



The Development of a Sustainable Bioinformatics Training Environment Within the H3Africa Bioinformatics Network (H3ABioNet)

Shaun Aron^{1*†}, Paballo Abel Chauke^{2†}, Verena Ras^{2†}, Sumir Panji^{2†}, Katherine Johnston² and Nicola Mulder² on behalf of the H3ABioNet Training and Education Work Package

¹Sydney Brenner Institute for Molecular Bioscience, University of the Witwatersrand, Johannesburg, South Africa,

²Computational Biology Division, Institute of Infectious Disease and Molecular Medicine, CIDRI Africa Wellcome Trust Centre, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa

OPEN ACCESS

Edited by:

Hugo Verli,
Federal University of Rio Grande do
Sul, Brazil

Reviewed by:

Peter Van Heusden,
University of the Western Cape,
South Africa
Erik Bongcam-Rudloff,
Swedish University of Agricultural
Sciences, Sweden

*Correspondence:

Shaun Aron
shaun.aron@wits.ac.za

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
STEM Education,
a section of the journal
Frontiers in Education

Received: 15 June 2021

Accepted: 27 August 2021

Published: 23 September 2021

Citation:

Aron S, Chauke PA, Ras V, Panji S,
Johnston K and Mulder N (2021) The
Development of a Sustainable
Bioinformatics Training Environment
Within the H3Africa Bioinformatics
Network (H3ABioNet).
Front. Educ. 6:725702.
doi: 10.3389/feduc.2021.725702

Bioinformatics training programs have been developed independently around the world based on the perceived needs of the local and global academic communities. The field of bioinformatics is complicated by the need to train audiences from diverse backgrounds in a variety of topics to various levels of competencies. While there have been several attempts to develop standardised approaches to provide bioinformatics training globally, the challenges encountered in resource limited settings hinder the adaptation of these global approaches. H3ABioNet, a Pan-African Bioinformatics Network with 27 nodes in 16 African countries, has realised that there is no single simple solution to this challenge and has rather, over the years, evolved and adapted training approaches to create a sustainable training environment, with several components that allow for the successful dissemination of bioinformatics knowledge to diverse audiences. This has been achieved through the implementation of a combination of training modalities and sharing of high quality training material and experiences. The results highlight the success of implementing this multi-pronged approach to training, to reach audiences from different backgrounds and provide training in a variety of different areas of expertise. While face-to-face training was initially required and successful, the mixed-model teaching approach allowed for an increased reach, providing training in advanced analysis topics to reach large audiences across the continent with minimal teaching resources. The transition to hackathons provided an environment to allow the progression of skills, once basic skills had been developed, together with the development of real-world solutions to bioinformatics problems. Ensuring our training materials are FAIR, and through synergistic collaborations with global training partners, the reach of our training materials extends beyond H3ABioNet. Coupled with the opportunity to develop additional career building soft skills, such as scientific communication, H3ABioNet has created a flexible, sustainable and high quality bioinformatics training environment that has successfully been implemented to train several highly skilled African bioinformaticians on the continent.

Keywords: LMIC, training modalities, Africa, bioinformatics training, professional development

INTRODUCTION

Bioinformatics training and education poses many challenges globally for several reasons; it is not a well-established subject that can easily be incorporated into traditional university curricular, it includes elements from multiple disciplines, application of bioinformatics requires practical skills, and the field is constantly and rapidly changing in response to the development of new data generating technologies and novel algorithms. There are also numerous different training audiences that need to acquire skills and knowledge, from wet-lab and other life science bioinformatics users to bioinformatics analysts and bioinformatics engineers, each of which require different levels of competencies (Mulder et al., 2018). Some scientists are even suggesting a nomenclature change from bioinformatics as it does not capture what the field entails (Bourne, 2021). The debate about labels is not within the scope of this paper as we present H3ABioNet's experiences in developing a holistic and diverse, multi-pronged training program in bioinformatics within an African environment. Bioinformatics skills are particularly lacking in low resourced settings, including many African countries. Ten years ago there were only a handful of institutions in Africa with bioinformatics infrastructure and skills, mostly in the middle income countries (Tastan Bishop et al., 2015). Therefore the continent was ill-equipped for a genomics revolution and to analyze and manage the ever-increasing availability of biological data. Though high-throughput data generation equipment is still limited on the continent, it is essential that data generated off samples derived from the continent be analysed by African scientists.

The Human Heredity and Health in Africa (H3Africa) Initiative is funded by the Wellcome Trust and National Institutes of Health from late 2012 to investigate the genetic and environmental basis for human diseases (Rotimi et al., 2014). With the promise of generating genomic data for thousands of African individuals, there was a need to develop bioinformatics capacity in Africa to manage, transfer, store, analyze and interpret the data in the continent. H3ABioNet, the Pan-African bioinformatics network for H3Africa was established in 2012 to develop the computing infrastructure, tools and skills required for this (Mulder et al., 2016; Mulder et al., 2017b). H3ABioNet developed a multi-faceted approach to address the vast bioinformatics training needs across diverse audiences in multiple African countries, each with different levels of infrastructure and expertise. To establish a sustainable bioinformatics infrastructure, it was necessary to build the underlying support mechanisms by training systems administrators in the effective use of computing infrastructures for bioinformatics applications, and local trainers and teaching assistants on how to run and support training interventions. In parallel, H3ABioNet focused on developing fundamental bioinformatics and computational skills in bioinformatics analysts and engineers, and in building introductory followed by specialized intermediate bioinformatics skills in the bioinformatics users. H3ABioNet also developed guidelines for establishing new bioinformatics degree programs at resource limited institutions and has supported the development of

such programs. Four years ago, the Fogarty International Center introduced a dedicated funding call through H3Africa for the establishment of bioinformatics postgraduate degree programs, enabling several new degree programs to be established in three African countries that had previously lacked these programs (Shaffer et al., 2019). H3ABioNet has continued to support the Fogarty postgraduate training program through providing trainer support and hosting a joint postgraduate seminar series to provide advice on student projects. H3ABioNet has subsequently focused on short term training interventions that enable professional development in specific topics outside of degree programs. This has been achieved using different modalities from face-to-face courses at various levels, hackathons, webinars and internships to mixed-model online teaching.

Success and Challenges With Traditional Face to Face Training

At the outset of the establishment of H3ABioNet, there was an initial need to develop a critical mass of skilled bioinformatics experts within Africa. While small pockets of skilled bioinformaticians did exist at some of the H3ABioNet nodes who had previously invested in bioinformatics skills development, the majority of the continent had little or no bioinformatics expertise. To tackle this initial lack of skills, a number of face-to-face training events were developed and run at the various nodes within the network across the African continent. Initially, international trainers were sourced where necessary to run earlier courses with the goal of developing local training expertise in specific data analysis topics. Some of these earlier training events took the form of a longer Postgraduate Introductory Bioinformatics workshop which covered different modules over 4–5 weeks to train bioinformaticians, and Train-the-Trainer events focused on developing regional training capacity. This formed the foundation to subsequently develop and host week-long workshops focusing on specialised bioinformatics analysis topics such as metagenomics, microbiome, genome-wide association studies (GWAS), data analytics and next generation sequencing (NGS). While one of the major focus areas of H3ABioNet was to develop bioinformatics capacity across the continent, it was soon realised that the design, planning and running of face-to-face workshops was not a sustainable long-term approach. Apart from the limited reach and immensely high costs associated with running face-to-face workshops, the challenges associated with working in an African context soon came to light.

While some of these challenges are specific to an African context, some may be encountered in other low-to-middle income countries (LMICs). The first challenge was internet access. While good internet connectivity is taken for granted as a necessity in developed countries, limited and slow internet access is a stark reality in most of Africa. Some universities have access to good local internet infrastructure, but issues with bandwidth and download speed restrictions are exacerbated by unexpected power outages and limited backup resources. H3ABioNet addressed this in some of the training courses by

running all practical analyses using the eBikit, which is a standalone Mac-based hardware resource that contains all the software and databases required for analyses, without the need for an internet connection to the outside world (Hernández-de-Diego et al., 2017). The second challenge, as already alluded to, is an unstable electricity supply, with planned and unplanned rolling blackouts a daily occurrence in several African countries. Slightly less common, but still observed were political unrest, visa bureaucracy, excessive costs for airline flights within Africa and endemic disease outbreaks. The disruptive nature of removing participants from their home institutes for extended periods of time coupled with participants not having access to local infrastructure to work on upon returning to their home institution, meant the face-to-face mode of training was not a long-term sustainable approach. This presented H3ABioNet with an opportunity to develop a new, more sustainable approach to build bioinformatics capacity across the African continent. Until the COVID-19 pandemic, these modes of training ran in parallel as the advantages of personal interaction, networking opportunities, community building and other benefits of face-to-face training are still important to exploit when feasible.

Need to Develop and Combine Alternative Modes of Training

Having built up a critical mass within the network, there was an opportunity to explore additional forms of training modalities. These training modalities and associated resources were borne out of necessity at the various stages of the development of the network and highlights the need for a flexible approach to developing capacity in any field. To complement the face-to-face training, H3ABioNet used training modalities that are targeted to different audiences with specific desired outcomes. Internships provide an opportunity for a researcher to be embedded in a host laboratory and work alongside experts to achieve a particular goal, usually to develop skills for analysis of their data. While the H3ABioNet internship program has been successful, it is a costly endeavour and is very limited in terms of reaching a wider audience. Therefore, H3ABioNet embarked on a new challenge to develop knowledge and skills in a much larger audience using an adaptation of traditional online courses. In our mixed-model approach, we developed courses running over a longer time period in an online format, but to multiple physical classrooms with local support. These have been introduced at both the introductory/overview and intermediate/specialized course levels and have been instrumental in addressing the enormous demand for accessible bioinformatics training across the continent.

The H3ABioNet training environment also includes longer term building of communities from the training courses, further career professional development in scientific communication and grant writing, and increased accessibility to our training resources by making them Findable, Accessible, Interoperable and Reusable (FAIR) (Garcia et al., 2020). Through a combination of all of these aspects of the training program, H3ABioNet has developed a sustainable training environment that has addressed many of the skills gaps on the continent with demonstrable impact on the

ability of African scientists to analyse their own data. In this paper we describe how H3ABioNet has developed and expanded its traditional training approaches to create a training environment suited to an African audience and aimed at developing an African bioinformatics community through the establishment of suitable training models and approaches. We further describe the components of this training environment, which has been developed to address multiple audiences and transfer multiple competencies to trainees based on their selected career paths. Finally, we present some results evaluating the impact and accessibility of our training program.

METHODS - DEVELOPING THE H3ABIONET TRAINING ENVIRONMENT

As previously mentioned, H3ABioNet uses a multi-pronged approach to training in both its target audiences and training modalities. The latter includes face-to-face, mixed-model online training and hackathons to enable the transfer of skills in various bioinformatics subject areas, systems administration and additional career development topics. We offer certificates of participation and acknowledgement for all of the workshops we host, and we have found that this encourages active participation and engagement through the value of adding these achievements to participants' CVs for job and fellowship applications.

Our course development uses competencies to drive the curriculum design and we strive to make the resulting training resources as accessible as possible through the curation of training materials. Below we provide further details about each of the modalities and the audiences and competencies that we aim to address with each one.

Switching to a Mixed-Model Training Approach

The need for basic bioinformatics skills in Africa is growing rapidly and H3ABioNet was faced with the challenge of finding a modality to train students and scientists "en masse" in bioinformatics topics, an endeavour that soon proved to be far more difficult than initially anticipated. Many trainees across Africa do not have access to powerful machines or stable internet/electricity, nor do they have access to expensive software or costly training materials. Additionally, many institutes attempting to establish bioinformatics programs do not have much experience in more advanced topics, and often lack experience with the organisational aspect of running these types of courses. Sourcing local trainers with the required expertise to teach specialized topics also proved difficult and costly, particularly where bioinformatics uptake within the region has already been low. This often results in data science courses, particularly within more specialized fields like bioinformatics, becoming very costly and highly competitive, reducing access to bioinformatics training significantly. It quickly became clear that any traditional approach to training would not suffice, and that a unique model would need to be implemented to meet this demand for training in Africa.

Online learning coupled with distance learning seemed the most cost-effective approach, but tackling other challenges like receiving adequate technical and teaching support at remote regions across Africa (at minimal cost) were more complex. It is well known that online courses that fail to provide additional teaching support, and even at times those that do, experience large drop-out and failure rates (Onah et al., 2014; Bawa, 2016). To combat this, the establishment of a local hosting classroom with some minimum infrastructural requirements such as a training room with computers that could connect to the internet and a projector to play lectures, would allow trainees the opportunity to learn (mostly) uninterrupted. An added human capital requirement of 1) someone to coordinate the classroom, 2) onsite support in the form of teaching assistant/s (TA) familiar with the content and, 3) a systems administrator (SA) for technical assistance ensures trainees receive the physical support required throughout the course. Any classroom across Africa that could meet these criteria would be eligible to join the course and become a local hosting site/classroom, once thoroughly vetted.

Since online and distance learning were to be used, course organisers could partner with international subject experts situated across Africa and abroad, to develop and teach course content. The first course to be developed using this mixed-model approach was H3ABioNet's flagship Introduction to Bioinformatics Training Course (IBT), piloted in 2016 (Gurwitz et al., 2017). Since the training was introductory, trainers could develop content around predominantly online tools ensuring infrastructural requirements remained low and participants could repeat at least some analyses once the course concluded. All content was designed to ensure participants on the course developed appropriate competencies (as described by the International Society for Computational Biology's (ISCB) core bioinformatics competencies) (Welch et al., 2014).

Participation in the course for all participants and staff is completely free of charge and voluntary to ensure the course reaches the widest possible audience. This has allowed not only thousands of participants to be trained over the last 5 years in introductory bioinformatics skills, but has also provided staff with the support required to implement this training locally. The detailed course model is described by Gurwitz et al. (2017).

Training an increasing number of scientists in introductory bioinformatics skills meant the need and demand for more specialized/advanced courses, implemented using a similar model, was also growing. The first course established using a similar mixed-model approach was the African Genomic Medicine Training Initiative (AGMT), targeted at nurses across Africa and first run in 2017 (Nembaware and Mulder, 2019). The program uses a similar mixed-model approach, but with an added requirement of a collaborative research project toward the end of the program. The AGMT course has been accredited and forms part of ongoing professional development for nurses. More importantly, AGMT allows for a wider uptake of genomic medicine within primary health care as nurses are trained in a range of relevant topics.

H3ABioNet then piloted an intermediate level microbiome 16 S rRNA data analysis course (IntBT) in 2019 which saw local

classrooms across Africa now needing stronger local technical expertise to access and maintain advanced software to perform more complicated and data intensive analyses, typically on a computing cluster environment using containers (Ras et al., 2021). The course allowed local staff the opportunity to pull and maintain containers while being supported by an experienced core team. These containers were then accessed by their local participants to perform in depth analyses of microbiome data. The use of containers had the added benefit of all software and tools remaining available for further analyses once the course ended and allowed local staff the opportunity to receive some systems administration training and support. Ongoing support was also made available via the H3ABioNet Helpdesk to ensure staff and participants could continue to receive support with local project analyses once the course ended (Kumuthini et al., 2019).

Feedback is collected at various time points (before, during and after) throughout the training to continually monitor the trainee's learning experience as well as the experiences of the staff to allow organisers to continually improve the model from year to year. The pilot phases of both the AGMT and the 16 S rRNA data analysis courses were successful (see (Nembaware and Mulder, 2019) and (Ras et al., 2021)) and demonstrates the flexibility and potential of the model for teaching more advanced and specialised topics. The success of the model with more advanced topics led to a partnership with Wellcome Connecting Science in the United Kingdom (UK) and the development of a Next Generation Sequencing Course which was run across 31 classrooms in Africa in 2021 with ~400 participants registered across them; an exciting "next step" in making high quality bioinformatics training more accessible to African scientists.

When the coronavirus pandemic struck during 2020, the flexibility of the mixed-model approach was tested even further as a result of lockdowns across most African countries. The pandemic meant nearly all face-to-face classrooms could not run, leaving the organisers with the difficult task of ensuring effective support was still provided from local staff teams, while also ensuring that the participants still gained the classroom experience without access to an actual classroom and the ability to meet physically. These courses also typically ran across many months with biweekly contact sessions which ended in a live, online Question & Answer session (Q&A) with the module trainer. While previously trainees were situated within a physical classroom, meaning only one online connection was required per classroom, each participant now needed to connect individually. For IBT, this meant over 1000 simultaneous connections, which quickly became a limiting factor as most online conferencing platforms at the time could not handle the large groups with a standard license, and the large number of connections quickly meant a bandwidth too large for trainees to maintain. While materials are made available via a Learning Management Site, Vula (Learning Management System used at the University of Cape Town which is based on the Sakai (<https://www.sakailms.org/>) platform), where participants may access the content at any time, the live Q&A was frequently not a component that could be shifted/modified. In some instances, trainees also struggled with internet accessibility/stability, which

often disrupted their access to the LMS, Vula and online sessions. Similarly, while Vula allowed structured forums to be made available to participants and staff, this was not enough to effectively support trainees who were now typically very isolated. The organisers thus worked alongside classrooms and found many creative solutions such as: the creation of breakout rooms on Zoom for more of a classroom feel which were accessible during the stipulated contact session times (in the case of IntBT); staggering sign on times to the online platforms with rolling 30 min slots for Q&As with the trainer (in the case of IBT); having some local classrooms create Whatsapp groups/independent Zoom rooms (used in both IBT & IntBT) and; even at times, using social media to stay in touch. In regions where trainees really struggled with internet accessibility, some classrooms opted to provide materials on removable storage drives or set up contact points where a head TA would meet with participants to transfer materials. H3ABioNet has also made use of Virtual Machines to ensure all trainees have access to a standardised system for more advanced courses. Both the IBT and IntBT courses still ran successfully in 2020 despite the various challenges, illustrating the effectiveness of the mixed-model approach to teaching.

The model has been used successfully for many years and while it holds a great deal of potential for training across many different topics/domains, it does not come without its share of challenges. It is beyond the scope of this paper to provide a detailed account of these challenges, however, a few are briefly discussed here (in addition to the COVID-19 related challenges discussed previously). As mentioned, a major benefit within the IBT course has been the use of predominantly online tools—this has allowed anyone with access to a browser and internet the ability to join the course. A major disadvantage of using mostly online tools, however, is that course materials need to be constantly updated as online resources tend to change and be updated fairly regularly. Course convenors are thus required to regularly engage with course trainers throughout the year to ensure materials are frequently updated and remain current for every iteration of the course. This is in contrast to the 16 S rRNA course (IntBT), where software and tools used within the course are packaged within containers and run locally. Since this course makes use of few online tools, materials mainly require updating when there have been significant updates or changes to the containers, software or workflows used within the course. It also requires course convenors to work closely with trainers as well as software developers, and to remain abreast of major changes to software and tools used within the field. Another key challenge within the current model and one that is faced by numerous training providers globally, is that all course content has historically been developed and provided in English. While this has resulted in a great uptake of materials within English-speaking African countries, it has meant little to no uptake within Francophile/French speaking countries. Accessibility to the course materials for those with other impairments i.e. hearing, visual are also major obstacles that need to be considered if the model is to reach as many people as possible (material accessibility is discussed in more detail later). Finally, the model constitutes a large amount of logistical and

administrative overhead, mostly undertaken by (typically) a single course convenor who manages and coordinates the various components of the model. Convenors need to have advanced coordination skills and need to have the ability to deal with, and address a wide range of issues while balancing many competing priorities—often managing cohorts and staff teams in the region of 500 to greater than 1000 people at any given time.

Applied Continuous Skills Development for Scientists, Engineers and Systems Administrators Through Hackathons

Training workshops, in their various formats, provide a controlled environment for the learning and use of different types of bioinformatics resources and tools which is beneficial to audiences being newly introduced to bioinformatics. Part of bioinformatics capacity development involves continuous learning and the practical application of new methods, tools, technologies and leveraging multi-disciplinary skills in a collaborative environment to solve various problems and produce outputs. Science hackathons provide a collaborative environment for scientists to work together over a short period of time fully concentrating on solving problems that enables concomitant skills transfer and learning (Groen and Calderhead, 2015; Aboab et al., 2016; Lyndon et al., 2018). To ensure new skills acquisition and continuous development through multi-disciplinary knowledge sharing via peer learning within the network, H3ABioNet has organized and held a variety of hackathons for audiences with diverse skill sets ranging from bioinformaticians, computer scientists, systems administrators, and statisticians to biologists (Ahmed et al., 2018; Ghouila et al., 2018; Fadlilmola et al., 2021). Each of the H3ABioNet hackathons have their specific target audiences that encourage multi-disciplinarity and are designed to achieve predetermined goals related to specific ongoing H3ABioNet projects that continue post-hackathon.

To address the issues of reproducibility and portability of executing bioinformatics software stacks in heterogeneous computing environments while keeping up with advances in technology, H3ABioNet organized a workflows and cloud computing hackathon (Ahmed et al., 2018; Baichoo et al., 2018). Introductory training on workflow languages (Common Workflow Language and Nextflow) and containerization of bioinformatics software applications for portability were provided to H3ABioNet personnel, including bioinformaticians, software developers and systems administrators (Baichoo et al., 2018; Ahmed et al., 2019). The goal of the H3ABioNet workflows and cloud computing hackathon was to create four containerized workflows that are portable for Genome Wide Associations Studies, imputation, variant calling from NGS data, and the processing and analysis of 16 S rRNA microbiome data. Specific workflow development teams consisted of bioinformaticians and software developers to develop the workflows, and a systems administrator for software containerization (Ahmed et al., 2018). Some participants of the hackathon have gone on to participate in

other Nextflow community organized hackathons outside of Africa, and have provided training on these topics in subsequent workshops. Skills gained from this hackathon are currently being applied to various H3ABioNet projects and have led to the development of workflows and containers used to provide training for advanced topics such as the 16 S rRNA data analysis course (IntBT) (Ras et al., 2021). The GWAS workflow and imputation service was used to provide training and analysis of data to H3Africa consortium members in a bring your own data (BYOD) face-to-face workshop.

Another goal-oriented hackathon, the DREAM of Malaria hackathon hosted by H3ABioNet and other organizations, brought together scientists at various career stages encompassing data generators, bioinformaticians, statisticians and data modellers to investigate various methods for predicting dihydroartemisinin (DHA) sensitivity of *P. falciparum* isolates using genome wide expression profiles (Ghouila et al., 2018). The multi-disciplinary composition of the various hackathon teams enabled more in-depth discussions and knowledge transfer via peer learning between the hackathon participants within and outside of their teams, while testing various methods in preparation for the Malaria DREAM Challenge (Davis et al., 2019). Data modellers gained a much better understanding and appreciation of why biological data can inherently be noisy, while data generators gained more knowledge of statistical modelling and analysis methods for their data (Ghouila et al., 2018). Apart from peer learning and knowledge transfer, hackathons are a great way of improving communication, fostering long term cohesion and collaboration between scientists and engineers while working towards common goals, as in the case of the H3ABioNet hackathon for the development of a genomic medicine and microbiome portal for African genomics data (Radouani et al., 2020; Fadlemola et al., 2021; Hamdi et al., 2021).

With the onset of the COVID-19 pandemic, H3ABioNet switched from a planned face-to-face hackathon to a virtual hackathon format utilizing Zoom and breakout rooms for a project aimed at creating an African genome reference graph. The virtual hackathon organization and format was quite different from the traditional face-to-face hackathons held by H3ABioNet. An advantage of having a virtual hackathon format that ran over 2 weeks for 4 to 5 hours a day was that more people within H3ABioNet could participate. The average cost per participant for a face-to-face hackathon (~USD 1,500) imposes a limit on the number of individuals to between 25–30 that can participate. Depending on the nature of the goals for a specific H3ABioNet hackathon, a predetermined selection criteria that places emphasis on existing pre-requisite skills is applied for a number of reasons. Some applicants mistakenly interpret a hackathon to primarily be a training workshop, rather than a goal-oriented event. Participation requires a high level of engagement and contribution that draws upon pre-existing knowledge at the scientist, engineer or systems administrator level, while rapidly learning and applying specialized skills. It can also be disheartening for a participant to attend a specific hackathon for which they are not equipped to make any meaningful contributions to at that particular point in time.

The prerequisite skills required to participate in some of the sub-projects for the virtual African genome graph hackathon included familiarity with the use of the version control system Github, graph building tools, Nextflow workflow language, NGS sequence formats and alignment tools and using High Performance Computing (HPC) environments. A number of applicants had some very basic skills or understanding of using Github, Nextflow and little knowledge about how to containerize software. As there was no financial constraint on participant numbers for the hackathon in its virtual format and to be as inclusive as possible, virtual training open to all network members was provided on using version control and Github. Introductory training on Nextflow and software containerization was also provided by participants who attended the previous H3ABioNet workflows and cloud computing hackathon. A specific sub-project within the African genome reference graph hackathon to containerize software used by the other sub-projects was aimed at systems administrators within H3ABioNet. A surprising outcome was the number of applicants who were not systems administrators that signed up specifically to be part of the containerization sub-project, rather than the other sub-projects using various bioinformatics tools. Further enquiries revealed that some of the applicants were bioinformatics students and scientists within H3ABioNet wanting to make their software and tools available in the form of containers. The virtual format of the hackathon provided an interactive environment for bioinformatics scientists to learn how to containerize software from systems administrators, and for interested users to work with experienced bioinformatics engineers to run novel tools in different HPCs while applying the skills learned during the pre-hackathon workshops. Overall, H3ABioNet has made effective use of the hackathon format to provide paths for continuous applied skills development and peer learning for African bioinformatics scientists, engineers and systems administrators as well as advanced users in multi-disciplinary, collaborative environments within Africa.

Internships as a Means to Bridge the Skills Divide

While training workshops and hackathons can impart and strengthen an individual's knowledge and skills base, they might not necessarily be a suitable bridge for a bioinformatics user to gain scientist level expertise in a specialized bioinformatics topic. To address this, H3ABioNet has implemented an internship program since its inception. Internship programs are a form of training that are essentially supervised work experiences over a dedicated period of time and have been found to be impactful, and in some cases essential for budding African scientists at postgraduate level (Mlotshwa et al., 2017). Recognizing that specialized bioinformatics skills are needed to conduct in-depth analysis of genomics data from a myriad of studies, H3ABioNet's internship program was developed specifically for the H3Africa and H3ABioNet consortium, where students and staff members from an H3Africa data generating project can apply to spend time at an H3ABioNet Node and learn to work with, analyse and interpret their data.

H3ABioNet staff and students can apply to spend dedicated time at another H3ABioNet Node or an external institution with established expertise for a specific computational analysis of data that they seek to better understand.

An important requirement for the H3ABioNet internship is that the individual applying for an internship should have a dataset which they will analyse for the duration of the internship, as this will enable them to gain practical in-depth expertise with this type of analysis. Additionally, the H3ABioNet internship program does not cover the costs of data generation, as H3ABioNet is a funded resources project with clear deliverables for capacity development and not research data generation. A proposed analysis plan, expected outcomes and motivation letters from the applicant's current supervisor and hosting institution and supervisor, as well as a plan to disseminate the skills they have learned at their home institution to ensure capacity development, are mandatory with the application. Additional information such as courses attended, presentations and publications are also required to determine whether an applicant has the foundational skills and knowledge to fruitfully undertake the proposed internship. While the internship program was directly affected by the COVID-19 travel restrictions, it proved a highly valuable and successful component of the H3ABioNet training endeavours and has resumed where possible. Further details on the internship program have been highlighted in the training impact section.

Developing Additional Career-Building Skills

Most early-career research scientists in Africa face continent specific and debilitating challenges which include poor access to libraries and online resources such as journals, lack of funds, inadequate support and lack of good research infrastructure and tools (Nchinda, 2002; Mccullough, 2010). It is from this perspective that the H3ABioNet training and educational initiatives are not only based on the technical or scientific knowledge transfer, but also focus on multiple, holistic, interdisciplinary, practical and diverse soft skills that scientists need for their scientific studies and careers, especially in a challenged continent like Africa. H3ABioNet focuses on experiential learning (learning by doing, which is hands-on in scope) that allows participants to benefit significantly from our offerings. Similar to Margaret Hostetter (Hostetter, 2002), the ideology that career development will enable our diverse, previously and currently disadvantaged scientists and especially bioinformaticians to become highly motivated and intensively trained scientists ready for their different careers across Africa is strongly espoused. We regularly offer face-to-face and online workshops (especially since the onset of COVID-19, all of our training has been virtual), webinars, conferences, career development, internships, mentorship and colloquiums all geared towards developing the soft skills of members and affiliates, be they professionals or students. Most of the soft skills training is offered in collaboration with a diverse group of scientists and trainers from across Africa working in different

organizations such as H3Africa and the African Academy of Sciences (AAS).

H3ABioNet provides short term and long term soft skills training depending on the context and temporality, for instance the H3ABioNet UCT CBIO Node contributes to an annual, all year postgraduate development program geared towards assisting students to learn how to be better academic writers and presenters. The shorter format training includes face-to-face workshops that are provided during H3Africa consortium meetings to H3Africa fellows, before the onset of the "new normal" imposed by COVID-19 restrictions. To quote from King (King, 2013), it is necessary to develop creative and sustainable ways to help "early career research scientists ascend the professional ladder." Regardless of the limitations inherent in online learning and training such as the lack of human contact, Brancaccio-Taras et al. have proven that effective scientific training can be offered to fellows with impactful results even on virtual platforms such as webinars (Brancaccio-Taras et al., 2016).

A recently run soft skills H3ABioNet workshop was the Scientific Communication Workshop for H3Africa fellows, staff and scientists, offered at the 17th Consortium Meeting held in April 2021. The main purpose of the workshop was to train attendees on different scientific communication platforms, styles and methods for specific needs. The workshop was facilitated by diverse volunteer trainers from across H3ABioNet nodes based in different African countries and the US. Similar to most of the training H3ABioNet provides, trainers conducted the training free of charge in order to offer necessary workshops to participants at no cost. The Scientific Communication Workshop had multiple learning outcomes aimed at addressing the often overlooked skills required in progressing in an academic environment. These included learning the importance of effective communication and writing, good practices for designing eye-catching posters, developing elevator pitches for sharing of research, valuable tips for providing constructive feedback, advice on the value of timely thesis preparation and referencing software, and guidance on how to effectively critique and summarise a research article.

Building Bioinformatics Communities Through Online Training and Support

The H3ABioNet training network spans many African countries (currently more than 16 countries). It has, and, continues to grow in large part due to our mixed-model training approach and blended learning courses. The Learning Management System used within the mixed-model courses, Vula, allows structured forums and a real time chatting tool to be made available to trainees which quickly enables participants from all over the world to connect in real time. Participants from across Africa and across all classrooms use these forums to connect and interact while joining one of our training courses. Since our training is often open to anyone with the prerequisite skills to apply, these courses double as a mechanism to connect local bioinformatics students and/or enthusiasts, to staff and peers at their local classrooms, institutions or regions. This, coupled with the

skills gained as part of various training events allows bioinformatics communities to begin forming, often at remote regions where bioinformatics uptake may not be as high. Many trainees return to act as staff i.e. TAs or SAs, or may even start up additional classrooms in future iterations of courses—further developing local capacity for bioinformatics teaching and training. Trainees are also encouraged to remain in touch with their local classmates and staff for ongoing training or further opportunities to participate in webinars, events and conferences. More recently we have attempted to implement “Open Learning Circles” (<https://h3abionet.github.io/LearningCircles/>), a coordinated “Mozilla open classrooms style” space where trainees that have attended any of our mixed-model courses could continue to come together, share new content, work on ongoing data challenges or simply continue to connect. H3ABioNet has also created Slack workspaces where participants from various training programmes could continue to connect, share resources and work on coding problems together, well beyond a course concluding. The creation of mailing lists and various social media accounts also ensures trainees are kept up to date on opportunities and allows for ongoing communication between H3ABioNet and these trainees.

One of the skills which is essential for future academics, but seldom included in traditional educational programs is training skills for trainers. There are several train-the-trainer programs developed for different contexts, but few specific to bioinformatics. The European Bioinformatics Institute (EBI) has run bioinformatics train-the-trainer courses for specific bioinformatics applications, and Wellcome Connecting Science has recently developed a train-the-trainer Massive Open Online Course. H3ABioNet has been developing trainers using different approaches. Several H3ABioNet members attended the EBI NGS train-the-trainer course and went on to assist at a Wellcome Connecting Science NGS course and are now trainers on the previously mentioned mixed-model NGS course. A group of African Carpentries (<https://carpentries.org/>) trainers is currently being established through working with the Carpentries community to host a Carpentries Instructor training course for H3ABioNet members. H3ABioNet has also developed a training guide that describes the processes followed in planning, designing and running courses, including relevant templates for the different steps (accessible at <https://doi.org/10.25375/uct.14337806.v1>). This is a valuable resource that can be used by others who wish to set up and run any type of training course, with a focus on bioinformatics resources. More informally, students who have participated in H3ABioNet courses are often given an opportunity to be teaching assistants on the next iteration of the course. Working alongside experienced trainers, they gain practical experience in providing training. The establishment of a strong training network has allowed a budding bioinformatics training community to begin forming in Africa which is now gaining traction across other continents too.

H3ABioNet members have also contributed to global efforts to develop training resources for bioinformatics trainers, such as competency frameworks, guidelines for course and curriculum development and a trainer portal, hosted by GOBLET (Global

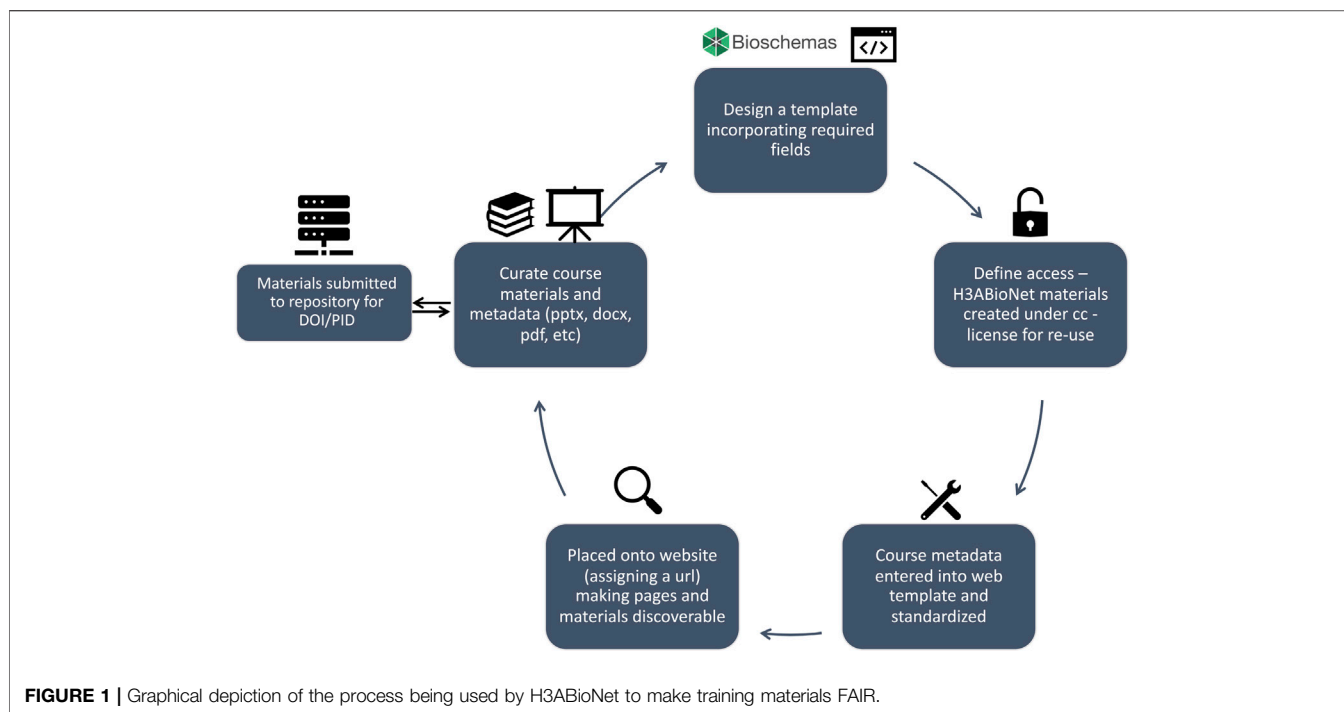
Organization for Bioinformatics Learning, Education and Training: <https://www.mygoblet.org/training-portal/trainer-resources/>). H3ABioNet initiated and hosted the first Bioinformatics Education Summit in Cape Town in 2019, which brought together bioinformatics trainers and educators from around the world to develop trainer resources. This community of trainers meets monthly and has since organized 2 further summits. Through these and other activities, such as the ISCB Education Committee and COSI (community of special interest), H3ABioNet has developed long term partnerships with key global bioinformatics trainers and training organizations, with whom we have co-organized several training events and are currently working with the global bioinformatics trainer community to develop an online train-the-trainer course which will be run using our mixed-model training approach.

Addressing Gender Inclusion in Data Science

According to Stads and Beintema (Stads and Beintema, 2006) “the science world appears to be greatly affected with gender barriers that disadvantage female scientists in their career development.” To address this gender disparity and inequality, H3ABioNet intentionally organizes spaces and training that targets African women in science. The focus on training African women is based on research and statistics that show that there are many barriers that African female scientists face, such as lack of funding, the work and family balance as well as patriarchal gatekeeping (Sonnert, 1999; Prozesky and Mouton, 2019). A few H3ABioNet female scientists are WiDS (Women in Data Science) ambassadors, and in 2021 we organized the inaugural WiDS Africa event which was hosted virtually and intended to engage women in data science across Africa. This was also an important collaborative space where women were encouraged by other women on ways to navigate the very exclusionary science fraternity. At the conference there was an echoing of the need for more women in science communities which has sparked future collaborations necessary for this endeavour within H3ABioNet (Chauke, 2021).

Curation of Training Materials and Improving Accessibility

Running several training events generates a wide range of training materials, from lecture slides to codebooks. H3ABioNet has thus embarked on a rather large effort to make our training materials (and supporting tools such as containers and workflows) more accessible, starting with the implementation of FAIR principles (Garcia et al., 2020) across our training materials and training pages available on our website. In order to make our training materials FAIR i.e. Findable, Accessible, Interoperable and Re-usable, H3ABioNet is ensuring our webpages reflect the most relevant information to ensure optimum findability and accessibility of our materials. The use of tags and keywords across H3ABioNet webpages to improve Search Engine Optimisation (SEO) along with submitting materials to a public repository to assign a permanent identifier, improve



data provenance and discoverability, and are two ways in which we are improving FAIR compliance across our materials. We have also recently begun the process of implementing bioschemas (<https://bioschemas.org/>) across our training webpages (**Figure 1**). Bioschemas allow for a standard schema markup to be applied across our pages and materials, making them more discoverable to web scrapers and training databases and repositories like GOBLET (<https://www.mygoblet.org/>) and TeSS (<https://tess.elixir-europe.org/>), thus increasing their findability. These tasks are currently underway with the aim to make all H3ABioNet training materials FAIR and freely accessible before the end of 2021 (many materials can already be accessed, freely and openly, via our website and course pages). H3ABioNet is also improving accessibility and reusability of materials by transcribing lectures and videos to add subtitles. The aim is to transcribe all materials into English in the first instance after which transcriptions can then be more easily translated to other common languages spoken in Africa like Arabic and French. Since most materials within H3ABioNet are generally released under a creative commons license, our materials are also free for re-use and distribution by anyone who needs them.

RESULTS AND DISCUSSION ON THE REACH AND IMPACT OF THE H3ABIONET TRAINING ENVIRONMENT

The different approaches discussed above were developed to address specific training needs and to adapt to local and global challenges. The H3ABioNet training environment, summarized in **Figure 2**, includes not only the training interventions for

imparting bioinformatics skills, but also career development opportunities, networking and community building, and access to training materials and trainer guidelines and resources. This is to ensure our training has a longer term impact with continued support. For all our training interventions we usually conduct pre- and post-training surveys and evaluations to determine what the participants learned, what they want to learn in future, and their impression of the event. The results of each training event's evaluations and feedback are discussed in post-course meetings and results are sent to trainers or guest lecturers to assess training successes, materials or implementations that may need to be improved. Monitoring, Evaluation and Learning (MEL) are imperative to all our training endeavours at H3ABioNet as we believe that it is necessary to constantly take stock and learn while iteratively providing alternative and additional training solutions. In the section below, we present selected results sourced from some of our experiences, in addition to feedback from a biannual long term evaluation survey sent out to each participant of an H3ABioNet training event. We discuss some of the results in terms of the impact of the various training modalities H3ABioNet has employed on the acquisition of knowledge and skills and the progression of participants' careers.

Training Audiences

Using a multi-pronged approach, H3ABioNet has developed a comprehensive bioinformatics training program for a diverse audience. To date (May 2021), 4,466 unique trainees have attended one or more H3ABioNet training course. Courses are tailored to the audience by designing the curriculum based on the competencies required by participants, taking into account their background and prerequisite skills. A summary of attendee backgrounds is presented in **Figure 3**. Since not all our

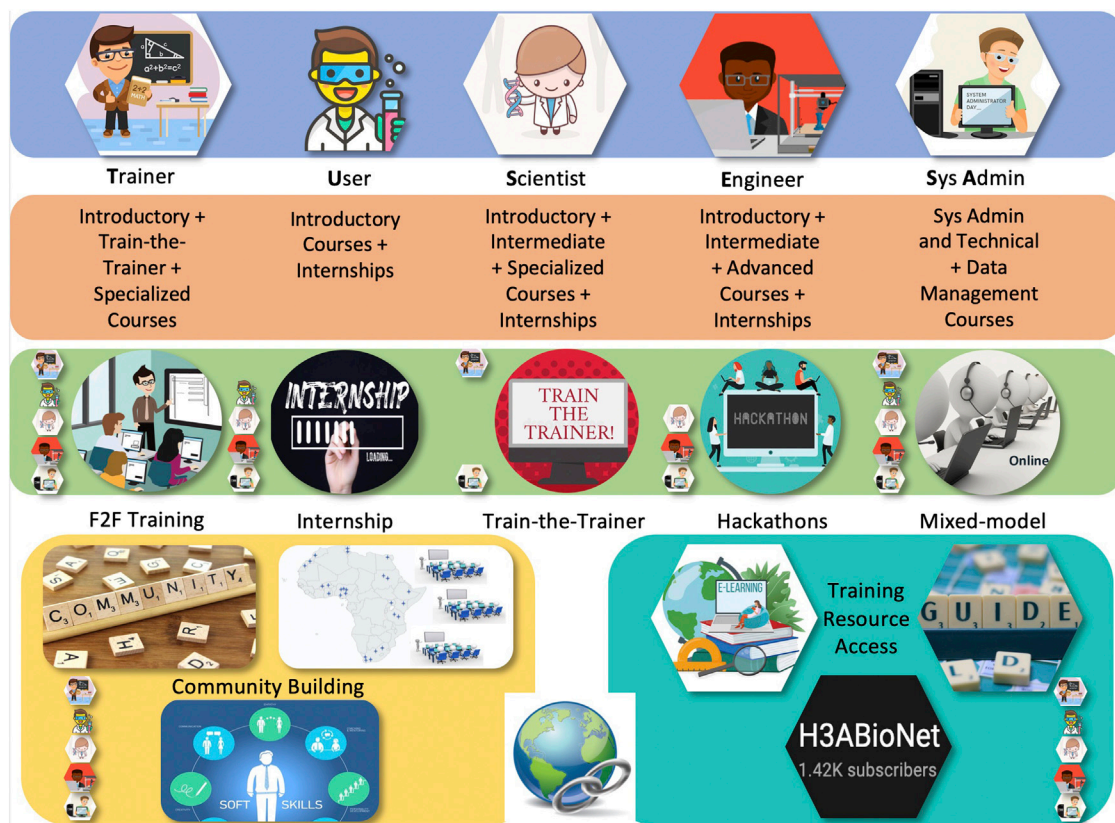


FIGURE 2 | A diagrammatic representation of the H3ABioNet training environment. The blue bar (first tier) represents the different audiences (personas) identified as necessary to develop within the network. The image of each persona is used to represent the choice of training approach/resource in the figure. Courses were designed and presented at varying levels of difficulty (red bar—second tier) to allow for progressive growth of individuals based on their desired backgrounds and planned career paths. Similarly, the complementary modes of teaching (green bar—third tier) presented the opportunity for individuals to explore and learn in unique environments that fostered the support and opportunities required to develop specific competencies and personas. The fourth tier (yellow block) of the training environment provides individuals with the opportunity to learn as part of a larger community as opposed to individual silos and further develop the complementary skills required to excel in building a successful academic career. The final component (teal block) highlights our efforts to make our training resources accessible beyond H3ABioNet to ensure that the efforts to continue training in bioinformatics are sustainable and continue to be utilised by the broader bioinformatics community.

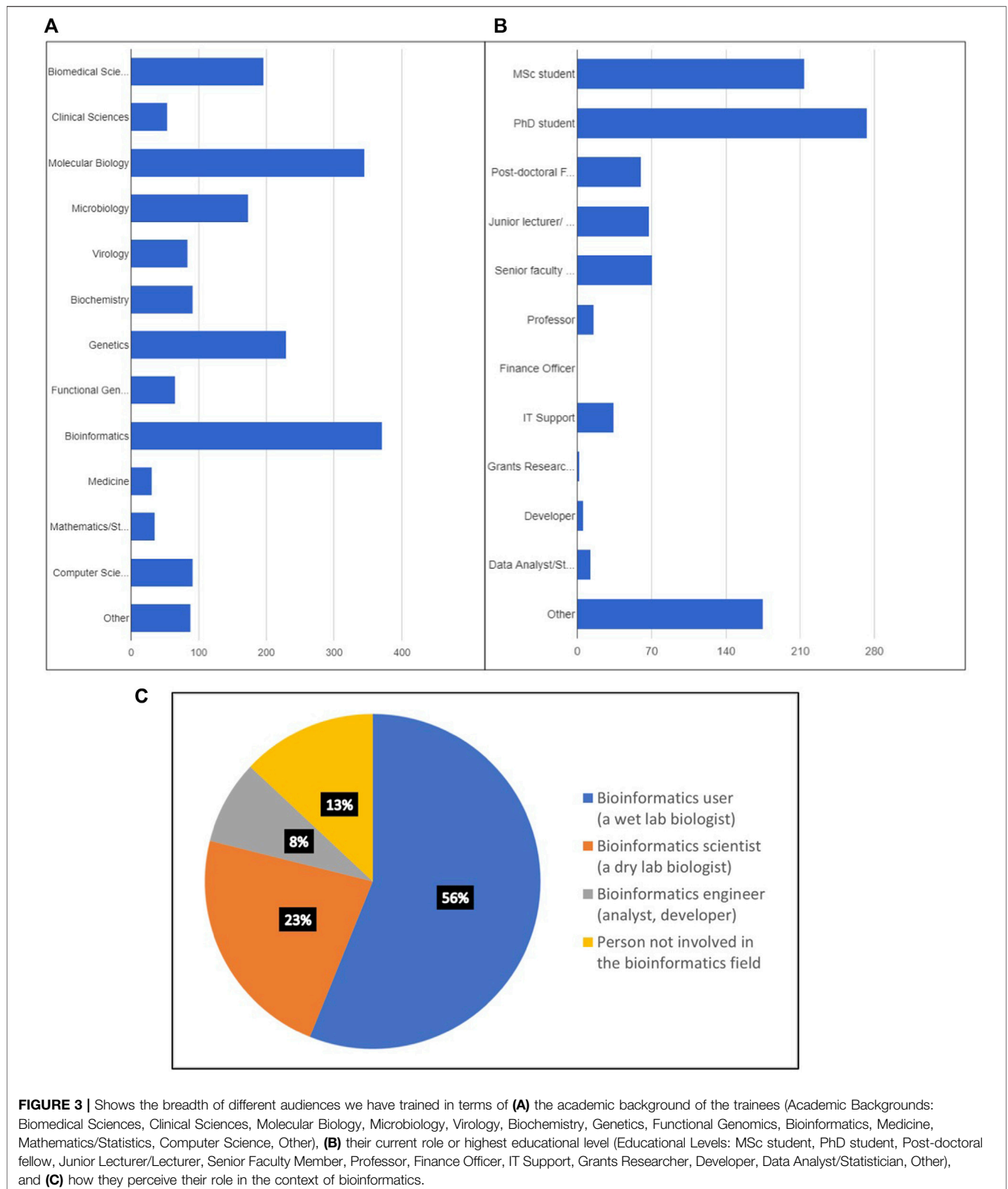
application forms captured information on the background of participants, this data was derived from those participants who completed the long term evaluation survey, but we believe this is generally representative of the broader participant pool. **Figure 3** shows that the majority of trainees are from a life sciences background, predominantly at postgraduate level, and consider themselves to be bioinformatics users. This reflects the large IBT audiences, as these courses are primarily aimed at life scientists who need to use some bioinformatics tools for their research.

In terms of geographical distribution, the initial courses were limited to locations that fulfilled the requirements to host face-to-face workshops. The transition to a mixed-model training approach resulted in an increased reach with several classrooms being able to be hosted for a single course. The interest in the mixed-model approach training courses such as IBT and IntBT has grown each year, with new countries enrolled during each iteration of the course. While we have not yet reached each country on the continent, we have provided training to attendees in a substantial number of African countries as shown in **Figure 4**. Attendees from

outside the continent, particularly from the USA and UK, were generally attendees of one of the hackathons. The mixed model courses, in addition to increasing the geographical reach, have resulted in a large increase in the annual number of attendees at our training events (**Figure 5**). Even at the height of the pandemic (2020), we trained over 1200 people.

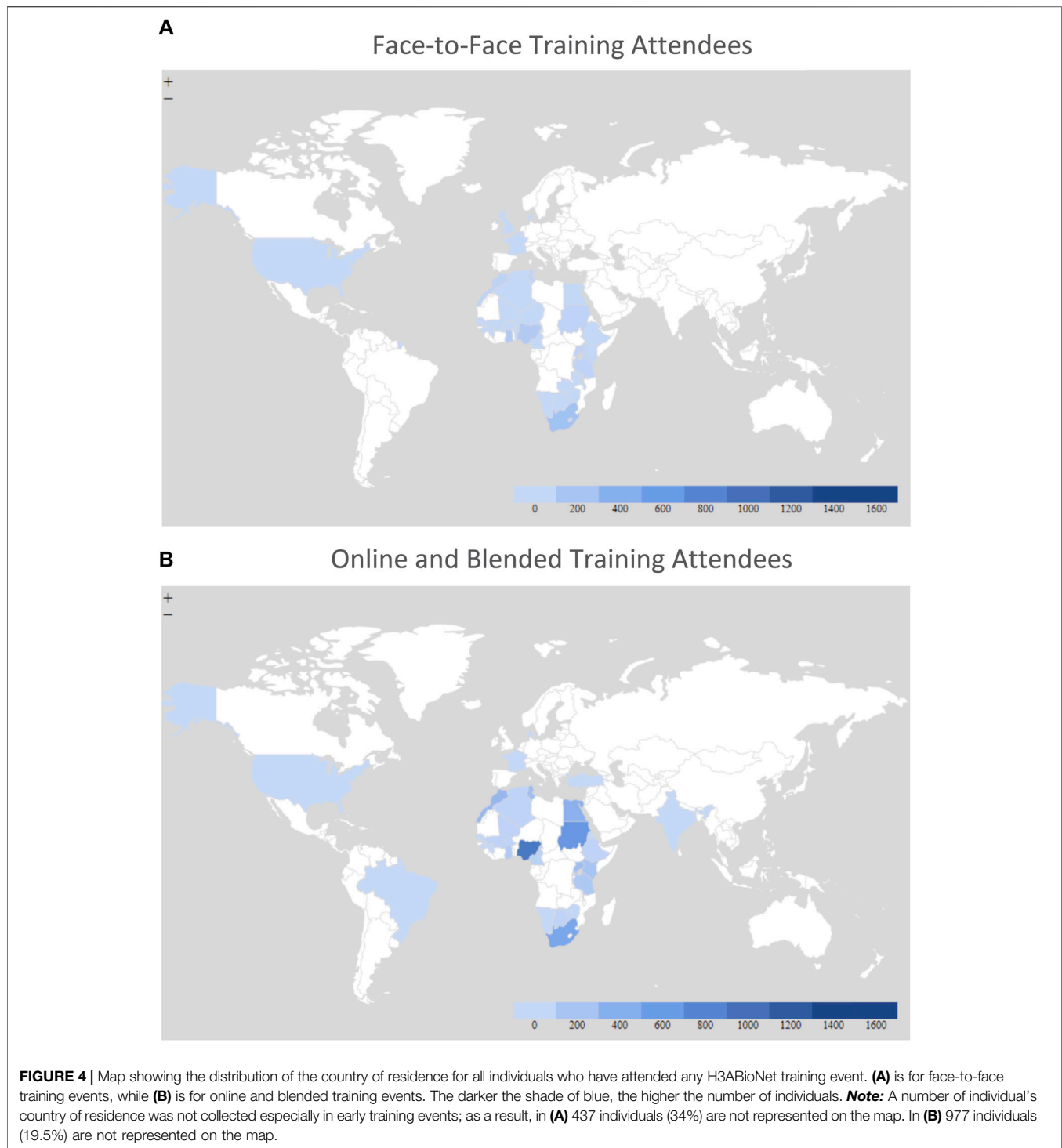
Training Topics and Course Attendance

The training events have covered a wide variety of topics at introductory, intermediate and advanced levels (**Supplementary Table S1**). The selection of training topics has primarily been driven by the perceived analysis needs within the H3Africa consortium as well as the need to develop highly skilled bioinformaticians within H3ABioNet. In addition to the annual Introduction to Bioinformatics course, which covers 6 different basic bioinformatics modules, the most common topics covered at varying levels of complexity are NGS and GWAS data analysis. While around 70% (3024) of the trainees attended one of our Introduction to Bioinformatics courses (these attract the largest audience per course because of the scalable training



model), several trainees have attended multiple courses (over 400 people have attended 2 courses and nearly 100 have attended 3 courses, see **Figure 6**) to gain general bioinformatics knowledge,

more specialized bioinformatics skills, and career-building skills such as grants management and scientific communication. Many people who attended an IBT course did not go on to do



further training, though around 200 trainees have also done one of the IntBT courses. It is worth noting that apart from the IBT courses, to cope with the demand, our events are generally limited to H3Africa or H3ABioNet members, so most IBT attendees were not eligible to attend the specialized courses. Nevertheless, the most common pairwise combination of courses attended by H3ABioNet/H3Africa

members was an IBT and the 16 S rRNA IntBT (**Figure 7A**), followed by an IBT and a career development course. When grouped by category (**Figure 7B**), the same most popular combinations were found, suggesting that more than 200 of our trainees have gained foundational bioinformatics skills, then specialized by increasing their expertise in a particular data analysis area, and supplemented these with scientific or

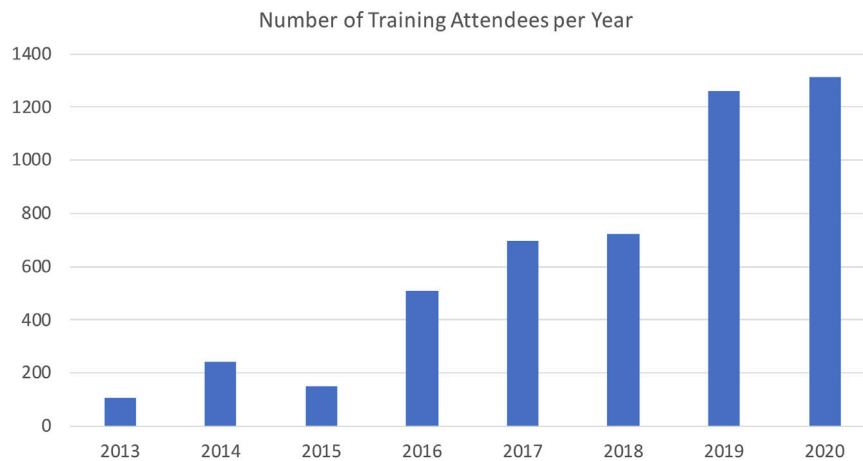


FIGURE 5 | The number of individuals attending H3ABioNet training events per year since the network's inception in 2013. The dramatic increase in 2016 was due to the inclusion of the mixed-model training approach.

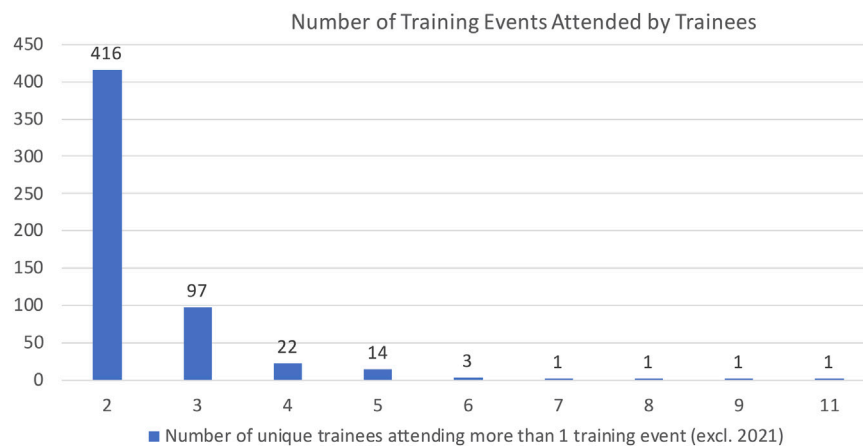


FIGURE 6 | Number of H3ABioNet training events attended by trainees. Most trainees attended just 1 training event (most often IBT, 3777 individuals), however, several went on to attend additional H3ABioNet training events.

grant writing skills. These combinations would serve to build researchers from novice bioinformatics users to more well-rounded academics able to analyse their own data. Many of those who attended face-to-face, intermediate or advanced courses did not also attend an IBT course, presumably because they already had the basic bioinformatics knowledge they required.

Training Modalities

As described earlier, training has been delivered using a variety of approaches and modalities. While there is no overall preference for a single mode of training we did observe that individuals from different backgrounds preferred different modes of training. Based on the experience of running the different modes of training, **Table 1** highlights these preferences as well as some of the advantages and challenges encountered when implementing the different training approaches.

Training Impact

Different training approaches are likely to have a different impact, which, in turn, will be affected by the trainee's background and skills coming into the training intervention as well as their expectations and their engagement during the training. For individuals, internships have enormous potential to provide long term impact on their research and careers, due to the hands-on, personal nature of an internship at an expert facility, and the opportunity to develop partnerships and future collaborations. H3ABioNet has awarded 20 internships to H3Africa and H3ABioNet consortium members that mainly comprise of postgraduate students and staff members who wished to develop expertise and analyse data in topics ranging from structural modelling, GWAS, human variant calling, large scale data transfers to metabolic modelling, pathogen informatics, human population genetics and microbiome studies. Of the 20 H3ABioNet internships

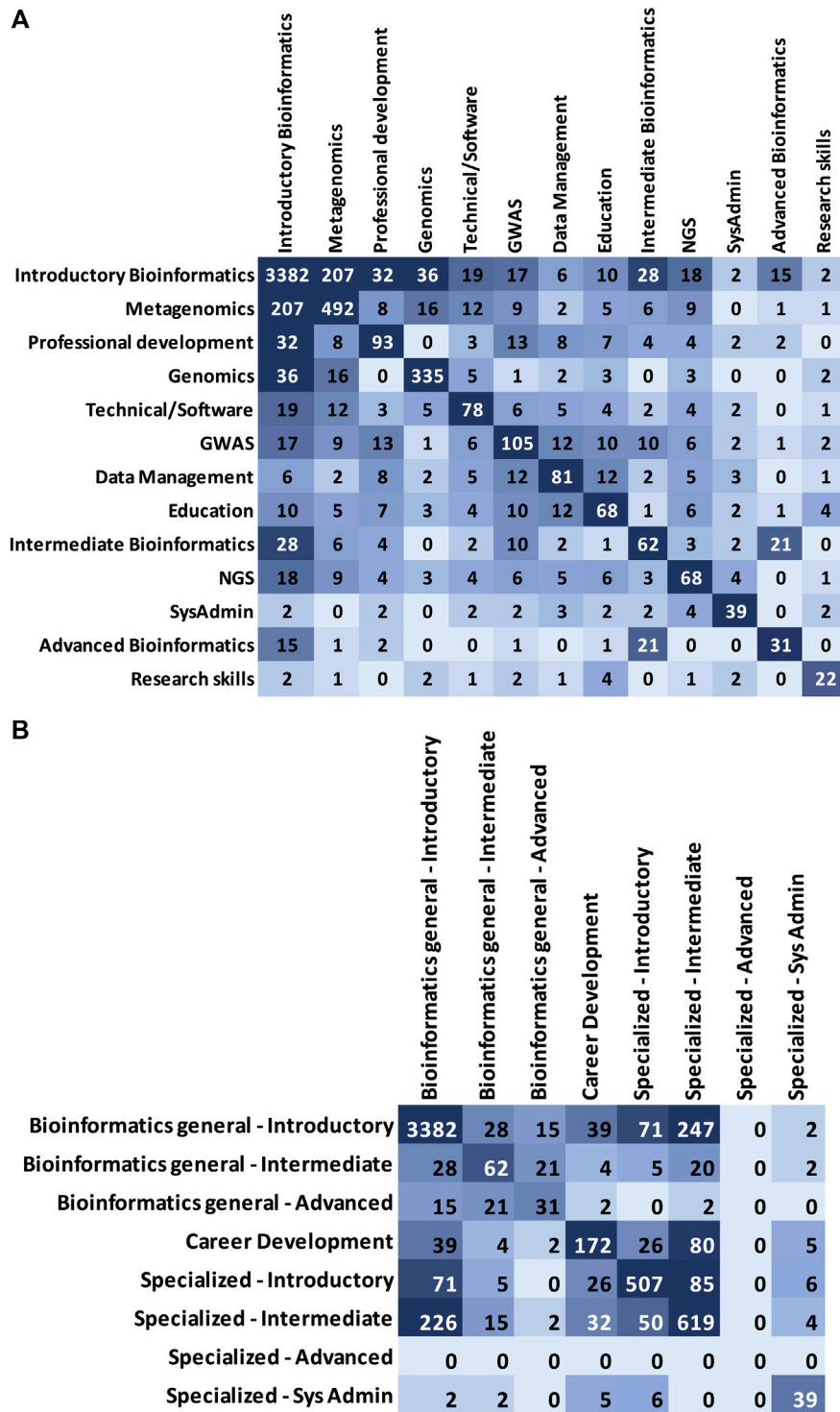


FIGURE 7 | Heatmaps showing the total number of trainees who attended combinations of training events (i.e. more than 1 training), the darker blues represent higher figures and therefore the most popular training combinations. **(A)** Shows which other training topics were commonly attended by the same trainee. **(B)** Also shows which other trainings were commonly attended by the same trainee but grouped by training category. Refer to **Supplementary Table S1** for further details on the training events under each training category and topic.

TABLE 1 | A description of the various training modalities preferred by the diverse audience types, highlighting the advantages and disadvantages associated with each modality.

Online Training <i>Preferred by: Developers and Junior Lecturers</i>	
Pros	Cons
Cost efficient Reach a large audience Easy to share material	Challenge to foster sense of community Challenge to form collaborations Licensing of materials could limit accessibility
Face-to-Face Workshops/Internships <i>Preferred by: MSc students, Doctoral and Post-doctoral students + IT Professionals + Professors + Senior Professionals</i>	
Dedicated time Close interaction Lots of support	Costly/unpredictable logistics Limited audience reached Trainees return to environments where they do not have access to software or tools
Hackathons <i>Preferred by: Developers, Junior Faculty, Data Analysts/Statisticians</i>	
Defined aims and output Develop practical skills Cross-disciplinary	Selection of participants crucial Base knowledge level required Limited audience
Mixed-Model Learning <i>Preferred by: MSc/PhD Students, Data Analysts/Statisticians and Professors</i>	
Cost efficient Reaches a wide audience Easily adaptable Develops local expertise	Huge administrative overhead on organisers Difficult to deal with local challenges remotely Regular updating of training material is challenging for courses with multiple modules and trainers, or content which changes frequently

awarded, 18 have been between H3ABioNet Nodes with 5 of the internships taking place at the H3ABioNet USA partner nodes, while 13 internships were within H3ABioNet African Nodes, and 2 of the internships have been to external institutions in Europe. The H3ABioNet internships have resulted in work being accomplished that has contributed to a number of publications by the interns, with many of the interns successfully completing their degree programs where applicable, or moving to new positions. In addition, selected internships have resulted in participants gaining strong applied skills in specific bioinformatics areas, creating the foundational knowledge required to contribute to several ongoing larger projects within the H3Africa consortium (Mulder et al., 2017a; Jongeneel et al., 2017; Azarian et al., 2018; Hamda et al., 2018; Ahmed et al., 2019; Choudhury et al., 2020; Sengupta et al., 2021). Due to the COVID-19 pandemic, the H3ABioNet internship program temporarily halted with internship applications having to be deferred to when international air travel was possible between different countries. The most recent H3ABioNet internship award was approved in April 2020, but due to travel restrictions the internship only commenced in March 2021.

The fact that internships are easily defined with agreed upon goals means they have an increasingly important role in education and bridging the gap between a bioinformatics user and a bioinformatics scientist. Internships are arguably one of the best ways for individuals to acquire practical skills and prepare for their careers as they provide many benefits such as learning how different scientific groups work, accessing

compute resources, setting up environments for their analysis, and working with dedicated bioinformatics scientists to produce outputs. Internships also provide intangible benefits that include being exposed to new environments, cultures and modes of working, while being immersed in a new environment, they enable networking and provide career-related experiences while facilitating skills transfer (Scott, 1992; Beard, 1998; Cook et al., 2004). These have benefited a number of H3ABioNet internship awardees (Scott, 1992; Beard, 1998; Cook et al., 2004).

Though face-to-face courses and hackathons are ideal for fostering interactions, these were not always possible when trying to reach a wide audience and in negotiating the global pandemic. Nevertheless, whatever the modality used, our training interventions had impact beyond just transferring skills. When trainees were surveyed, nearly 90% (1142 of 1289 responses) indicated that they had shared their new knowledge or the training materials with at least 3 other people, thus extending the reach of our training. While some passed on their knowledge through informal discussions or practical demonstrations, a few did this through hosting workshops. Additionally, approximately 33% (421 of 1289 responses) said the training led to, or facilitated a publication mostly by improving their knowledge and understanding of bioinformatics, through exposure to new ideas or topics, or due to improved data analysis skills. About 30% (413 of 1289 responses) indicated that the training facilitated submission of their thesis, for the same reasons as above. Interestingly, more than 45% (602 of

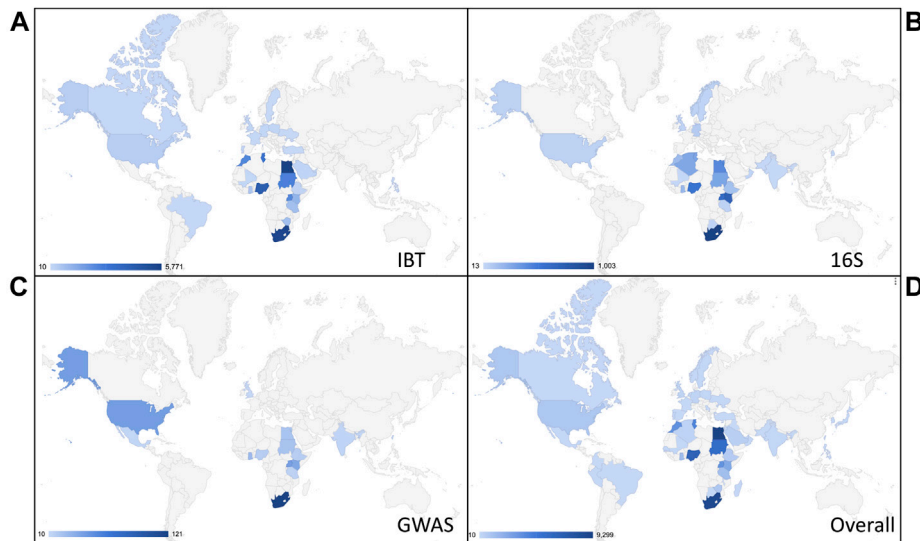


FIGURE 8 | Geographical locations (countries) where H3ABioNet YouTube playlists are being accessed from for the **(A)** Introduction to Bioinformatics Training (IBT) playlist, **(B)** 16 S rRNA data analysis Intermediate Bioinformatics Training Course (IntBT) playlist, **(C)** Genome Wide Association Study (GWAS) playlist and **(D)** H3ABioNet channel views overall. Maps do not depict views from countries with a total number of views <10 and represent the number of overall playlist views (clicks). Individual video views and analytics will thus differ. Colour scales on the maps represent the actual number of playlist clicks/views.

1289 responses) said the training led to a new collaboration, demonstrating the opportunity that training events provide for networking and building relationships with peers.

Accessibility of Training Materials

Although the H3ABioNet training materials have only recently been curated and some of the lectures and videos have been on YouTube for just a couple of years, all materials have been viewed to varying extents after the training event. For one of our GWAS courses, we ran a series of webinars providing theoretical content, followed by a week-long face-to-face hands-on workshop to ensure maximum use of the time together to focus on the practical components of the analysis. These webinar lectures have each been viewed over 1100 times from multiple countries both within and outside of Africa (**Figure 8C**). The IBT video series has between 100 and 3000 views per country on YouTube (**Figure 8A**) and draws a large number of viewers from outside of Africa, with our channel now having well over 100 000 individual video views overall. Despite videos only being available since 2018, they have drawn interest from the global community with a large number of views from the United States, Canada and India (**Figures 8A–D**). In addition, the training pages on our website (www.h3abionet.org/training) have been accessed several thousands of times. Our flagship IBT course pages alone have been accessed between 795 and 15,324 times between 2016 and 2020. Similarly our GWAS training pages have been accessed >1500 times. As we will be one of the first organizations to make our training pages Bioschemas compliant, all H3ABioNet training material should become even more easily findable and accessible. These results demonstrate the immediate impact and growth in uptake of materials by making materials more openly accessible and FAIR.

CONCLUSION

Here we have presented the H3ABioNet training environment which aims to provide a holistic training experience. We target bioinformatics users for introductory training and bioinformatics users, scientists and engineers for more specialized training. A parallel effort to develop local trainers and build grant writing and scientific communication skills ensures sustainability and improved prospects for career progression. While the COVID-19 pandemic has impacted our ability to take advantage of the personal interactions offered by face-to-face events, we were able to reasonably easily adapt to virtual training and continue intensive training throughout the lockdowns. In addition to transferring skills and knowledge, our training environment works towards making training materials FAIR and more accessible to those who are not native English speakers, and enables us to share our experience in training through our training guide and templates. Through the creation of a sustainable training environment, H3ABioNet has provided the foundation for further development of bioinformatics capacity across the continent beyond the lifetime of the network. Though funding for H3ABioNet will end soon, the systems and processes are well documented, and materials are available for the network of trainers that have been developed to pick up and continue at minimal cost. Already, the infrastructure is being leveraged in a separately funded project to roll out training for pathogen surveillance. The enormous demand for bioinformatics training in Africa will hopefully ensure that the legacy continues. While there are many successful training programs world-wide, H3ABioNet

has overcome many of the challenges of working in resource-limited settings and provided a multidisciplinary training program that has reached a very wide audience both directly and indirectly.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Centre for Higher Education Development and the Faculty of Health Sciences Human Research Ethics Committee at the University of Cape Town. The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

SA, PC, SP, VR, and NM contributed to the conception and design of the study and wrote the first draft of the manuscript. SA, PC, VR, SP, and NM contributed towards the development and running of the various training endeavours of H3ABioNet. KJ analysed the training evaluation results and generated the related figures for the manuscript. All authors contributed to

REFERENCES

- Aboab, J., Celi, L. A., Charlton, P., Feng, M., Ghassemi, M., Marshall, D. C., et al. (2016). A “Datathon” Model to Support Cross-Disciplinary Collaboration. *Sci. Transl. Med.* 8, 333ps8. doi:10.1126/scitranslmed.aad9072
- Ahmed, A. E., Heldenbrand, J., Asmann, Y., Fadlelmola, F. M., Katz, D. S., Kendig, K., et al. (2019). Managing Genomic Variant Calling Workflows With Swift/T. *Plos one.* 14, e0211608. doi:10.1371/journal.pone.0211608
- Ahmed, A. E., Mpangase, P. T., Panji, S., Baichoo, S., Souilmi, Y., Fadlelmola, F. M., et al. (2018). Organizing and Running Bioinformatics Hackathons Within Africa: The H3ABioNet Cloud Computing Experience. *AAS Open Res.* 1, 9. doi:10.12688/aasopenres.12847.1
- Azarian, T., Mitchell, P. K., Georgieva, M., Thompson, C. M., Ghouila, A., Pollard, A. J., et al. (2018). Global Emergence and Population Dynamics of Divergent Serotype 3 CC180 Pneumococci. *Plos Pathog.* 14, e1007438. doi:10.1371/journal.ppat.1007438
- Baichoo, S., Souilmi, Y., Panji, S., Botha, G., Meintjes, A., Hazelhurst, S., et al. (2018). Developing Reproducible Bioinformatics Analysis Workflows for Heterogeneous Computing Environments to Support African Genomics. *BMC Bioinformatics.* 19, 457–513. doi:10.1186/s12859-018-2446-1
- Bawa, P. (2016). Retention in Online Courses: Exploring Issues and Solutions—A Literature Review. *SAGE Open.* 6, 1–11. doi:10.1177/2158244015621777
- Beard, D. F. (1998). The Status of Internships/Cooperative Education Experiences in Accounting Education. *J. Account. Education.* 16, 507–516. doi:10.1016/s0748-5751(98)00021-9

the manuscript and read and approved the final submitted version.

FUNDING

H3ABioNet is supported by the National Human Genome Research Institute of the National Institutes of Health under Award Number U24HG006941. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

ACKNOWLEDGMENTS

We would like to express our gratitude to the members of the H3ABioNet Education and Training Work Package who have contributed to much of the work discussed in this paper. We would also like to thank all the trainers who have contributed towards the development and presentation of the H3ABioNet training events as well as all the course participants. Finally we would like to thank Wisdom Akuguru for assisting in extracting and compiling the statistics for the H3ABioNet YouTube account.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/feduc.2021.725702/full#supplementary-material>

Supplementary Table S1 | | List of H3ABioNet training events by topic and category.

- Ben Hamda, C., Sangeda, R., Mwita, L., Meintjes, A., Nkya, S., Panji, S., et al. (2018). A Common Molecular Signature of Patients With Sickle Cell Disease Revealed by Microarray Meta-Analysis and a Genome-Wide Association Study. *Plos one.* 13, e0199461. doi:10.1371/journal.pone.0199461
- Bourne, P. E. (2021). Is “Bioinformatics” Dead? *Plos Biol.* 19, e3001165. doi:10.1371/JOURNAL.PBIO.3001165
- Brancaccio-Taras, L., Gull, K. A., and Ratti, C. (2016). The Science Teaching Fellows Program: A Model for Online Faculty Development of Early Career Scientists Interested in Teaching. *J. Microbiol. Biol. Education.* 17 (3), 333–338. doi:10.1128/jmbe.v17i3.1243
- Chauke, P. (2021). *Data Science Needs Input from Women*. Johannesburg: Mail and Guardian. Available at: <https://mg.co.za/opinion/2021-04-11-data-science-needs-input-from-women/>.
- Choudhury, A., Aron, S., Botigué, L. R., Sengupta, D., Botha, G., Bensellak, T., et al. (2020). High-Depth African Genomes Inform Human Migration and Health. *Nature.* 586, 741–748. doi:10.1038/s41586-020-2859-7
- Cook, S. J., Parker, R. S., and Pettijohn, C. E. (2004). The Perceptions of Interns: A Longitudinal Case Study. *J. Education Business.* 79, 179–185. Available at: <https://www.proquest.com/scholarly-journals/perceptions-interns-longitudinal-case-study/docview/202821430/se-2?accountid=15083>.
- Davis, S., Button-Simons, K., Bensellak, T., Ahsen, E. M., Checkley, L., Foster, G. J., et al. (2019). Leveraging Crowdsourcing to Accelerate Global Health Solutions. *Nat. Biotechnol.* 37, 848–850. doi:10.1038/s41587-019-0180-5
- Fadlelmola, F. M., Ghedira, K., Hamdi, Y., Hanachi, M., Radouani, F., Allali, I., et al. (2021). H3ABioNet Genomic Medicine and Microbiome Data Portals

- Hackathon Proceedings. *Database: J. Biol. databases curation*. 2021, 1–13. doi:10.1093/database/baab016
- Garcia, L., Batut, B., Burke, M. L., Kuzak, M., Psomopoulos, F., Arcila, R., et al. (2020). Ten Simple Rules for Making Training Materials FAIR. *Plos Comput. Biol.* 16, e1007854. doi:10.1371/journal.pcbi.1007854
- Ghouila, A., Siwo, G. H., Entfellner, J. D., Panji, S., Button-Simons, K. A., Davis, S. Z., et al. (2018). Hackathons as a Means of Accelerating Scientific Discoveries and Knowledge Transfer. *Genome Res.* 28, 759–765. doi:10.1101/gr.228460.117
- Groen, D., and Calderhead, B. (2015). Science Hackathons for Developing Interdisciplinary Research and Collaborations. *eLife*. 4, e09944. doi:10.7554/eLife.09944
- Gurwitz, K. T., Aron, S., Panji, S., Maslamoney, S., Fernandes, P. L., Judge, D. P., et al. (2017). Designing a Course Model for Distance-Based Online Bioinformatics Training in Africa: The H3ABioNet Experience. *Plos Comput. Biol.* 13, e1005715. doi:10.1371/journal.pcbi.1005715
- Hamdi, Y., Zass, L., Othman, H., Radouani, F., Allali, I., Hanachi, M., et al. (2021). Human OMICs and Computational Biology Research in Africa: Current Challenges and Prospects. *OMICS*. 25, 213–233. doi:10.1089/omi.2021.0004
- Hernández-de-Diego, R., de Villiers, E. P., Klingström, T., Gourel, H., Conesa, A., and Bongcam-Rudloff, E. (2017). The eBioKit, a Stand-Alone Educational Platform for Bioinformatics. *Plos Comput. Biol.* 13, e1005616–14. doi:10.1371/journal.pcbi.1005616
- Hostetter, M. K. (2002). Career Development for Physician-Scientists: The Model of the Pediatric Scientist Development Program. *J. Pediatr.* 140, 143–144. doi:10.1067/mpd.2002.121584
- Jongeneel, C. V., Achinike-Oduaran, O., Adebisi, E., Adebisi, M., Adeyemi, S., Akanle, B., et al. (2017). Assessing Computational Genomics Skills: Our Experience in the H3ABioNet African Bioinformatics Network. *Plos Comput. Biol.* 13, e1005419–10. doi:10.1371/journal.pcbi.1005419
- King, L. (2013). Helping Early Career Research Scientists Ascend the Professional Ladder. *Trends Biochem. Sci.* 38, 373–375. doi:10.1016/j.tibs.2013.06.001
- Kumuthini, J., Zass, L., Panji, S., Salifu, S. P., Kayondo, J. K., Nembaware, V., et al. (2019). The H3ABioNet Helpdesk: an Online Bioinformatics Resource, Enhancing Africa's Capacity for Genomics Research. *BMC Bioinformatics*. 20, 741–747. doi:10.1186/s12859-019-3322-3
- Lyndon, M. P., Cassidy, M. P., Celi, L. A., Hendrik, L., Kim, Y. J., Gomez, N., et al. (2018). Hacking Hackathons: Preparing the Next Generation for the Multidisciplinary World of Healthcare Technology. *Int. J. Med. Inform.* 112, 1–5. doi:10.1016/j.ijmedinf.2017.12.020
- Mccullough, H. (2010). Using Personal Development Planning for Career Development With Research Scientists in Sub-saharan Africa PhD thesis, University of Liverpool, 4–61. doi:10.17037/PUBS.03894557
- Mlotshwa, B. C., Mwesigwa, S., Mboowa, G., Williams, L., Retshabile, G., Kekitiinwa, A., et al. (2017). The Collaborative African Genomics Network Training Program: A Trainee Perspective on Training the Next Generation of African Scientists. *Genet. Med.* 19, 826–833. doi:10.1038/gim.2016.177
- Mulder, N., Schwartz, R., Brazas, M. D., Brooksbank, C., Gaeta, B., Morgan, S. L., et al. (2018). The Development and Application of Bioinformatics Core Competencies to Improve Bioinformatics Training and Education. *Plos Comput. Biol.* 14, e1005772. doi:10.1371/journal.pcbi.1005772
- Mulder, N., Adebamowo, C. A., Adebamowo, S. N., Adebayo, O., Adeleye, O., Alibi, M., et al. (2017a). Genomic Research Data Generation, Analysis and Sharing - Challenges in the African Setting. *Data Sci. J.* 16, 49. doi:10.5334/dsj-2017-049
- Mulder, N. J., Adebisi, E., Adebisi, M., Adeyemi, S., Ahmed, A., Ahmed, R., et al. (2017b). Development of Bioinformatics Infrastructure for Genomics Research. *Glob. Heart*. 12, 91–98. doi:10.1016/j.gheart.2017.01.005
- Mulder, N. J., Adebisi, E., Alami, R., Benkahla, A., Brandful, J., Doumbia, S., et al. (2016). H3ABioNet, a Sustainable Pan-African Bioinformatics Network for Human Heredity and Health in Africa. *Genome Res.* 26, 271–277. doi:10.1101/gr.196295.115
- Nchinda, T. C. (2002). Research Capacity Strengthening in the South. *Soc. Sci. Med.* 54, 1699–1711. doi:10.1016/s0277-9536(01)00338-0
- Nembaware, V., Mulder, N., and Mulder, N. (2019). The African Genomic Medicine Training Initiative (AGMT): Showcasing a Community and Framework Driven Genomic Medicine Training for Nurses in Africa. *Front. Genet.* 10, 1209. doi:10.3389/fgene.2019.01209
- Onah, D. F. O., Sinclair, J., and Boyatt, R. (2014). “Dropout Rates of Massive Open Online Courses : Behavioural Patterns MOOC Dropout and Completion: Existing Evaluations,” in Proceedings of the 6th International Conference on Education and New Learning Technologies EDULEARN14, 1–10. doi:10.13140/RG.2.1.2402.0009
- Prozesky, H., and Mouton, J. (2019). A Gender Perspective on Career Challenges Experienced by African Scientists. *S. Afr. J. Sci.* 115, 40–44. doi:10.17159/sajs.2019/5515
- Radouani, F., Zass, L., Hamdi, Y., Rocha, J. D., Sallam, R., Abdelhak, S., et al. (2020). A Review of Clinical Pharmacogenetics Studies in African Populations. *Per Med.* 17, 155–170. doi:10.2217/pme-2019-0110
- Ras, V., Botha, G., Aron, S., Lennard, K., Allali, I., Claassen-Weitz, S., et al. (2021). Using a Multiple-Delivery-Mode Training Approach to Develop Local Capacity and Infrastructure for Advanced Bioinformatics in Africa. *Plos Comput. Biol.* 17, e1008640. doi:10.1371/JOURNAL.PCBI.1008640
- Rotimi, C., Rotimi, C., Abayomi, A., Abimiku, A., Adabayeri, V. M., Adebamowo, C., et al. (2014). Research Capacity. Enabling the Genomic Revolution in Africa. *Science* 344, 1346–1348. doi:10.1126/science.1251546
- Scott, J. (1992). Scotland. *Adoption & Fostering* 16, 59. doi:10.1177/030857599201600412
- Sengupta, D., Choudhury, A., Fortes-Lima, C., Aron, S., Whitelaw, G., Bostoen, K., et al. (2021). Genetic Substructure and Complex Demographic History of South African Bantu Speakers. *Nat. Commun.* 12, 2080. doi:10.1038/s41467-021-22207-y
- Shaffer, J. G., Mather, F. J., Wele, M., Li, J., Tangara, C. O., Kassogue, Y., et al. (2019). Expanding Research Capacity in Sub-saharan Africa through Informatics, Bioinformatics, and Data Science Training Programs in Mali. *Front. Genet.* 10, 331–413. doi:10.3389/fgene.2019.00331
- Sonnert, G. (1999). Women in Science and Engineering: Advances, Challenges, and Solutions. *Ann. NY Acad. Sci.* 869, 34–57. doi:10.1111/j.1749-6632.1999.tb08353.x
- Stads, G.-J., and Beintema, N. M. (2006). *Women Scientists in Sub-saharan African Agricultural R&D. Brief Prepared for the USAID Meeting on Women in Science: Meeting the Challenge Lessons for Agricultural Sciences in Africa*. Washington, D.C: The Agricultural Science and Technology Indicators (ASTI). June 21.
- Tastan Bishop, Ö., Adebisi, E. F., Alzohairy, A. M., Everett, D., Ghedira, K., Ghouila, A., et al. (2015). Bioinformatics Education-Perspectives and Challenges Out of Africa. *Brief Bioinform.* 16, 355–364. doi:10.1093/bib/bbu022
- Welch, L., Lewitter, F., Schwartz, R., Brooksbank, C., Radivojac, P., Gaeta, B., et al. (2014). Bioinformatics Curriculum Guidelines: Toward a Definition of Core Competencies. *Plos Comput. Biol.* 10, e1003496. doi:10.1371/journal.pcbi.1003496

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Aron, Chauke, Ras, Panji, Johnston and Mulder. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.