American Journal of Laboratory Medicine
2025, Vol. 10, No. 1, pp. 20-31
https://doi.org/10.11648/j.ajlm.20251001.12

Research Article

Decentralising Bioinformatics Capacity: Lessons from Training Hospital Staff and Field Epidemiologists in Nigeria

Erkison Ewomazino Odih[1], **Odion O. Ikhimiukor**[1], **Rotimi Ayodeji Dada**[1], **Ifeoluwa Akintayo**[1], **Faith Ifeoluwa Oni**[1], **Elshama Queen Adanna Nwoko**[1], **Anthony Underwood**[2], **Iruka N. Okeke**[1], **Ayorinde Oluwatobiloba Afolayan**[1,*]

[1]Global Health Research Unit for the Genomic Surveillance of Antimicrobial Resistance, Department of Pharmaceutical Microbiology, Faculty of Pharmacy, University of Ibadan, Ibadan, Nigeria

[2]Centre for Genomic Pathogen Surveillance, Big Data Institute, University of Oxford, Oxford, United Kingdom

Abstract

Genomics is increasingly utilised across Africa to address pressing public health challenges, including antimicrobial resistance (AMR). However, the continent's capacity for sequence data analysis and interpretation remains limited, particularly outside research institutions. To bridge this gap, we conducted a virtual bioinformatics training program in Nigeria, targeting hospital laboratory staff, medical personnel, and field epidemiologists, to build capacity for genome sequence analysis and interpretation. The training consisted of two modules: Module 1 introduced sequencing technologies, AMR prediction, bacterial typing, and phylogenetics using web-based tools, while Module 2 focused on command-line tools and piloted Nextflow Tower for decentralized sequence analysis. Post-course and follow-up surveys assessed the program's impact. Twenty-two participants from nine institutions completed Module 1, with ten from six institutions progressing to Module 2. Interactive sessions facilitated knowledge retention, with 83.3% of participants rating Module 2 as highly relevant. The Nextflow Tower platform facilitated cost-effective bacterial genome analysis ($0.0026 per genome). While confidence in web-based tools improved significantly, challenges remained in adopting command-line tools. Feedback highlighted the effectiveness of pre-recorded lectures, interactive engagement, and decentralized analysis platforms. This training program enhanced bioinformatics capacity among Nigerian public health professionals, highlighting the feasibility of implementing centralized sequencing with decentralized bioinformatics analysis in resource-limited settings. Furthermore, it highlights the importance of prioritizing introductory bioinformatics and web-based tools as a foundation for building long-term genomic surveillance capacity.

Keywords

Antimicrobial Resistance, Bioinformatics, Capacity Building, Public Health, Nigeria, Africa

1. Introduction

The rapid advancement of next-generation sequencing (NGS) technologies has transformed microbial genomics, providing unmatched insights into the epidemiology, transmission, and evolution of infectious diseases [1-3]. The global

---

[*] Corresponding author: afolayanayorinde@gmail.com (Ayorinde Oluwatobiloba Afolayan)

Received: 7 February 2025; **Accepted:** 22 February 2025; **Published:** 7 March 2025

significance of NGS became particularly evident during the COVID-19 pandemic, where its integration into public health systems enhanced outbreak detection, informed infection prevention strategies, and accelerated progress toward personalized medicine and improved patient care [4-6]. As these technologies continue to evolve, NGS is expected to play an increasingly prominent role in clinical decision-making and public health [7, 8].

Despite its potential, the adoption of NGS in low- and middle-income countries (LMICs), including Nigeria, remains limited due to barriers such as inadequate bioinformatics capacity, insufficient infrastructure, and lack of institutional support [9-12]. Among these challenges, building bioinformatics capacity stands out as a critical but solvable issue. Training a workforce skilled in bioinformatics is essential for empowering LMICs to conduct genomic-informed pathogen surveillance and monitor antimicrobial resistance (AMR) [8, 10].

National and international initiatives have increasingly prioritized this need, with programs like H3Africa, the Africa Pathogen Genomics Initiative (Africa PGI), and the African Bioinformatics Institute leading efforts to expand bioinformatics expertise across the continent [13-15]. These initiatives, supported by organizations such as the Africa CDC, have enabled countries to leverage genomic data for public health interventions. However, persistent challenges such as visa restrictions, limited funding, and high travel costs often hinder access to these opportunities, particularly for individuals in remote regions [16, 17]. Within Nigeria, even domestic travel costs can prevent participation in training programs outside local areas.

The Nigerian National Action Plan to combat AMR identified bioinformatics capacity development as a cornerstone for achieving robust genomic surveillance within hospitals and communities [18]. The Global Health Research Unit for Genomic Surveillance of Antimicrobial Resistance (GHRU-GSA), Nigeria, hosted at the University of Ibadan, plays a pivotal role in supporting the Nigeria Centre for Disease Control and Prevention's (NCDC) AMR surveillance efforts. By generating valuable sequencing data and associated reports, GHRU Nigeria drives genomic surveillance of multidrug-resistant and high-priority pathogens, contributing significantly to the country's AMR response [19-22]. Typically, isolates collected from sentinel healthcare facilities nationwide are processed at the National AMR Reference Laboratory at the University College Hospital, Ibadan, before being sent to the GHRU Nigeria genomics laboratory for whole genome sequencing. The resulting sequences are analyzed by resident GHRU Nigeria bioinformaticians, adhering to national AMR surveillance protocols and the specific requirements of sentinel sites. Despite these efforts, bioinformatics expertise within Nigeria remains largely centralized within GHRU Nigeria and a handful of other genomics centres. Disseminating these skills across Nigeria is critical to ensuring that genomic data is efficiently analyzed and translated into actionable insights that inform national policies and guidelines. Furthermore, equitable distribution of expertise is vital to addressing biases toward hospital-derived data, ensuring that surveillance efforts encompass all geographic regions and include non-human niches.

Historically, bioinformatics capacity-building initiatives in Nigeria have overlooked key personnel such as clinical microbiologists, who manage patients directly, medical laboratory scientists, who generate data in clinical microbiology laboratories, and field epidemiologists, who are critical to public health surveillance [23, 24]. This gap has limited hospitals and public health institutions from fully utilizing genomic information for infection prevention and control. To ensure that genomic technologies make a meaningful and sustainable impact, they must be accompanied by training programs tailored to local contexts. Such programs must equip individuals with the skills to independently analyze and interpret NGS data, fostering long-term capacity at all levels.

We organized a targeted, virtual bioinformatics training course in Nigeria for hospital laboratory staff, medical personnel, and field epidemiologists. This training aimed to bridge the knowledge gap by equipping participants with foundational skills for genome sequence analysis and interpretation. The goal was to enable local experts to generate biologically relevant insights that could guide infection prevention and control practices in their institutions. Focusing on in-country capacity building, the initiative provided an entry point for key personnel to contribute directly to genomic-informed AMR surveillance.

2. Methodology

2.1. Ethical Considerations

We taught a bioinformatics training course held in two modules between August 16, 2021, and October 27, 2021, and initiated research to study the impact of the course on August 1, 2022. The research study received ethical approval from the UI/UCH Ethics Committee, with reference number UI/EC/22/0084, before any research tasks commenced. Written informed consent was obtained from all participants prior to their involvement in the study.

2.2. Participant Recruitment

The Nigeria Centre for Disease Control (NCDC) oversees antimicrobial resistance (AMR) surveillance activities in sentinel hospital laboratories as part of the national action plan on AMR. In collaboration with NCDC, sentinel hospital laboratories were invited to nominate at least one staff member to participate in a virtual bioinformatics training workshop. Selected participants in the Nigeria Field Epidemiology and Laboratory Training Program (NFELTP) were also invited. The workshop was hosted by GHRU-GSA Nigeria, Department of Pharmaceutical Microbiology, Faculty of Pharmacy, University of Ibadan, Oyo State, Nigeria, to equip participants with the skills to integrate genomics into clinical microbiology.

2.3. Training Module 1: Virtual Bioinformatics Workshop (Web-based)

The first virtual bioinformatics training course (Module 1) was conducted from August 16 to 25, 2021 (Figure 1). It served as an entry point for hospital laboratory scientists and epidemiologists to gain proficiency in online bioinformatics tools for AMR pathogen surveillance. The workshop curriculum introduced participants to key bioinformatics topics essential for AMR surveillance, including whole-genome sequencing (WGS), quality control, species identification, phylogenetic analysis, bacterial sub-typing, plasmid detection, and AMR and virulence gene prediction. The content was designed to be relevant, up-to-date, and of high quality, leveraging sound educational theories learned by trainers during a train-the-trainer course at the Wellcome Sanger Institute in 2019 [25].

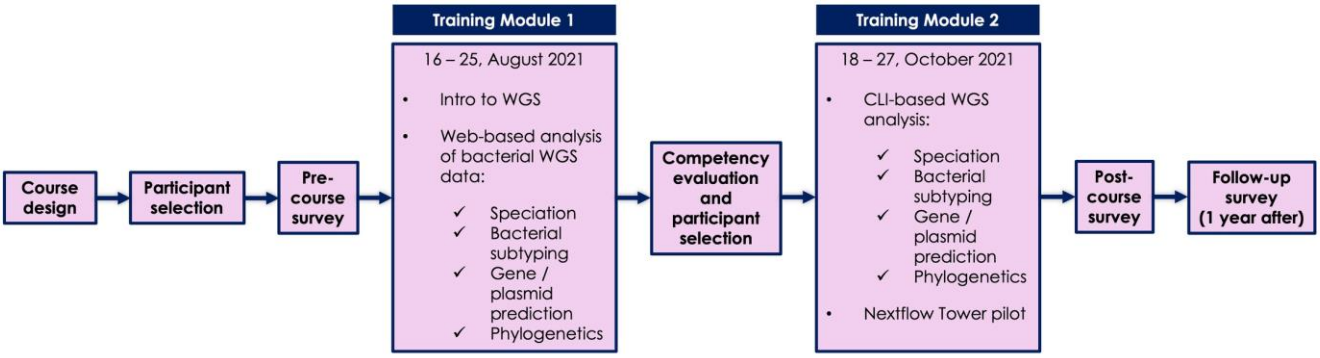


Figure 1. Flowchart of activities from course design to the follow-up survey.

2.4. Training Format and Delivery

To address potential challenges with internet connectivity or other unforeseen occurrences during the virtual training, a hybrid delivery approach was adopted:

Pre-recorded lectures: Trainers recorded their sessions in advance, ensuring high-quality content delivery. Furthermore, trainers activated webcams during recordings to maintain a personal connection and motivate participants.

Live demonstrations and hands-on sessions: Live sessions, conducted via Zoom or Gather Town, included practical demonstrations followed by hands-on exercises.

Recorded sessions: All Zoom sessions were recorded, and links were shared with participants at the end of each workshop day for review.

Participants used publicly available web-based tools such as the Centre for Genomic Epidemiology Platform [26, 27] and Pathogenwatch [28, 29] during this module. These tools were selected for their high quality, active maintenance and development, usefulness for analysing a wide variety of species of public health importance, and the comprehensive bioinformatics functionalities they offer. In addition, their accessibility, user-friendly interfaces, and minimal computational and connectivity requirements makes them suitable for settings with limited internet bandwidth. Participants from sentinel facilities in the Nigeria Antimicrobial Surveillance Network (NASN) were provided with whole genome sequence data of isolates submitted by their respective institutions to GHRU. This approach ensured that the exercises were both relevant and personalized. For participants from institutions not yet enrolled in NASN, publicly available genome datasets were provided. Assignments were given at the end of each training day to reinforce learning, and trainers were available on Gather Town during work hours (8:00 to 17:00) to address participant queries. Each training session lasted approximately three hours per day (8:45 – 11:30/12:00).

Participants who completed assignments during or after Module 1 were considered for continued training, as assignment completion was used as a metric for engagement and commitment.

2.5. Training Module 2: Virtual Bioinformatics Workshop (Command-line Based)

The second training module was conducted from October 18 to 27, 2021 (Figure 1). This module built on the foundational concepts introduced in Module 1, focusing on the use of bioinformatics tools and pipelines on the command-line and on Nextflow Tower (now Seqera Platform). Participants were introduced to the benefits and practical use of command-line tools for bioinformatics analysis, replicating tasks previously performed in the first training module.

2.6. Nextflow Tower Architecture for Decentralised Bioinformatics Analysis

During Module 2, we piloted a decentralised sequence data dissemination and analysis architecture as an alternative to the current model of sequence data flow and analysis in the Nigerian AMR surveillance system. This architecture utilised Amazon Web Services (AWS) for computational and data storage services, alongside the pipeline deployment and execution functionalities offered by the Nextflow Tower platform (now Seqera Platform). Using custom AWS *IAM* policies and access restrictions, users from all sentinel sites were granted programmatic and password-based access to their respective sentinel site data directories within a dedicated S3 bucket. A scalable compute environment was also configured using AWS batch, with default settings except for setting instances to 'spot' and the allocation strategy set to 'SPOT_CAPACITY_OPTIMIZED'. Next, to facilitate programmatic pipeline deployment and execution, we set up an administrative user with full access to the S3 buckets and configured compute environment. This user was granted permissions specified in the Nextflow Tower AWS Batch Forge policy (available at https://github.com/seqeralabs/nf-tower-aws/tree/master/forge).

For the pilot, we forked the existing GHRU AMR prediction pipeline (https://gitlab.com/cgps/ghru/pipelines/dsl2/pipelines/amr_prediction) and modified it to ensure compatibility with the Nextflow Tower set up (https://gitlab.com/bioinfo_erkison/amr-prediction-twr). This pipeline used the ARIBA tool [30] for AMR prediction, using the acquired AMR gene and PointFinder databases of the National Center for Biotechnology Information [31] Additionally, the hAMRonization tool [32] was integrated into the pipeline to enable standardized AMR gene reporting. The modified pipeline was then deployed to the Nextflow Tower platform.

In the decentralised sequence data flow and analysis model (Figure 2), bioinformaticians at the central genomics reference laboratory upload raw sequence data for each sentinel site to their respective access-protected directories on AWS, ensuring secure, decentralised access to their sequence data while maintaining data privacy. The central bioinformaticians also develop, deploy, and update Nextflow pipelines for sequence analysis within the Nextflow Tower workspace. Users with appropriate access credentials can then launch these deployed pipelines on specified datasets through the Nextflow Tower interface, configuring the pipeline parameters as needed. To retrieve analysis results, users can connect to the AWS S3 buckets using tools such as Cyberduck (https://cyberduck.io/), which supports secure file transfer protocols (e.g., SFTP). Importantly, users can also access their raw sequence data via FTP using Cyberduck or other compatible FTP clients.

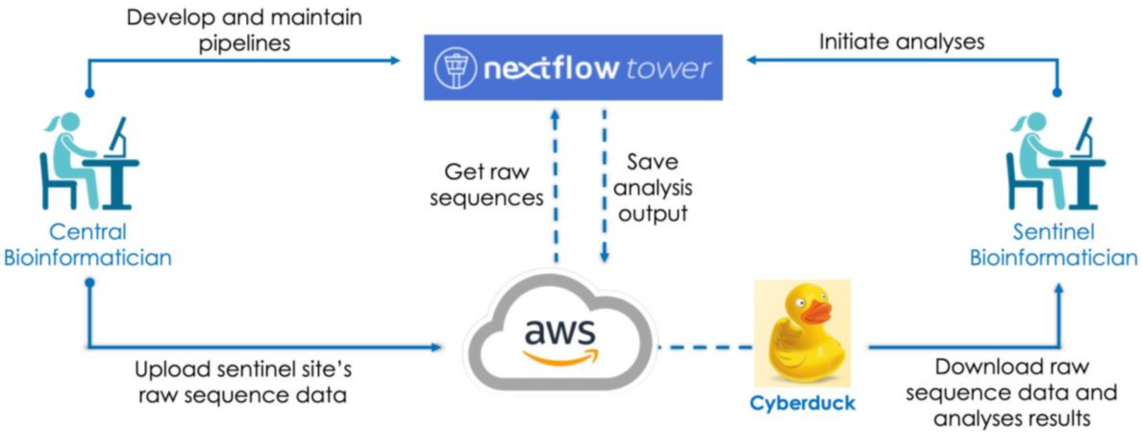


Figure 2*. The Nextflow Tower (Seqera Platform) Architecture. This involves deployment and maintenance of Nextflow pipelines on AWS Cloud computing by the Central Bioinformatician, and the initiation of analysis and retrieval of analysis results by the sentinel bioinformatician.*

Using the Nextflow Tower platform, all Module 2 participants securely accessed and analyzed their genome sequence data and generated standardized AMR reports. As in Module 1, we used Gather Town to facilitate hands-on exercises and real-time trainer support, collaboration among participants to exchange AMR genomic surveillance results, and the visualization of the potential impact of regular data exchange beyond training sessions.

2.7. Post-Course Survey and Follow-Up Survey

Two surveys were conducted at different intervals to evaluate the impact and effectiveness of the training. After Module 2, a post-course survey was distributed via SurveyMonkey to assess the usefulness of the topics covered, participants' level of satisfaction, whether their expectations were met, and their confidence in applying the skills gained.

A follow-up survey was conducted one year after the completion of Module 2 to assess the long-term impact of the training. Likert-scale and binary responses were visualised using the HH (v3.1.52) package [33] in R (v4.2.3) [34], providing insights into the training's strengths and areas for enhancement.

2.8. Provision of Equipment

To address the challenge of inadequate resources for genomic surveillance in sentinel hospitals, laptops and essential accessories were provided to each sentinel site and to NFELTP participants at the end of the training modules. Each laptop featured an Intel(R) Core™ i5-10210U processor (1.60GHz – 2.11GHz), a 64-bit operating system, and x64-based architecture, ensuring compatibility with bioinformatics tools. The availability of laptops ensured participants were adequately prepared to begin integrating genomics into routine clinical microbiology practices.

3. Results

3.1. Participant's Socio-demographic Characteristics and Professional Qualification

A total of 22 participants from ten public health institutions across six states participated in the first virtual bioinformatics training (Figure 3A). Of these, ten (~45%) were medical professionals, while the rest were either field epidemiologists (n = 3), medical laboratory scientists (n = 8), or research scientists (n = 1). Ten (~45%) of the twenty-two participants were female. All participants had a minimum educational qualification of a BSc, demonstrating high academic achievement.

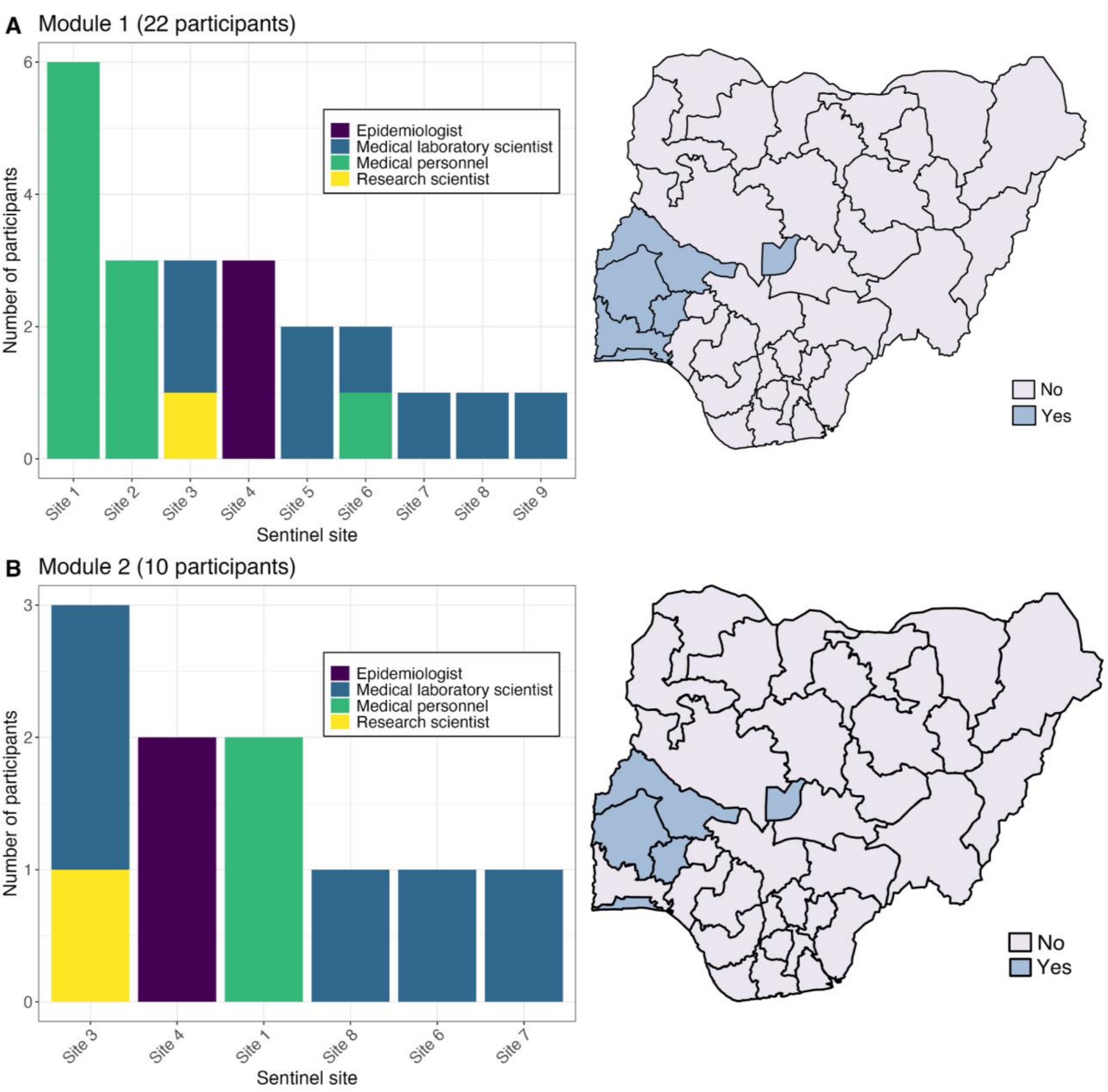


Figure 3. Number of individuals from sentinel sites in the country that participated in the first (A) and second (B) virtual bioinformatics workshop.

Of the twenty-two participants who joined the first virtual bioinformatics training, ten moved on to the second virtual bioinformatics training, and they were from 6 public health institutions in five states (Figure 3B). This group included medical laboratory scientists (n = 5), medical professionals (n = 2), field epidemiologists (n = 2), and a research scientist (n = 1). Of these, six participants (60%) were female.

3.2. Pilot Implementation of the Nextflow Tower Platform

Using the Nextflow Tower platform on AWS cloud computers, the average cost to complete the AMR prediction pipeline for one *Escherichia coli* genome was $0.0026 USD, with an average completion time of 2.5 minutes (Figure 4).

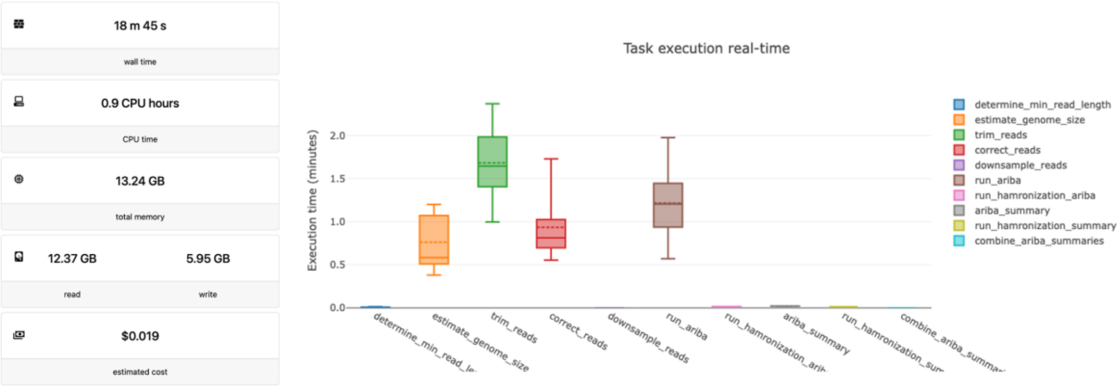
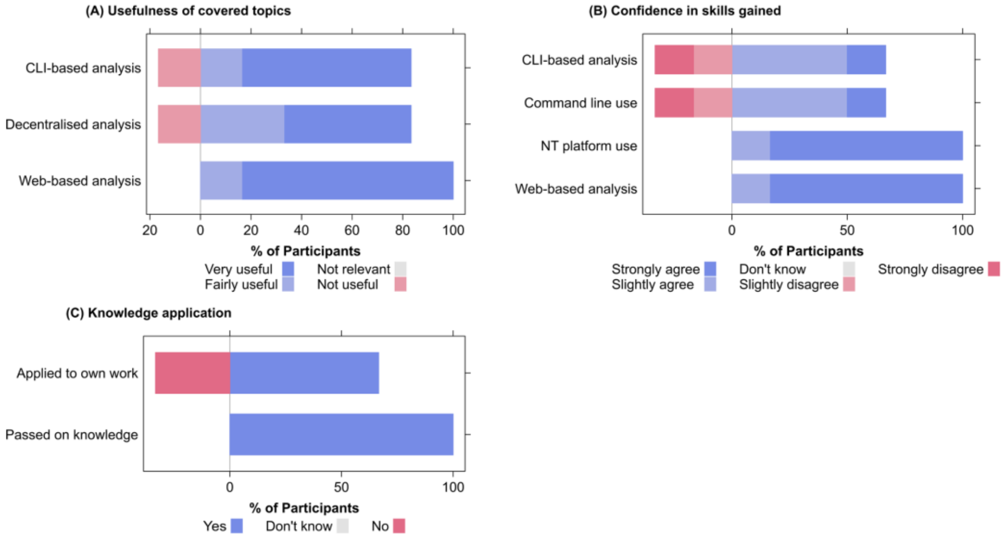


Figure 4*.* *Computational efficiency in the AWS cloud infrastructure. Boxplots show the execution times for each task in the AMR prediction pipeline run on nine (9) Escherichia coli genomes. Overall execution time, physical memory usage, single core CPU usage, and computation costs are highlighted on the left of the plots.*

3.3. Post-Course Survey

Seven out of ten participants responded to the post-course survey, providing valuable insights into their experiences (Figure 5). Overall, participants reported high satisfaction with the workshop content, delivery, and structure. The interactive format, coupled with practical exercises using freely available web-based and command-line tools, significantly enhanced engagement and learning. On a scale of 1 to 10, the perceived benefit of the course averaged 9.57.



CLI – Command Line; NT: Nextflow Tower

Figure 5*.* *Summary of post-workshop survey results showing A.) Usefulness of covered topics, B.) Confidence in skills gained, and C.) Knowledge application.*

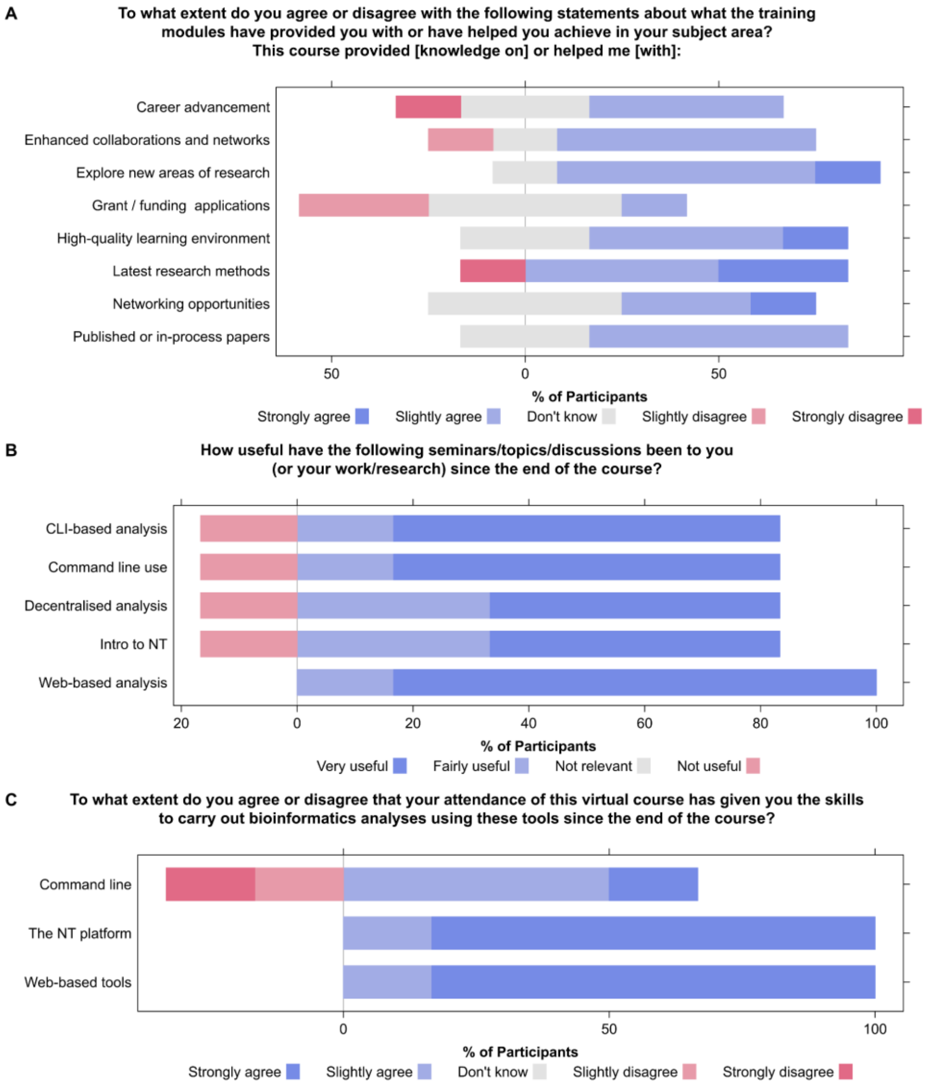
Participants described the workshops as "insightful," "simple," "hands-on," and "well-explained," noting that the tools and resources were highly applicable to their work. The Gather Town platform was particularly appreciated for fostering collaboration and ensuring no participant was left behind, especially during sessions where Zoom lectures could feel overwhelming. Trainers were commended for their patience and willingness to address questions thoroughly.

A strong majority (>90%) found the web-based tools highly useful. However, the utility and confidence in using command-line tools were lower (~50%), with many participants expressing the need for additional practice to build proficiency. Confidence in using the Nextflow Tower platform also improved significantly by the end of the course, though some participants suggested extending training sessions on both the command-line interface and Nextflow to solidify their skills.

All participants reported sharing the knowledge gained with colleagues in their respective institutions, with over 60% successfully applying these skills in their professional and personal work. One participant particularly appreciated the short duration of daily sessions, which helped maintain focus and prevent fatigue.

Participants also highlighted a few areas for improvement. One suggestion was to provide notifications for the course well in advance, allowing participants to secure training leave or arrange for duty exemptions. Several participants preferred physical training sessions to address the challenges posed by unreliable internet connections, which occasionally hindered full participation during virtual sessions.

3.4. Follow-up Survey



CLI – Command Line; NT: Nextflow Tower

Figure 6*. Summary of 1-year follow-up survey results showing A.) Application of knowledge gained B.) Usefulness of covered topics in work/research C.) Confidence in skills gained.*

A follow-up survey conducted a year after the workshop had six out of ten participants responding. Respondents reported continued utility of the training, with most applying the knowledge gained to their work, sharing it with colleagues, and enhancing AMR surveillance efforts. Five participants rated the course as highly useful for providing practical experience and insights into modern research (Figure 6A).

Almost all respondents agreed that the training positively impacted their career progression, aiding in manuscript publication, fostering collaboration, and expanding professional networks (Figure 6A). Examples of skill application included assisting colleagues with sequence data analysis using online tools, interpreting molecular datasets from literature, drafting WGS protocols for postgraduate dissertations, and interpreting Global Antimicrobial Resistance and Use Surveillance System (GLASS; https://www.who.int/initiatives/glass) reports for their facilities.

Despite these successes, some participants faced challenges, particularly in applying command-line tools (Figure 6B and Figure 6C). Many indicated a preference for online tools due to their ease of use. However, one participant demonstrated the potential for practical application by using command-line skills during an observership at a reference laboratory in 2022. Participants also disseminated their knowledge, sharing skills such as building phylogenetic trees, developing protocols for the Nextflow platform, and organizing training sessions using web-based tools like Microreact [35] and Pathogenwatch.

4. Discussion

This study aimed to evaluate the outcomes of two virtual bioinformatics training workshops conducted in 2021, designed to provide foundational bioinformatics skills to hospital laboratory staff and field epidemiologists in Nigeria. The overarching goal was to bridge the gap between the growing need for genomic surveillance of AMR and the lack of computational infrastructure, bioinformatics expertise, and institutional support. This initiative was informed by our experience co-teaching two SEQAFRICA (https://www.food.dtu.dk/english/topics/antimicrobial-resistance/seqafrica) bioinformatics training workshops in early 2021, where we recognized the enthusiasm of Nigerian public health workers to acquire bioinformatics expertise despite systemic challenges. Additionally, we recognized a gap in bioinformatics training within Nigeria, as many previous programs focused on researchers or academic institutions, with little emphasis on public health personnel or field epidemiologists, who are essential for AMR surveillance and genomic public health interventions. Most training initiatives also lacked a practical framework for real-world applications in resource-limited settings and did not explore scalable solutions, such as cloud-based platforms, for decentralized bioinformatics analysis. This study sought to fill this gap by offering tailored, practical bioinformatics training that directly addressed the needs of public health professionals in Nigeria.

One of the workshop's key successes was the involvement of highly skilled and experienced trainers. Several trainers had benefited from participation in the Wellcome Genome Campus Train-the-Trainer program in 2019 [25], while others were seasoned bioinformatics practitioners with strong backgrounds in molecular biology and next-generation sequencing. This foundational expertise contributed significantly to the quality of the workshops. The workshops also provided opportunities for capacity building among volunteer staff, many of whom were PhD candidates or postdoctoral researchers. Their involvement in behind-the-scenes logistics and teaching roles enriched their professional development [36] and self-confidence [37], aligning with the broader goal of expanding educational capacity within the bioinformatics field. We shared our success story on the Public Health Alliance for Genomic Epidemiology’s website (https://pha4ge.org/articles/pha4ge-amr-sub-award-success-story-team-nigeria/).

The integration of pre-recorded lectures and interactive sessions via Zoom and Gather Town addressed the challenges posed by poor internet connectivity, a common barrier in resource-limited settings. This approach not only reduced redundancy in content delivery but also allowed participants to revisit lecture materials as needed, enhancing their learning experience. By combining this innovative instructional method with the use of cloud-based platforms like Nextflow Tower for bioinformatics analysis, the training demonstrated the feasibility of leveraging modern, cost-effective tools for genomic research in resource-constrained environments. While the specific functionalities of Nextflow Tower have evolved since the study, the overarching model of integrating centralized sequencing hubs with decentralized bioinformatics capacity remains valid, underscoring its adaptability for scalable genomic surveillance. This learning model holds promise for broader adaptation to similar settings globally.

Despite the successes, several areas for improvement were identified. One notable issue was the pre-course quiz, administered via SurveyMonkey. While it was intended to gauge participants' readiness and knowledge of molecular biology, it is likely that some participants might have used internet resources to complete the quiz. This could have led to the selection of individuals who lacked foundational molecular biology knowledge, which subsequently made it challenging for them—and at times their peers—to keep pace during the training. While monitoring completion times provided some insights, future iterations would benefit from more robust and secure assessment methods to ensure that participants possess the requisite background knowledge. To address this gap, we developed a pre-assessment and a foundational "Module Zero" to better prepare future participants that self-select to take it and have potential participants take pre-course quiz after module 0 has been delivered.

Another issue was the low retention rate, with only 10 out of 22 participants advancing to Module 2. Participants who completed assignments during or after Module 1 were considered for continued training, as assignment completion was used as a metric for engagement and commitment. Only 10 met the criteria, suggesting that workload, competing professional responsibilities, or difficulties in keeping up with the course material may have contributed to attrition. To improve retention, future workshops will incorporate structured mentorship, flexible pacing, and additional support resources to help participants balance training with professional responsibilities.

Participant feedback revealed that while web-based tools were widely appreciated, confidence in using command-line tools remained low. This suggests the need for longer and more focused sessions to build participants' confidence and competence in this area. Additionally, limited access to datasets hindered some participants from applying their skills, highlighting a systemic barrier that training programs alone cannot resolve. Notably, one participant successfully implemented command-line tools during an observership, illustrating the potential for practical application when adequate support is provided. Expanding the duration of training programs and incorporating more structured, hands-on practice sessions could further bridge the gap between basic users and advanced bioinformatics practitioners.

Looking ahead, we aim to organize regular training sessions to expand the pool of skilled personnel capable of genomic surveillance of AMR. Equivalent Modules 1 and 2 have been taught to other audiences, preceded by a module 0. A third bioinformatics module focusing on advanced sequence data analysis and interpretation is planned, further enhancing participants' capabilities. Following this, a dedicated train-the-trainer program will empower participants to become instructors, creating a multiplier effect to expand bioinformatics expertise at local, regional, and national levels. To ensure standardization and quality, we plan to implement external quality assurance measures and work toward the accreditation of sentinel laboratories. Efforts are also underway to make analysis pipelines more accessible via the evolved Seqera platform (formerly Nextflow Tower), which will cater to public health practitioners with limited command-line experience. Ensuring sustainability and scalability will require securing funding for storage and computation costs on cloud platforms like Amazon Web Services.

Our study bridges the critical gap in bioinformatics training by targeting hospital laboratory staff, medical personnel, and field epidemiologists—key stakeholders in AMR surveillance. By integrating foundational and advanced genomic analysis training with practical applications, we have demonstrated a replicable model for enhancing genomic capacity in public health. Furthermore, the introduction of the Nextflow Tower platform (now Seqera platform) offers an innovative solution to decentralizing bioinformatics analysis while leveraging centralized sequencing hubs. This approach enables cost-effective and scalable genomic data processing tailored to the needs of resource-limited settings. Our focus on web-based tools as entry points for professionals with varying expertise ensures broader accessibility and inclusivity. The study also emphasizes sustainability by fostering train-the-trainer models and planning for future advanced modules, laying the groundwork for long-term capacity building in Nigeria and beyond.

This study highlights the importance of tailored bioinformatics training programs for public health personnel in low- and middle-income countries, especially those engaged in AMR surveillance and outbreak response. The successful implementation of our training program underscores the potential for scaling similar initiatives regionally and globally, fostering knowledge dissemination and creating a network of skilled professionals to address critical public health challenges. Policymakers and global health stakeholders can leverage the findings of this study to invest in sustainable bioinformatics capacity building as a cornerstone of genomic public health preparedness. Future research should focus on evaluating the long-term impact of such training initiatives on genomic surveillance outcomes and exploring advanced modular training for more complex analyses.

Finally, we emphasize that with creativity and dedication, initiatives like this can succeed in resource-limited settings. By focusing on accessible introductory tools and leveraging the growing availability of advanced web-based platforms, we can further democratize bioinformatics education and prepare the next generation of professionals to respond effectively to future outbreaks and pandemics.

Abbreviations

Africa PGI	Africa Pathogen Genomics Initiative
AMR	Antimicrobial Resistance
AWS	Amazon Web Services
GHRU-GSA	Global Health Research Unit for Genomic Surveillance of Antimicrobial Resistance
GLASS	Global Antimicrobial Resistance and Use Surveillance System
LMICs	Low- and Middle-Income Countries
NFELTP	Nigeria Field Epidemiology and Laboratory Training Program
NASN	Nigeria Antimicrobial Surveillance Network
NGS	Next-Generation Sequencing
NCDC	Nigeria Centre for Disease Control and Prevention
UI/UCH	University of Ibadan - University College Hospital
WGS	Whole-Genome Sequencing
CLI	Command-line
NT	Nextflow Tower

Acknowledgments

We acknowledge Jolaade O Ajiboye, the program officer of GHRU-GSA Nigeria for working with the trainers to ensure seemless communication between the trainers and participants.

Author Contributions

Erkison Ewomazino Odih: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Visualization, Writing – original draft, Writing – review and editing

Odion O. Ikhimiukor: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Writing – review and editing

Rotimi Ayodeji Dada: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Writing – review and editing

Ifeoluwa Akintayo: Data curation, Formal analysis, Investigation, Methodology, Writing – review and editing

Faith Ifeoluwa Oni: Data curation, Formal analysis, Investigation, Methodology, Writing – review and editing

Elshama Queen Adanna Nwoko: Data curation, Formal analysis, Writing – review and editing

Anthony Underwood: Conceptualization, Investigation, Methodology, Validation, Software, Writing – review and editing

Iruka N. Okeke: Conceptualization, Funding acquisition, Methodology, Validation, Supervision, Resources, Writing – review and editing

Ayorinde Oluwatobiloba Afolayan: Conceptualization, Data curation, Formal analysis, Funding acquisition, Project Administration, Investigation, Methodology, Visualization, Supervision, Writing – original draft, Writing – review and editing

Funding

This training was funded by the Public Health Alliance for Genomic Epidemiology (Reference: PHA4GE/2021-01) as a sub-grant promoting sustainable development in bioinformatics to support public health and by Official Development Assistance (ODA) funding from the National Institute of Health Research (grant number 16_136_111). INO is a Calestous Juma Science Leadership fellow supported by the Bill & Melinda Gates Foundation (INV-036234). A. O. A. is a post-doctoral researcher supported by the Federal Ministry of Education and Research (BMBF) under 01KI2018.

Data Availability Statement

All training materials can be found at https://github.com/ayoraind/PHA4GE_bioinformatics_training_materials

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet 2016; 17: 333–351. https://doi.org/10.1038/nrg.2016.49

[2] Satam H, Joshi K, Mangrolia U, Waghoo S, Zaidi G, et al. Next-Generation Sequencing Technology: Current Trends and Advancements. Biology (Basel) 2023; 12: 997. https://doi.org/10.3390/biology12070997

[3] Gardy JL, Loman NJ. Towards a genomics-informed, real-time, global pathogen surveillance system. Nat Rev Genet 2017; 19: 9–20. https://doi.org/10.1038/nrg.2017.88

[4] Dylus D, Pillonel T, Opota O, Wüthrich D, Seth-Smith HMB, et al. NGS-Based S. aureus Typing and Outbreak Analysis in Clinical Microbiology Laboratories: Lessons Learned From a Swiss-Wide Proficiency Test. Front Microbiol 2020; 11: 2822. https://doi.org/10.3389/fmicb.2020.591093

[5] Argimón S, Masim MAL, Gayeta JM, Lagrada ML, Macaranas PKV, et al. Integrating whole-genome sequencing within the National Antimicrobial Resistance Surveillance Program in the Philippines. Nat Commun 2020; 11: 1–15. https://doi.org/10.1038/s41467-020-16322-5

[6] Chen X, Kang Y, Luo J, Pang K, Xu X, et al. Next-Generation Sequencing Reveals the Progression of COVID-19. Front Cell Infect Microbiol 2021; 11: 632490. https://doi.org/10.3389/fcimb.2021.632490

[7] Rossen JWA, Friedrich AW, Moran-Gilad J. Practical issues in implementing whole-genome-sequencing in routine diagnostic microbiology. Clinical Microbiology and Infection 2018; 24: 355–360. https://doi.org/10.1016/j.cmi.2017.11.001

[8] Jauneikaite E, Baker KS, Nunn JG, Midega JT, Hsu LY, et al. Genomics for antimicrobial resistance surveillance to support infection prevention and control in health-care facilities. Lancet Microbe 2023; 4: e1040–e1046. https://doi.org/10.1016/S2666-5247(23)00282-3

[9] Inzaule SC, Tessema SK, Kebede Y, Ogwell Ouma AE, Nkengasong JN. Genomic-informed pathogen surveillance in Africa: opportunities and challenges. Lancet Infect Dis 2021; 21: e281–e289. https://doi.org/10.1016/S1473-3099(20)30939-7

[10] Onywera H, Ondoa P, Nfii F, Ogwell A, Kebede Y, et al. Boosting pathogen genomics and bioinformatics workforce in Africa. Lancet Infect Dis 2024; 24: e106–e112. https://doi.org/10.1016/S1473-3099(23)00394-8

[11] Getchell M, Wulandari S, de Alwis R, Agoramurthy S, Khoo YK, et al. Pathogen genomic surveillance status among lower resource settings in Asia. Nature Microbiology 2024 9: 10 2024; 9: 2738–2747. https://doi.org/10.1038/s41564-024-01809-4

[12] Akintola AA, Aborode AT, Hamza MT, Amakiri A, Moore B, et al. Bioinformatics proficiency among African students. Frontiers in Bioinformatics 2024; 4: 1328714. https://doi.org/10.3389/fbinf.2024.1328714

[13] Bishop ÖT, Adebiyi EF, Alzohairy AM, Everett D, Ghedira K, et al. Bioinformatics Education—Perspectives and Challenges out of Africa. Brief Bioinform 2015; 16: 355–364. https://doi.org/10.1093/bib/bbu022

[14] Rotimi C, Abayomi A, Abimiku A, Adabayeri VM, Adebamowo C, et al. Research capacity. Enabling the genomic revolution in Africa. Science 2014; 344: 1346. https://doi.org/10.1126/science.1251546

[15] Mulder NJ, Adebiyi E, Adebiyi M, Adeyemi S, Ahmed A, et al. Development of Bioinformatics Infrastructure for Genomics Research. Glob Heart 2017; 12: 91. https://doi.org/10.1016/j.gheart.2017.01.005

[16] Gurwitz KT, Aron S, Panji S, Maslamoney S, Fernandes PL, et al. Designing a course model for distance-based online bioinformatics training in Africa: The H3ABioNet experience. PLoS Comput Biol;13. Epub ahead of print 1 October 2017. https://doi.org/10.1371/journal.pcbi.1005715

[17] Fatumo S, Shome S, Macintyre G. Workshops: A Great Way to Enhance and Supplement a Degree. PLoS Comput Biol 2014; 10: e1003497. https://doi.org/10.1371/journal.pcbi.1003497

[18] Okeke IN, Aboderin AO, Egwuenu A, Underwood A, Afolayan AO, et al. Establishing a national reference laboratory for antimicrobial resistance using a whole-genome sequencing framework: Nigeria's experience. Microbiology (United Kingdom) 2022; 168: 001208. https://doi.org/10.1099/mic.0.001208

[19] Ikhimiukor OO, Oaikhena AO, Afolayan AO, Fadeyi A, Kehinde A, et al. Genomic characterization of invasive typhoidal and non-typhoidal Salmonella in southwestern Nigeria. PLoS Negl Trop Dis 2022; 16: e0010716. https://doi.org/10.1371/journal.pntd.0010716

[20] Afolayan AO, Oaikhena AO, Aboderin AO, Olabisi OF, Amupitan AA, et al. Clones and Clusters of Antimicrobial-Resistant Klebsiella from Southwestern Nigeria. Clinical Infectious Diseases 2021; 73: S308–S315. https://doi.org/10.1093/cid/ciab769

[21] Afolayan AO, Aboderin AO, Oaikhena AO, Odih EE, Ogunleye VO, et al. An ST131 clade and a phylogroup A clade bearing an O101-like O-antigen cluster predominate among bloodstream Escherichia coli isolates from South-West Nigeria hospitals. Microb Genom 2022; 8: 000863. https://doi.org/10.1099/mgen.0.000863

[22] Odih EE, Oaikhena AO, Underwood A, Hounmanou YMG, Oduyebo OO, et al. High Genetic Diversity of Carbapenem-Resistant Acinetobacter baumannii Isolates Recovered in Nigerian Hospitals in 2016 to 2020. mSphere; 8. Epub ahead of print 22 June 2023. https://doi.org/10.1128/msphere.00098-23

[23] Fatumo SA, Adoga MP, Ojo OO, Oluwagbemi O, Adeoye T, et al. Computational Biology and Bioinformatics in Nigeria. PLoS Comput Biol; 10. Epub ahead of print 2014. https://doi.org/10.1371/journal.pcbi.1003516

[24] Nashiru O, Huynh C, Doumbia S, Kissinger JC, Isokpehi RD, et al. Building bioinformatics capacity in West Africa. Afr J Med Med Sci 2007; 36 Suppl: 15–18.

[25] Abrudan M, Matimba A, Nikolic D, Hughes D, Argimón S, et al. Train-the-Trainer as an Effective Approach to Building Global Networks of Experts in Genomic Surveillance of Antimicrobial Resistance (AMR). Clinical Infectious Diseases 2021; 73: S283–S289. https://doi.org/10.1093/cid/ciab770

[26] Joensen KG, Scheutz F, Lund O, Hasman H, Kaas RS, et al. Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic Escherichia coli. J Clin Microbiol 2014; 52: 1501–1510. https://doi.org/10.1128/JCM.03617-13

[27] Tetzschner AMM, Johnson JR, Johnston BD, Lund O, Scheutz F. In Silico Genotyping of Escherichia coli Isolates for Extraintestinal Virulence Genes by Use of Whole-Genome Sequencing Data. J Clin Microbiol; 58. Epub ahead of print 1 October 2020. https://doi.org/10.1128/JCM.01269-20

[28] Argimón S, Yeats CA, Goater RJ, Abudahab K, Taylor B, et al. A global resource for genomic predictions of antimicrobial resistance and surveillance of Salmonella Typhi at pathogenwatch. Nat Commun 2021; 12: 2020. 07. 03. 186692. https://doi.org/10.1038/s41467-021-23091-2

[29] Argimón S, David S, Underwood A, Abrudan M, Wheeler NE, et al. Rapid Genomic Characterization and Global Surveillance of Klebsiella Using Pathogenwatch. Clinical Infectious Diseases 2021; 73: S325–S335. https://doi.org/10.1093/cid/ciab784

[30] Hunt M, Mather AE, Sánchez-Busó L, Page AJ, Parkhill J, et al. ARIBA: Rapid antimicrobial resistance genotyping directly from sequencing reads. Microb Genom; 3. Epub ahead of print 1 October 2017. https://doi.org/10.1099/mgen.0.000131

[31] Zankari E, Allesøe R, Joensen KG, Cavaco LM, Lund O, et al. PointFinder: a novel web tool for WGS-based detection of antimicrobial resistance associated with chromosomal point mutations in bacterial pathogens. https://doi.org/10.1093/jac/dkx217

[32] Mendes I, Griffiths E, Manuele A, Fornika D, Tausch SH, et al. hAMRonization: Enhancing antimicrobial resistance prediction using the PHA4GE AMR detection specification and tooling. bioRxiv 2024; 2024. 03. 07. 583950. https://doi.org/10.1101/2024.03.07.583950

[33] Heiberger RM, Robbins NB. Design of Diverging Stacked Bar Charts for Likert Scales and Other Applications. J Stat Softw 2014; 57: 1–32.

[34] R Core Team. A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna (2022). https://www.r-project.org/ (2022).

[35] Argimón S, Abudahab K, Goater RJE, Fedosejev A, Bhai J, et al. Microreact: visualizing and sharing data for genomic epidemiology and phylogeography. Microb Genom 2016; 2: e000093. https://doi.org/10.1099/mgen.0.000093

[36] de Ridder J, Meysman P, Oluwagbemi O, Abeel T. Soft Skills: An Important Asset Acquired from Organizing Regional Student Group Activities. PLoS Comput Biol 2014; 10: e1003708. https://doi.org/10.1371/journal.pcbi.1003708

[37] Jongeneel CV, Achinike-Oduaran O, Adebiyi E, Adebiyi M, Adeyemi S, et al. Assessing computational genomics skills: Our experience in the H3ABioNet African bioinformatics network. PLoS Comput Biol 2017; 13: e1005419. https://doi.org/10.1371/journal.pcbi.1005419