool Development**

The H3ABioNet Research and Tool Development Working Group (RTDWG) has been focusing on the various research projects that are taking place within H3ABioNet.

The RTDWG will be re-assessing the various research milestones and deliverables required for the NIH reports and for the H3ABioNet research projects over the coming weeks. The RTDWG will also work with the various research project groups to re-assess the delivery dates for milestones that were originally submitted to the NIH as these were based on best estimates. Reassessment of the milestones will be done for each working group over the coming period and will be vital for the Network to achieve its maximum potential, identify areas of synergy between the various working groups and re-focus the different working groups on delivery of their goals.

If Consortia members would like to propose a project no matter how large or small e.g a 1 week research project, a 2 week coding project or database implementation or a large scale project, the RTDWG has adopted a Research Proposal template that can be found at:
http://www.h3abionet.org/attachments/article/41/H3ABIONET_project_proposal_template.docx

The H3ABioNet Project Proposal can be filled in and submitted to info@h3abionet.org by any individual that would like to propose a project and are looking for specific skill sets and the RTDWG will do its best to facilitate the project.

In terms of research projects, there are different on-going projects within the H3ABioNet consortium which will be highlighted within the "H3ABioNet Research Focus" section of the H3ABioNet newsletter in order for Consortia members to keep abreast of H3ABioNet research activities.

The first project to be highlighted in this month's issue is the NetCapDB project being undertaken by the South African National Bioinformatics Institute (SANBI) Node driven by postdoctoral fellow Dr. Jean-Baka Domelevo Entfellner, software developer Dr. Adam Dawe under the supervision of Dr. Nicki Tiffin with input from Prof. Alan Christoffels and Dr. Junaid Garmieldien. NetCapDB involves the design and implementation of a database to capture H3ABioNet metrics and will ultimately provide automated reporting for each of the Nodes enabling them to easily compile their annual reports. NetCapDB will aim to measure the impact and success of H3ABioNet for the various Nodes involved.

Do take the time to familiarise yourself with NetCapDB information requirements and get involved as it is compulsory for all H3ABioNet Nodes to fill in the required information.

Dr. Julie Makani.   Prof. Ezekiel Adebiyi.

H3ABioNet

- **User Support**

User Support

The H3ABioNet was announced at the Society for the Advancement of Science in Africa (SASA) conference and H3ABioNet flyers were distributed from the CPGR booth at the venue. H3ABioNet was presented at the second meeting of the H3Africa consortium which took place in Accra, Ghana on 16-18 of May (see). Brochures for H3ABioNet and the helpdesk were designed and distributed to the various members of the H3Africa consortium and to attendees of the African Society for Human Genetics via the CPGR booth.

So far there has been little traffic to the H3ABioNet helpdesk, but this is probably due to the fact the various PIs within the H3Africa consortium would need to filter the helpdesk information to their laboratory members. The USWG anticipates the students and members of the various H3Africa PI laboratories would be the most likely target audience to use the helpdesk as opposed to the actual Principal Investigators (PIs). To this effect, the USWG members have taken upon themselves to publicise the helpdesk when they attend any workshops or conferences such as the recently concluded Systems Administration Workshop in Pretoria and the upcoming workshops at ICIPE and NABDA as well as recruit any additional expertise to add to the existing user support personnel.

The USWG is also helping to support the anticipated roll out of the NetCapDB project by creating a special category for any NetCapDB related queries within the H3ABioNet helpdesk system for registered H3ABioNet website users and will help to review the NetCapDB documentation being compiled.

The USWG will be focusing on determining its next set of goals and milestones based on the responses of the surveys conducted earlier this year by H3ABioNet for the H3ABioNet Nodes and the H3Africa projects. Similar to the other working groups, the USWG will be re-working its milestone deliverables for the next NIH reporting period and will be looking at synergies for the various deliverables with the other working groups.

Dr. Judit Kumuthini.   Dr. Jonathan Kayondo.

• Node Accreditation

• Node
• Accreditation

The Node Accreditation exercise that this Work Group has been tasked to develop will test the ability of the nodes to successfully analyse datasets of the types that the H3Africa research projects will be producing. These will primarily be high-throughput sequences derived from human exomes (later whole genomes) and genotyping data from SNP chips. One of the crucial components of the exercise will be the choice of suitable input datasets that satisfy the following criteria:

1. The datasets should be different for each candidate Node, but similar enough so that no single Node will be favoured.
2. They should not be easily identifiable by matching them to the published literature.
3. They should be small enough not to require extensive computational resources for analysis.
4. To the extent possible, the variants or phenotype associations documented by the datasets should be known in advance, providing a yardstick by which to assess candidates' results.

While no final decision has been made, the NAWG is leaning towards the use of synthetic datasets rather than real ones for a number of reasons:

- The synthetic datasets can be easily produced in unlimited numbers.
- The genomic variants documented by the data will be known in advance and therefore, we will have an objective measure for the quality of the analysis.
- The datasets will be consistent across candidate nodes as they can be made to contain the same density of information e.g. frequency of variants, types of variation etc.
- The datasets will not be identical across candidate nodes and are impossible to match to known data, thus eliminating any potential biasness or perception of collusion.
- The NAWG will be able to introduce specific features into the datasets (e.g. population admixture or related individuals, or known disease-causing variants) for discovery by the candidates.

At present, different tools are currently being used and assessed by the various laboratories that are part of the NAWG to produce artificial NGS datasets and SNP genotypes on which we will expand on in the next H3ABioNet newsletter.

Dr. Victor Jongeneel.

- **H3ABioNet Research Focus**

H3ABioNet Research Focus - NetCapDB

H3ABioNet launched an ambitious project with NetCapDB, a comprehensive database-oriented tool for monitoring capacity building activities taking place within the network. NetCapDB has been primarily designed to assess capacity development within the H3ABioNet project, but it would potentially suit other needs.

NetCapDB is structured around a central database hosted at the SANBI node (Cape Town, South Africa). Individual Nodes of H3ABioNet (scaling from small emerging research groups to larger institutes) are responsible for entering their own data (e.g. how many people work in that node, how much money they get from external grants, how many PhD students are supervised or have recently graduated there, how often and how far do people travel to attend conferences, etc). The data entry process is carried out through a web-based interface (Xataface-based design), which makes it at the same time quicker and less cumbersome to enter data because it is neatly organised in so-called "dashboards".

These data are stored in a complex MySQL database (126 tables with a cumulative total of 1067 fields). Data will then be retrieved and processed through MySQL queries in order to generate accurate and rich reports made out of raw or synthetic figures, graphics and maps of Africa representing all the aspects of sustainable capacity building within the H3ABioNet initiative, ranging from increased research output to strengthened education and training activities, as well as steady growth in terms of personnel and computing / educational infrastructure.

The reports generated will be useful not only to the H3ABioNet Nodes themselves, but also to the H3ABioNet Scientific Advisory Body (SAB), the Management Committee (MC) and for each Nodes' annual report to the NIH providing them with an objective and comprehensive activities assessment as an input to the decision-making process about where and for which projects should funding go.

NetCapDB has been designed, implemented and tested from December 2012 to June 2013. Data entry from H3ABioNet Nodes will soon be launched (early July 2013). We kindly request that all the H3ABioNet Nodes to have completed the data entry process by August 15th. Meanwhile, the reporting interface is to be developed over the months of July, August and September 2013. A paper should be ready for publication around October/November 2013, with the first results showcased at the 3rd H3Africa Consortium meeting to be held in Johannesburg between the 3rd – 5th of October, 2013.

- **H3ABioNet Research Focus**

H3ABioNet Research Focus - NetCapDB

The NetCapDB project is driven at the SANBI Node by postdoctoral fellow Dr. Jean-Baka Domelevo Entfellner and software developer Dr. Adam Dawe, under the supervision of Dr. Nicki Tiffin and with input from Prof. Alan Christoffels and Dr. Junaid Garmieldien. A couple of user feedback meetings have also taken place with participants from the Central UCT node.
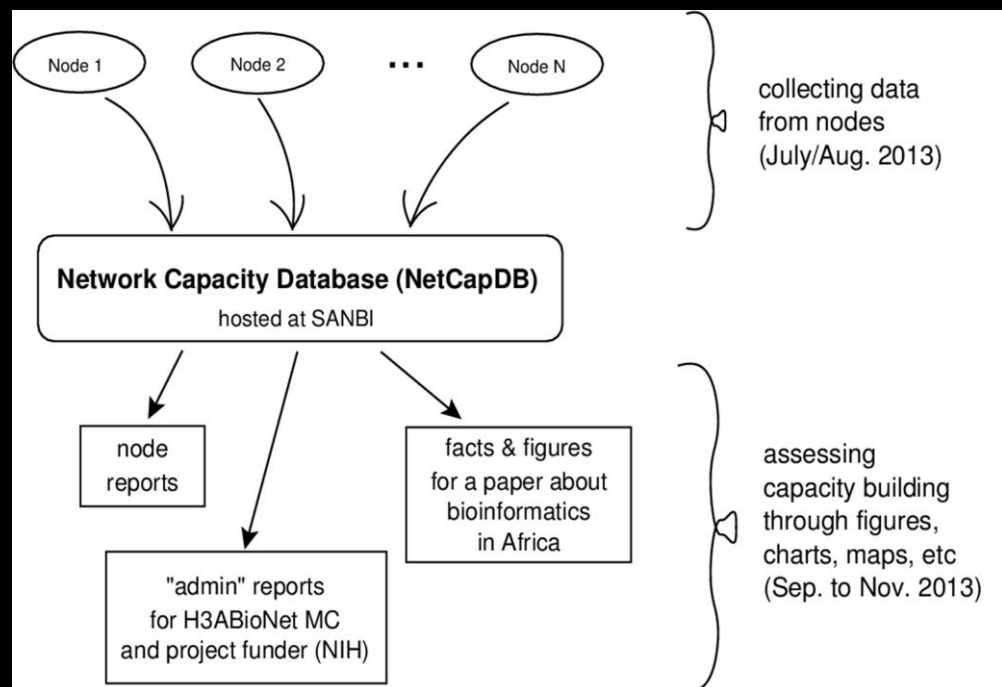
Schematic of NetCapDB anticipated data entering timelines and report outputs.

How can the H3ABioNet Nodes get involved?

Download and have a look at the NetCapDB data entry checklist available at: http://www.h3abionet.org/attachments/article/9/NetCapDB_data_entry_chec klist.pdf and start preparing / delegating members of your Node to accumulate the information required. Also expect a User Name / Login password combination to be sent to each Node within two weeks for data entry as we move from testing to production phase.

In the interim, a NetCapDB user guidelines document is being devised and if any of the Nodes do have any queries they can be directed to the H3ABioNet helpdesk by logging into the H3ABioNet website, selecting the Helpdesk Dashboard, Submit a new ticket and under Classification, choose NetCapDB from the categories.

Dr. Jean–Baka Domelevo Entfellner.

- **H3ABioNet Systems Administration Workshop - Pretoria 2013**

H3ABioNet Systems Administration Workshop - Pretoria 2013

The H3ABioNet Systems Administration workshop was attended by 18 participants from 11 different African countries. The participants covered a variety of topics which included Scientific Linux installation and Systems Administration, Galaxy installation and configuration, cloud and high performance computing, virtual machines, data management and security, genome databases and browsers. The different modules were taught by Peter van Heusden, (South African National Bioinformatics Institute), Prof. Fourie Joubert (University of Pretoria), Christopher Fields and Daniel LaPine (University of Illinois). All participants found the quality of teaching to be high and the biggest plus of all was the ability to interact and network with their peers. Hence, the techie mailing list that was created for the course has been left active to provide a forum for all participants to communicate and interact with each other.
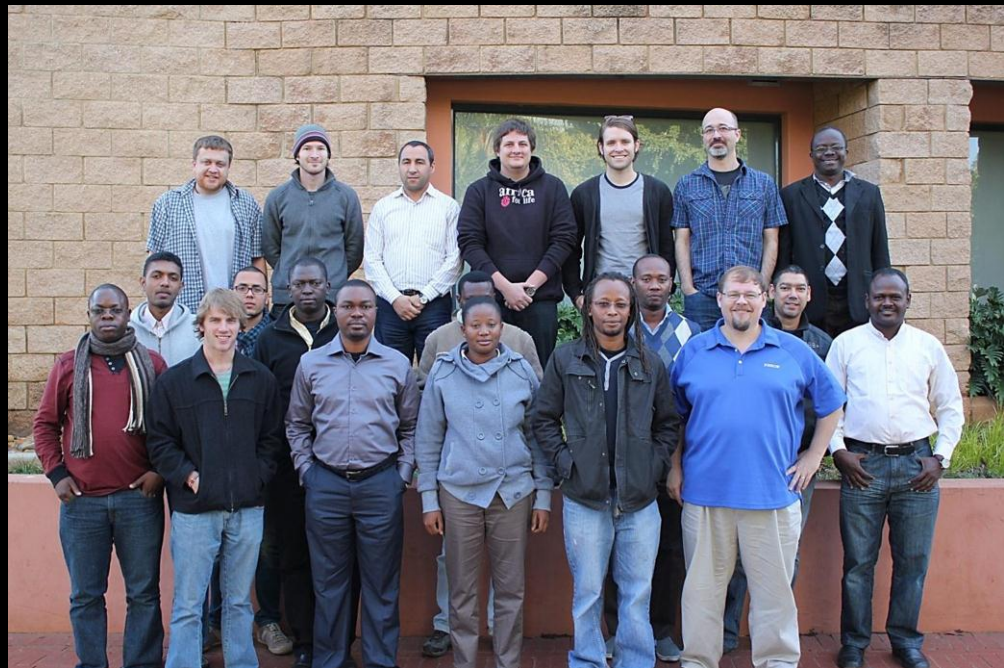
Left to right, **Top Row:** Ayton Meintjes, Jon Ambler, Tayalati Yahya, Burger van Jaarsveld, Timothy Reid, Christopher Fields, Timothy Kimbowa Wamala
Left to right, **Middle Row:** Olivier Sheik Amamuddy, Mohamed Alibi, Alio Sanda Abdel Kader, David Indome, Kelvin Kiritta, Suresh Maslamoney, Mutawakil Saad
Left to right, **Bottom Row:** Benjamin Kumwenda, David Brown, Oladipo Olaleye, Flora Mlyango, Frederick Mbuya, Daniel LaPine

- ## H3Africa Consortium Meeting - Accra, Ghana 2013

H3Africa Consortium Meeting - Accra, Ghana 2013

The second H3Africa consortium meeting was jointly held together with the African society for human genetics (ASHG) conference in Accra, Ghana on 16-18 of May. The H3Africa meeting was funded by the National Institute of Health (NIH) and Wellcome Trust and well attended by participants from Wellcome Trust Research Projects, NIH Pilot Biorepository Projects, NIH Research Projects, NIH Collaborative Center Research Projects, NIH Bioinformatics Network, Independent Expert Committee Members, NIH and Wellcome Trust Staff. Overall the meeting was a success! All the H3Africa consortia working groups met and presented their work, followed by open discussions. The second day of the meeting biorepository, data sharing and access and ethics work group presented consolidated recommendations. The overall aims and objectives of the meeting were:

- Update the H3Africa Consortium, its Independent Expert Committee and NIH and Wellcome Trust staff about the progress of the H3Africa grants and projects and for awardees to discuss specific issues with their funders.
- Provide the H3Africa Working Groups (WGs) an opportunity to meet face-to-face to finalize their deliberations.
- Share start-up experiences, successes, and lessons learned across projects.
- Provide an opportunity for the H3Africa working groups and Steering Committee to discuss and develop policies, materials and recommendations to standardize and support activities across the H3A consortium.
- Identify opportunities for collaborations across the initiative and discuss how to build research capacity across the initiative and access new opportunities.

Group picture of H3Africa Consortium Meeting Attendees.

Dr. Judit Kumuthini.

• Announcements

Announcements

- H3ABioNet has joined the Global Alliance to "enable the sharing of genomic and clinical data to accelerate progress in medicine. Over 70 leading health care, research, and disease advocacy organizations that together involve colleagues in over 40 countries have taken the first steps to form an international alliance dedicated to enabling secure sharing of genomic and clinical data. The cost of genome sequencing has fallen one-million fold, and more and more people are choosing to make their genetic and clinical data available for research, clinical, and personal use. This alliance is now in the beginning stages of formation. Following the circulation of a White Paper, over 70 organizations from Africa, Asia, Australia, Europe, and North and South America, have joined together to form a non-profit global alliance that will work to develop this common framework, enabling learning from data while protecting participant autonomy and privacy." For more information on the Global Alliance please see: http://www.h3abionet.org/h3abionet-joins-the-global-alliance

- Thanks to H3ABioNet, one of many great outcomes is the establishment of an Egyptian Center of Bioinformatics and Genomics (ECBAG), in Egypt. The Center established in Zagazig University started in 2012 with only three members (PI, Co-PI and MSc Student). Within a few months, ECBAG got more funding from Zagazig University through other projects to get 6 more IT specialists, 1 PhD and 3 BSc Bioinformatics students. They have also been provided with suitable infrastructure and furniture from the University, and have produced one publication and developed a new software tool (Gene Tracer). In addition, they are developing cooperation with many African countries from within the H3ABioNet. They are currently going through accreditation from the supreme council of Universities in Egypt and have many official visits from different scientific sectors in the council (below). ECBAG members are participating actively in all the training courses and meetings of H3ABioNet.

H3ABioNet Montly Newsletter Issue 2 ; June 2013

H3ABioNet

- **Announcements**

- **Important Dates**

- The CPGR node has initiated the new Knowledge Transfer Programme (KTP) which is open to all H3Africa consortium members. It has recruited various scientific members of H3ABioNet and other attendees as part of the scientific advisory and review committee. For more information on the KTP please contact Dr. Kumuthini (Jkumuthini@gmail.com).

Important Dates

- 15th of July, 2013 – Closing date for Applications for the Visual Analytics of Human Genome Variation Datasets workshop, NABDA, Abuja, Nigeria.

- 8th – 26th of July, 2013 - H3ABioNet Train the Trainer Bioinformatics Workshop, ICIPE, Nairobi, Kenya.

- 29th July – 2nd August, 2013 - Introduction to Bioinformatics using eBioKits, ICIPE, Nairobi, Kenya.

- 1st - 2nd of October, 2013 - H3ABioNet Annual Meeting, Johannesburg, South Africa

- 3rd - 5th of October, 2013 - H3Africa Consortium Meeting, Johannesburg, South Africa

- 3rd - 5th of October 2013, South African Society for Human Genetics Conference, Johannesburg, South Africa