H3ABioNet

Pan African Bioinformatics Network for H3Africa

Issue 19: October 2016

H3ABioNet

In This Issue

Continue the conversation:

#bioinformatics #Africa
#H3ABioNet @H3ABionet

Foreword

Welcome to the final newsletter for 2016. The year has flown by and we will rapidly be into our last 6 months of the project for the current funding period! Now that we have recovered from submission of our 857 page renewal proposal for the H3Africa Informatics Network, we can focus on achieving the final deliverables for the current project. This involves wrapping up a number of projects and packaging the outputs to disseminate the products to make them more accessible in a user friendly manner. As an example, for training materials we are curating these and making them more consistent and easily searchable, as discussed in the training report below.

The ISWG has continued to work on key infrastructural projects which are enabling rapid data transfer and facilitating submission of data to the archive and subsequently to the public repositories. We should now be better prepared for the deluge of H3Africa data that is being generated. The Research Working Group continues to generate new ideas for collaboration and sharing of knowledge. Some examples include the Bioinformatics Research Directory and COSI/SIG representatives discussed in their report below. The IBT course is now complete, congratulations to Kim and team! However much work remains to assess the evaluations, review the course materials and start preparing to run it again in 2017!

Finally, the newsletter reports on the H3ABioNet/IBM Research Africa Hackathon on malaria drug resistance for a DREAM challenge. I attended the first day of this hackathon and was again impressed by the enthusiasm and commitment of participants. Bringing scientists with different skills together to work on a common project drives innovation and skills transfer and also achieves a significant outcome in terms of analysis done on a project. We plan to use this kind of forum more often in the future.
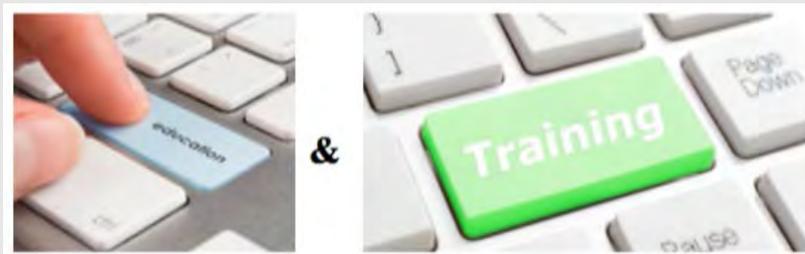
It remains for me to wish you well over the festive season, enjoy reading our final newsletter for 2016!

Issue 19: October 2016

H3ABioNet

Education and Training Working Group

Education and Training Working Group

As 2016 winds down, we reflect on the last workshop planned for the year, which was presented in collaboration with the International Network for Data Analysis (INDA) at the Institut Pasteur in Senegal at the beginning of October. While we mentioned this workshop in last month's newsletter, we have been sent some photos of the event and couldn't pass up the opportunity to share them with you all (please see on next page). Briefly, the workshop brought together participants from various backgrounds interested in learning more about NGS statistics and data analysis. There was a great interest in the first week of the course, which dealt with the theoretical aspects of NGS data analysis, while the second week focused on practical hands-on data analysis sessions. Almost 100 participants attended the first week of the course, while 24 selected participants stayed on for week two. We are glad to have hosted this workshop in collaboration with INDA and are planning to host future workshops with the network.

As mentioned in last month's newsletter, a task force has been established to curate and review all H3ABioNet training material. The task force has begun the process of reviewing the material available for each of the training courses and tagging the courses with *EDAM ontology* terms. EDAM is a simple ontology of well established, familiar concepts that are prevalent within bioinformatics, including: types of data and data identifiers, data formats, operations, and topics. In addition to attaching standardised descriptors to each course, this process will also help us to determine if there is material missing for any of the courses and to standardise the format and presentation of the material for each course. Once this step is completed, each course will be entered into the *eGenomics catalogue* to allow for the courses to be searchable, via the catalogue, based on specific descriptors.

Another project that the working group is currently tackling is the review and update of the MSc curriculum development resource that was developed in 2014. The MSc curriculum development resource is a detailed collection of information of suggested core and elective modules for an MSc degree programme in bioinformatics. The original task force did an excellent job in putting the resource together, however, it has not been updated or curated since the first version was completed. The task force has been reconvened to complete this task and has already begun working on updating the resource. This is a valuable and useful resource that has already been used as the basis for developing a bioinformatics MSc curriculum, as well as the development of other H3ABioNet course content, so we would like to ensure that it is a high quality up-to-date resource.

Continue the conversation:

#bioinformatics #Africa
#H3ABioNet @H3ABionet

Issue 19: October 2016

H3ABioNet

Education and Training Working Group

Collaboration at its best

Group photo of the INDA NGS data analysis course participants

#H3ABioNetEducationAndTraining

Nicky Mulder and Shaun Aron

Continue the conversation:

#bioinformatics #Africa
#H3ABioNet @H3ABionet

Back to Contents

Issue 19: October 2016

H3ABioNet

Infrastructure Working Group

Infrastructure Working Group

The ISWG kept itself busy in October with steady, if low profile work:

- Data Management Task Force: Good progress is being made on the design and implementation of the data archive. SOPs are being refined for the submission of data from H3Africa research groups to H3ABioNet and onward submission to the EGA. One area we need to work on is our link to the H3Africa Biorepositories LIMS group.

- The H3Africa Recruitment Database was used successfully for a third round of data collection in order to monitor progress of recruitment of H3Africa projects. The funders have expressed their gratitude for this work done. Congratulations to Mamana Mbiyavanga and the team.

- System Administration Task Force: The Globus Online roll out has proved its success. We have successfully transferred many hundreds of terabytes of data. This has been particularly important for the work on the H3Africa Chip design and for the genome analysis working group. Adoption by new sites of Netmap and Globus Online has slowed down, but we encourage sites who do not yet have GO/Netmap to contact us - we are always willing to help. We are cleaning up and refining our our system administration documentation, and testing the new eBiokit images.

- Cloud Task force: After the hackathon, our pipeline teams are busy refining and completing their pipelines. We hope that the pipeline workflows will be in a publishable form by the end of November.

We always welcome new members to join us!

#H3ABioNetInfrastructure

Scott Hazelhurst and Suresh Maslamoney

Back to Contents

Continue the conversation:

#bioinformatics #Africa
#H3ABioNet @H3ABionet

H3ABioNet

Research Working Group

Research Working Group

As new members joined the Research Working Group (RSWG) due to the recent rescheduling of the working groups, we started the October meeting with an overview of RSWG milestones and deliverables for the 5th year in order to provide all the new-comers with a solid understanding of the vision and mission of the working group so that they may know exactly which activities they could be involved in or volunteer for.

During the month of October, the H3ABioNet webinars series covered one topic and the theme of the month was "Cloud Computing Pipelines relevant to Bioinformatics". The presenter of the seminar was Dr Liudmila Sergeevna Mainzer who is a Senior Research Scientist at the National Center for Supercomputing Applications and a Research Assistant Professor at the Institute for Genomic Biology, both at the University of Illinois at Urbana-Champaign, USA. Dr Mainzer's talk was entitled "Methods of measuring performance of Bioinformatics software on HPC infrastructure". This interesting topic and well-articulated presentation attracted around 23 audience members who attended the seminar from beginning to end. In this talk Dr Liudmila explored methods for measuring, optimizing and monitoring performance of bioinformatics software on computer clusters.

Dr Liudmila Sergeevna Mainzer

Continue the conversation:

#bioinformatics #Africa #H3ABioNet @H3ABionet

H3ABioNet

Research Working Group

The RSWG has also discussed methods to evaluate the H3ABioNet webinars. There are many key metrics that the Seminars Series Task Force needs to introduce in order to monitor audience members so that we may determine how to attract a greater audience and get them to register and attend the webinars. But, before we do this we need to create a webinar assessment baseline from the existing situation in order to assess the webinars performance going forward. It's worth mentioning that a significant improvement in terms of attendance has been noticed since the last re-organization of the seminars.

An announcement and call for contributions to the H3ABioNet Bioinformatics Research Digest (BioRD) was sent to all members of the H3ABioNet consortium. The announcement explained the background, approach and important deadlines. Briefly, BioRD involves the compilation of a list of the recent bioinformatics articles pertinent to the H3ABioNet/ H3Africa projects. The Digest will publish a summary of selected articles as the BioRD newsletter on a quarterly basis. We are excepting the first edition to be published before the end of the year.

An announcement and call for contributions and volunteers was sent to all members of the H3ABioNet consortium seeking staff and graduate students, who are keen and active, to represent H3ABioNet at the Communities of Special Interests/Special Interest Groups (COSIs/SIGs) within the Intelligent Systems for Molecular Biology (ISMB) and International Society of Computational Biology (ISCB). As these members will represent H3ABioNet, we thought that we should select the best candidates for this programme. Therefore, we requested every volunteer to provide us with a written strategy note (half page) explaining why s/he thinks s/he is the best candidate and what approach s/he proposes to contribute to this activity. The chairs of RSWG, with the support of the H3ABioNet Management Committee members, will assess and evaluate the submitted written strategy notes thoroughly and will follow pre-set criteria in this process. Subsequently, the chairs will nominate the best candidates to be the H3ABioNet representatives for the various COSIs/SIGs.

#H3ABioNetResearch

Faisal Fadlelmola and Amel Ghouila

Back to Contents

Continue the conversation:

#bioinformatics #Africa
#H3ABioNet @H3ABionet

Issue 19: October 2016

H3ABioNet

H3Africa Participant Recruitment Database

H3Africa Participant Recruitment Database

To minimize delays in participant recruitment for H3Africa projects, we have designed a participant recruitment database, *H3APRDB*, to help monitor the progress of participant recruitment in the consortium. Recruitment of the target number of participants for a clinical study is a critical step to effectively answer important questions that may lead to a better understanding of disease and an improvement of public health. Failure to achieve the targeted enrolment and retention rates can have negative scientific and financial impacts. For example, studies that fail to achieve the target number of participants or that fail to recruit participants in a timely manner can suffer from reduced statistical power and increased false positive rates. It can also lead to premature termination of the study with financial implications both in terms of the money and efforts invested by the investigators and to participants enrolled in the study. As such, the H3APRDB is an important tool for the consortium.

The H3APRDB is a reporting interface designed for H3Africa principal investigators (PIs) and funders to track participant recruitment at the site, project, and consortium levels. It also provides built-in graphical output summaries, allowing stakeholders to conveniently assess progress of enrolment and retention of their target participants as well as identify problem areas.

The first component of the H3APRDB is the reporting module, which is built in *REDCap* and takes advantage of REDCap features such as survey login, survey queue, automated survey invitations, and reminders. Project-specific forms are designed in REDCap in order to track key metrics (e.g., participants recruited, samples collected, gender, case/control, HIV status, etc.). For each project, the total number of recruitments at each site is collected with 4-month intervals, starting from November 2015. The second component of H3APRDB is the dashboard, which is a very simple interface that allows both PIs and funders to visualize, in a very simplistic but user-friendly and intuitive fashion, the progress achieved, with regards to recruitment of participants, in their project or of the projects they fund.

The H3APRDB is developed and supported as a collaborative project with extensive inputs from various H3ABioNet nodes including CBIO, CPGR, Wits, CUBRe, and IPT. Though no sample details are collected and retained, the H3APRDB provides brief information regarding the location and informative descriptors of recruited participants. Thus, using the H3APRDB, one can request the number of case/control participants collected amongst all the projects in a particular country or region in Africa, for example. This can highly increase the reusability of data, for example, a project might invest in collecting case participants in a region where control participants are already being collected by another project.

Continue the conversation:

#bioinformatics #Africa
#H3ABioNet @H3ABionet

H3ABioNet

H3Africa Participant Recruitment Database

H3APRDB reporting interface built as a survey using REDCap

H3APRDB dashboard showing the progress of participant recruitment of all H3Africa projects. The number of participants expected, the number of participants recruited, and the percentage of recruitment are also provided

#H3APRDB

Mamana Mbiyavanga

Continue the conversation:

#bioinformatics #Africa
#H3ABioNet @H3ABionet

IBT corner

October 10th marked the end of 3 months of intensive training with the Introduction to Bioinformatics (IBT) course. We are happy to report that 195 participants (52% of total enrolled) met all the course requirements (submit at least 90% of practical assignments and obtain a minimum of 60% average for the assessments overall) and received the letter of completion. The average pass grade for the assessments was 75.3% +-8.2%. While roughly half of enrolled participants stuck it out for the entire course, many participants honed in on the modules that interested them. Typically for each assignment we received 230-280 submissions indicating that the vast majority of participants engaged with the course content.

While retention of participants and number of 'passes' is an obvious measure of impact, it is not necessarily the most useful or most informative. We also employed self-rated measures to determine impact - we asked a series of questions relating to self-measured confidence with certain basic bioinformatics skills both before the course began and after each respective module. For all modules, we found that participants were more confident in their bioinformatics abilities after each module (see below for a few examples of the trend).
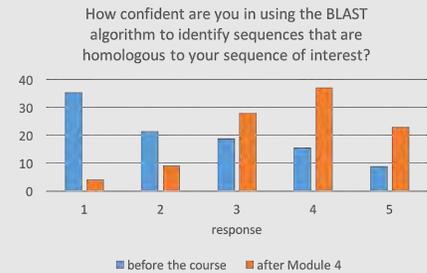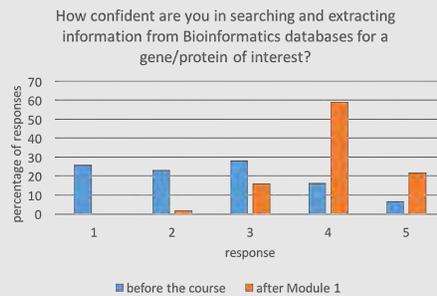
**'On a scale from 1 (not at all confident) to 5 (very confident), how confident are you in…'
Module 1: Databases and Resources, Module 4: Sequence Alignment Theory and Application**

We are in the process of analyzing the course feedback and can share that overall the response to the course has been very positive:

We need more and more courses like these, they are so crucial and thank you.

The course is really informative and provides a good foundation one can build up on as he/she advances in their bioinformatics analyses.

I enjoyed the course and value the skills I have learned. Overall I believe it was well worth attending and will advise others to do so next year if it runs again.

Read more about the IBT participants and volunteer IBT staff on the IBT course website.

#IBT_2016

Kim Gurwitz

Back to Contents

H3ABioNet

H3ABioNet / IBM Research Africa Hackathon on Malaria Drug Resistance - DREAM Challenge

H3ABioNet / IBM Research Africa Hackathon on Malaria Drug Resistance for a DREAM Challenge

H3ABioNet, jointly with the University of Notre Dame and IBM Research Africa, organized the first scientific hackathon on Malaria Drug Resistance for a DREAM challenge (6th-10th September 2016) in Johannesburg, South Africa.

The main aim of this hackathon was to pre-analyze a genomic dataset, comprising a range of malaria parasites, that had been generated at the University of Notre Dame (USA), and to establish an analysis baseline for a DREAM challenge that will be opened to the international community in mid-2017. This later challenge will focus on the identification of the next potential drug combinations for Malaria treatment.

Hackathon participants
During this hackathon, ten H3ABioNet participants were selected from the different H3ABioNet nodes. They teamed up with scientists and engineers from IBM research Africa and worked on analyzing the preliminary microarray dataset. Various African countries were represented during this hackathon including: Egypt, Ghana, Malawi, Morocco, Nigeria, Senegal, South Africa, Sudan, Tanzania and Tunisia. The hackathon was hosted by IBM Research Africa at their friendly work premises in Johannesburg. This environment stimulated the participants who were motivated to work until late in the evenings.

Work organization during the Hackathon
Three heterogeneous teams (named Snakes, Tigers and Monkeys), with diverse and complementary areas of expertise, were formed (as shown in the figures below). Three researchers from Notre Dame University, who were involved in the dataset generation, worked and interacted with the different teams.

Teams Tigers and Monkeys working on their plans

Continue the conversation:

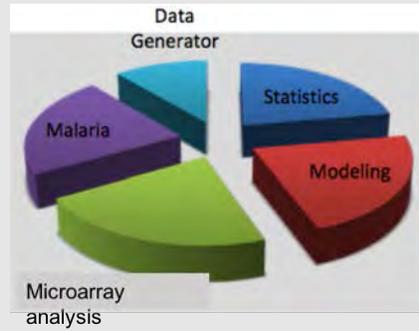#bioinformatics #Africa #H3ABioNet @H3ABionet

H3ABioNet

H3ABioNet / IBM Research Africa Hackathon on Malaria Drug Resistance - DREAM Challenge
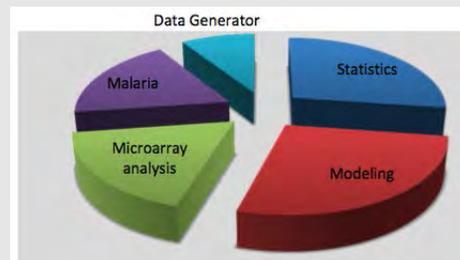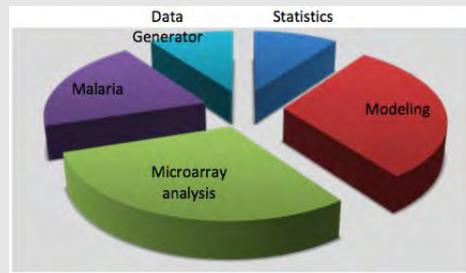
Continue the conversation:

#bioinformatics #Africa #H3ABioNet @H3ABionet

Team Monkey's expertise represented as a pie chart

Team Snake's expertise represented as a pie chart

Team Tiger's expertise represented as a pie chart

Pre-hackathon online meetings took place to get the discussion started. Each team outlined an initial roadmap and worked together on the different steps of the analysis during the hackathon itself. During the first day of the hackathon, the data generators from University of Notre Dame gave a detailed presentation about the dataset, the background, and then went through a quick literature review about Malaria drug resistance. During the following days, each team progressed on their suggested analysis plan before reporting back to the whole group. This generated interesting open discussions that allowed the teams to get constructive feedback in order to refine their plans after these report-back sessions.

Issue 19: October 2016

H3ABioNet

H3ABioNet / IBM Research Africa Hackathon on Malaria Drug Resistance - DREAM Challenge

#Monkeys , #Tigers, and #Snakes reporting back to the whole group at the end of Day 3 of the hackathon

Hackathon outcomes

This hackathon provided an ideal environment for generating robust ideas and transferring skills between team members. The different teams had the chance to explore, in parallel, different methods and tools. This led to a deep pre-analysis of the data in a very short period of time by 20 people from different backgrounds. It also offered the data generators valuable feedback that will be of great use for refining the data generation process and for releasing datasets of good quality for the international crowdsourcing challenge. Data generators typically don't have access to this kind of feedback about the quality of the data.

In the hackathon evaluation survey, the majority of the participants highlighted the fact that they enjoyed working in multi-disciplinary teams and that they learned many skills thanks to this experience.

Anita, a research fellow with a biology background *"liked the exposure of this kind of data and having to manipulate new software and packages."*

Katie, a PhD student involved in the data generation process commented on *"the ability to work with so many analysis experts for such an extended length of time."*

Continue the conversation:

#bioinformatics #Africa #H3ABioNet @H3ABionet

Hackathon participants and organizers at IBM Research Africa, WITS, Johannesburg, September 2016

H3ABioNet

H3ABioNet / IBM Research Africa Hackathon on Malaria Drug Resistance - DREAM Challenge

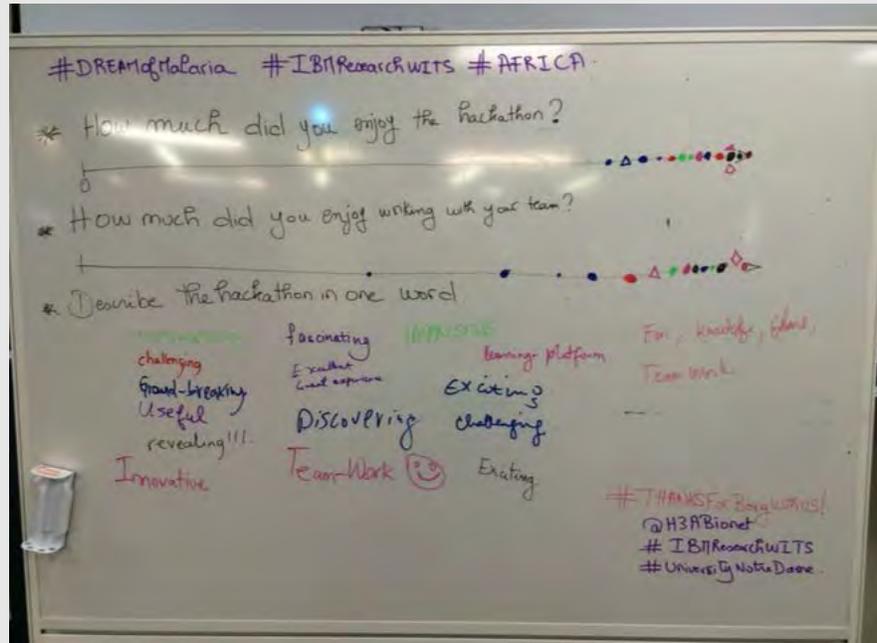**Participant feedback**

During the last day of the hackathon, we asked the participants to rate how much they liked the hackathon and how much they enjoyed working with their team. We came up with the following evaluation:

Participant feedback

Take home message

This hackthon allowed us to create an environment for open innovation and learning - gathering scientists from all over Africa to make attempts and efforts towards solving an important scientific challenge of high positive impact on the continent.

A *blog post* about this hackathon was published by IBM Research.

#DREAMofMalaria #OpenScience #IBMResearchAfrica

Amel Ghouila, Jean-Baka Domelevo Entfellner, Sumir Panji, Faisal M. Fadlelmola, Nicola Mulder and Geoffrey Siwo

Back to Contents

Continue the conversation:

#bioinformatics #Africa
#H3ABioNet @H3ABionet

H3ABioNet

Collaborating to Strengthen the Research Workforce of Tomorrow

Collaborating to Strengthen the Research Workforce of Tomorrow

The H3ABioNet Node at the University of Illinois, together with the KnowEnG Big Data to Knowledge Center at the Carl R. Woese Institute for Genomic Biology (IGB), is partnering with Fisk University in Nashville, Tenn., in an innovative program to promote diversity in the biomedical, behavioral, and clinical research workforce. The collaboration aims to prepare underrepresented minority undergraduates for entry into competitive M.D. or graduate programs by teaching them to apply computational thinking and statistical and informatics tools to address biomedical research challenges.

KnowEnG is one of 11 NIH-funded Big Data to Knowledge (BD2K) Centers. NIH funds collaborations between institutions with BD2K centers and institutions serving students from backgrounds that are underrepresented in research. These collaborations allow students to have learning and research experiences at the intersection of computer science and biology, biochemistry, molecular biology, and mathematics.

Fisk is a natural fit for this program. As Lee Limbird, Dean of Fisk's School of Natural Sciences, Mathematics and Business explained, *"Fisk has adopted a strategic STEM research and research training development plan that affirms the cross-disciplinary and enhanced quantitative and computational components of contemporary discovery. We have intentionally fostered an interdisciplinary research culture that crosses all of our STEM areas of biology, chemistry, biochemistry and molecular biology, computer science, mathematics, and physics/materials science."*

The partnership extends Fisk's bioinformatics course offerings by broadcasting bioinformatics seminars taught at the U. of I. to Fisk students and faculty. Training tools developed by the BD2K center are being integrated in Fisk undergraduate courses and laboratories.

"Collaborating with UIUC ... will give our students a chance to do research with top-notch scientists ... The program will also improve biology, chemistry, computer science and mathematics curricula at Fisk and enhance research collaboration between Fisk and Illinois faculty," said Fisk computer scientist Lei Qian.

The Fisk students work on research projects during the academic year under the mentorship of faculty whose research aligns with the BD2K initiative. They also have the opportunity to do a summer internship with the U. of I. High Performance Biological Computing (HPCBio) group, which forms the core of the Illinois Node. Next year, Mayo Clinic, a KnowEnG partner, will also offer internships.

Carleigh Frazier and Shelby Clark were the first two Fisk interns to come to the U. of I.. Frazier hopes to be a dermatologist and do research in her own practice. Clark wants to be an obstetrician/gynecologist helping women in developing countries. They both believe that what they learn through the program will help them in their future careers.

Frazier and Clark also participated in the U. of I. Graduate College's Summer Research Opportunity Program (SROP), which helps students from populations underrepresented in graduate study to explore careers in research. SROP activities include writing workshops, GRE prep, and participation in research teams.

Fisk faculty attended summer short courses in data carpentry and computational genomics at U. of I., allowing them to establish collaborations that will provide them and their students with research opportunities, foster relationships, and work together on innovations in course design.

Ultimately, the collaboration partners hope to increase diversity in Big Data biomedical research discovery through an evidence- and experience-driven program that will prepare professionally confident minority trainees.

#collaboration #bioinformaticstraining

Victor Jongeneel

Continue the conversation:

#bioinformatics #Africa
#H3ABioNet @H3ABionet

H3ABioNet

Upcoming Events

Upcoming Events

- **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.

- **March 23rd to March 25th 2017**: *SMC'2017: Data Engineering In Bioinformatics, Image and Data Analysis*, organized by the Moroccan Classification Society, will take place in Tangier, Morocco.

- **April 2nd to April 12th 2017**: *1st Big Data Africa Summer School*. The school will focus on practical Data Science Applications to Astronomy and Bioinformatics and will take place in Cape Town, South Africa.

- **June 4th to June 16th 2017**: *EMBO Practical Course in Bioinformatics and Genomics*, will take place in Thermi, Greece.

- **July 19th to July 21st 2017**: *NextComp 2017: Next Generation Computing Application Conference*, will take place in Republic of Mauritius.

- For a comprehensive list of bioinformatics and genomics conferences, please consult: *Conference service - Bioinformatics*

Any upcoming event notifications may be sent to:

kim.gurwitz@uct.ac.za AND sumir.panji@uct.ac.za

We look forward to updating you on further H3ABioNet happenings at the beginning of 2017. The Working Group schedules will also be included in the first newsletter of 2017.

For now, we wish you well over the festive season. Happy holidays!

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

Back to Contents

Continue the conversation:

#bioinformatics #Africa
#H3ABioNet @H3ABionet