

Development of Bioinformatics Infrastructure for Genomics Research in H3Africa

Nicola J. Mulder^{*}, Ezekiel Adebisi^{†,‡}, Marion Adebisi^{†,‡}, Seun Adeyemi^{†,§}, Azza Ahmed^{||}, Rehab Ahmed^{||}, Bola Akanle^{‡,§}, Mohamed Alibi[¶], Don L. Armstrong[#], Shaun Aron^{**}, Efejiro Ashano^{‡,††}, Shakuntala Baichoo^{‡‡}, Alia Benkahla[¶], David K. Brown^{§§}, Emile R. Chimusa^{*|||}, Faisal M. Fadlelmola^{¶¶}, Dare Falola[‡], Segun Fatumo^{††}, Kais Ghedira[¶], Amel Ghouila^{###}, Scott Hazelhurst^{**}, Iunu Isewon^{‡,‡}, Segun Jung^{***}, Samar Kamal Kassim^{†††}, Jonathan K. Kayondo^{‡‡‡}, Mamana Mbiyavanga^{*}, Ayton Meintjes^{*}, Somia Mohammed^{||}, Abayomi Mosaku[‡], Ahmed Moussa^{§§§}, Mustafa Muhammd^{||}, Zahra Mungloo-Dilmohamud^{‡‡}, Oyekanmi Nashiru^{††}, Trust Odia[‡], Adaobi Okafor[‡], Olaleye Oladipo^{‡,|||}, Victor Osamor^{‡,‡}, Jellili Oyelade^{‡,‡}, Khalid Sadki^{¶¶¶}, Samson Pandam Salifu^{###,****}, Jumoke Soyemi^{††††}, Sumir Panji^{*}, Fouzia Radouani^{‡‡‡‡}, Oussama Souiai[¶], Özlem Tastan Bishop^{§§} : and The H3ABioNet Consortium, as members of the H3Africa Consortium

Cape Town, South Africa; Ota, Nigeria; Khartoum, Sudan; Tunis, Tunisia; Champaign, IL, USA; Johannesburg, South Africa; Abuja, Nigeria; Grahamstown, South Africa; Tunis-Belvédère, Tunisia; Chicago, IL, USA; Cairo, Egypt; Entebbe, Uganda; Tangier, Morocco; Omu-Aran, Nigeria; Rabat, Morocco; Kumasi, Ghana; Ilaro, Nigeria; and Casablanca, Morocco

ABSTRACT

Background: Although pockets of bioinformatics excellence have developed in Africa, generally, large-scale genomic data analysis has been limited by the availability of expertise and infrastructure. H3ABioNet, a pan-African bioinformatics network, was established to build capacity specifically to enable H3Africa (Human Heredity and Health in Africa) researchers to analyze their data in Africa. Since the inception of the H3Africa initiative, H3ABioNet's role has evolved in response to changing needs from the consortium and the African bioinformatics community.

Objectives: H3ABioNet set out to develop core bioinformatics infrastructure and capacity for genomics research in various aspects of data collection, transfer, storage, and analysis.

Methods and Results: Various resources have been developed to address genomic data management and analysis needs of H3Africa researchers and other scientific communities on the continent. NetMap was developed and used to build an accurate picture of network performance within Africa and between Africa and the rest of the world, and Globus Online has been rolled out to facilitate data transfer. A participant recruitment database was developed to monitor participant enrollment, and data is being harmonized through the use of ontologies and controlled vocabularies. The standardized metadata will be integrated to provide a search facility for H3Africa data and biospecimens. Because H3Africa projects are generating large-scale genomic data, facilities for analysis and interpretation are critical. H3ABioNet is implementing several data analysis platforms that provide a large range of bioinformatics tools or workflows, such as Galaxy, the Job Management System, and eBiokits. A set of reproducible, portable, and cloud-scalable pipelines to support the multiple H3Africa data types are also being developed and dockerized to enable execution on multiple computing infrastructures. In addition, new tools have been developed for analysis of the uniquely divergent African data and for downstream interpretation of prioritized variants. To provide support for these and other bioinformatics queries, an online bioinformatics helpdesk backed by broad consortium expertise has been established. Further support is provided by means of various modes of bioinformatics training.

Conclusions: For the past 4 years, the development of infrastructure support and human capacity through H3ABioNet, have significantly contributed to the establishment of African scientific networks, data analysis facilities, and training programs. Here, we describe the infrastructure and how it has affected genomics and bioinformatics research in Africa.

The authors report no relationships that could be construed as a conflict of interest.

H3ABioNet is supported by the National Institutes of Health Common Fund (grant number U41HG006941).

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

From the ^{*}Computational Biology Division, Department of Integrative Biomedical Sciences, Institute for Infectious Disease and Molecular Medicine, University of Cape Town, Cape Town, South Africa; [†]Department of Computer and Information Sciences, Covenant University, Ota, Nigeria; [‡]Covenant University Bioinformatics Research (CUBRe), Covenant University, Ota, Nigeria; [§]Center for System and Information Service, Covenant University, Ota, Nigeria; ^{||}Centre for Bioinformatics and Systems Biology, Faculty of Science, University of Khartoum, Khartoum, Sudan; [¶]Laboratory of Bioinformatics, Biomathematics and Biostatistics (BIMS), Institut Pasteur de Tunis, Tunis, Tunisia; [#]Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Champaign, IL, USA; ^{**}Sydney Brenner Institute for Molecular Bioscience, University of the Witwatersrand, Johannesburg, South Africa; ^{††}H3Africa Bioinformatics Network (H3ABioNet) Node, National Biotechnology Development Agency (NABDA), Federal Ministry of Science and Technology (FMST), Abuja, Nigeria; ^{‡‡}University of Mauritius, Moka, Mauritius; ^{§§}Research Unit in Bioinformatics (RUBi), Department of Biochemistry and Microbiology, Rhodes University,

Grahamstown, South Africa; ||||Division of Human Genetics, Department of Pathology, Faculty of Health Sciences, Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Cape Town, South Africa;

¶¶Future University of Sudan, Khartoum, Sudan; ##Institut Pasteur de Tunis, LR11IPT02, Laboratory of Transmission, Control and Immunobiology of Infections (LTClI), Tunis-Belvédère, Tunisia;

***Computation Institute, University of Chicago and Argonne National Laboratory, Chicago, IL, USA; †††Medical Biochemistry and Molecular Biology Department, Faculty of Medicine, Ain Shams University, Abbaseya, Cairo, Egypt; ‡‡‡Uganda Virus Research Institute (UVRI), Entebbe, Uganda;

§§§LabTIC Laboratory, ENSA, Abdelmalek Essaadi University, Tangier, Morocco;

||||Center for System and Information Service, Landmark University, Omu-Aran, Nigeria; ¶¶¶School of Sciences, Mohammed V University of Rabat, Rabat, Morocco;

####Department of Biochemistry and Biotechnology, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana; ****Kumasi Centre for Collaborative Research, South End Asougya Road, KNUST Campus, Kumasi, Ghana; ††††Department of Computer Science, Ilaro Polytechnic, Ilaro, Nigeria; and the ††††Chlamydiae and Mycoplasma Laboratory, Institut Pasteur du Maroc, Casablanca, Morocco. Correspondence: N.J. Mulder (Nicola.mulder@uct.ac.za).

GLOBAL HEART
© 2017 World Heart Federation (Geneva). Published by Elsevier Ltd. All rights reserved.
VOL. ■, NO. ■, 2017
ISSN 2211-8160/\$36.00.
<http://dx.doi.org/10.1016/j.gheart.2017.01.005>

Africa is currently undergoing an epidemiological transition, with endemic infection and a rapidly growing burden of cardiometabolic and other noncommunicable diseases. Understanding the genetic determinants of diseases can lead to novel insights into disease etiology, which may identify novel therapeutic targets, and the potential for better disease prognosis and management. Genomics research holds great promise for medical and health care research and is gaining global momentum as we transition to an era of precision medicine, whereby treatment of individual patients is driven by a greater understanding of the clinical diagnosis through interpretation of underlying genomic variation. With sequencing costs dropping, whole genome sequencing as an aid to diagnosis is becoming more affordable. However, this does not consider the hidden costs required for analysis and interpretation of the data, which is substantial [1]. The decreasing costs associated with next-generation sequencing (NGS) technologies have been accompanied by increasing size and complexity of the sequence data. To deal with such complex and voluminous data, existing data storage and transfer mechanisms as well as public repositories and data processing technologies have had to adapt, and skills in data science have had to be developed [1]. As a consequence, bioinformaticians have become essential to biomedical research projects. This is also true of other large-scale technologies, such as genotyping by arrays, which may not generate quite the same data sizes as NGS, but with data for millions of single nucleotide polymorphisms (SNPs) being generated and analyzed, the processing and downstream analysis require substantial computing resources and associated skills. Muir et al. [1] describe 4 key adaptations that have been required for embracing the genomics era, particularly in relation to NGS data: development of algorithms to handle short reads and long reference genomes, new compression formats for facilitating efficient data storage, adoption of distributed and parallel computing, and increased data security protocols. Bioinformaticians with data science skills are essential for processing, analyzing, and integrating genomics data, but there is also a need to train geneticists and clinicians to interpret and translate the results.

Although large-scale genomics projects have been undertaken for several years internationally, Africa has lagged behind due to limited infrastructure for implementing such large projects. Initiatives such as the H3Africa (Human Heredity and Health in Africa) (<http://h3africa.org/>) [2] are accelerating genomics research on the continent by funding research as well as building capacity and infrastructure. One component of the H3Africa initiative is H3ABioNet [3], a pan-African bioinformatics network for H3Africa (<http://www.h3abionet.org>), which is building capacity on the bioinformatics front to enable genomics research on the continent. Here, we describe how this bioinformatics capacity development has been undertaken, demonstrated with examples and the potential effect on genomics research.

APPROACH AND IMPLEMENTATION

H3ABioNet is an extensive network covering over 30 institutions in 15 African countries and 2 partners outside of Africa, and it includes a large diversity of skills. However, the task of bioinformatics capacity development in most African countries is large, and due to limited resources, infrastructure, and expertise, requires careful coordination and pooling of efforts. With limited infrastructure in place prior to the H3Africa initiative, the network had to work on all fronts from genomics data management to building capacity for local analysis of the data.

Tools that are being developed and implemented are addressing the diverse bioinformatic needs of multisite clinical research projects. This includes tools for monitoring recruitment of participants, optimization of data transfer between project sites and to public repositories, harmonization of data, and establishment of standard and custom workflows for data analysis. H3ABioNet has made these available via several web-based, easy to use platforms such as REDCap, Galaxy, or WebProtégé. The network also provides access to experts from a variety of domains to address questions about the established infrastructure and to provide support to H3Africa and other genomics projects. The development of bioinformatics capacity and user support undertaken by H3ABioNet can be divided into several categories related to data transfer and storage; data collection, management, and integration; data analysis and development of associated tools; and training on all of these aspects. Further details about these developments are described in the following text.

GENOMICS SKILLS TRAINING

Genomics, being a multidisciplinary field, requires cross-disciplinary skills involving a combination of knowledge and proficiency from the fields of biology, computer science, mathematics, and statistics. There are numerous challenges in the field of genomics, including the major challenge of processing the massive amount of data and extracting biological meaning from them [4]. In view of these challenges, there is an increased demand for highly trained and experienced bioinformatics experts who can handle the data inundation as well as interface with biologists. Thus, at the core of enabling genomic research is the development of human capacity and the promotion of interdisciplinary training. H3ABioNet has undertaken a multifaceted approach to training, including integrating training with other activities, such as webinars, data analysis, or development hackathons, and the inclusion of shadow teams in projects. H3ABioNet delivers formal training through internships (enabling 1-on-1 skills transfer), short specialized courses, hackathons, and online distributed courses. The training has covered various aspects of bioinformatics from general introductory topics to specialized subjects such as NGS and GWAS analyses. Since 2013, H3ABioNet has conducted over 25 bioinformatics courses across Africa (<http://h3abionet.org/training-and-education/h3abionet-courses>). Follow-up surveys have

Download English Version:

<https://daneshyari.com/en/article/5602339>

Download Persian Version:

<https://daneshyari.com/article/5602339>

[Daneshyari.com](https://daneshyari.com)