A Department of Defense Laboratory Consortium Approach to Next Generation Sequencing and Bioinformatics Training for Infectious Disease Surveillance in Kenya

Irina Maljkovic Berry[1]*, Wiriya Rutvisuttinunt[1,2], Logan J. Voegtly[3,4], Karla Prieto[5,6], Simon Pollett[1], Regina Z. Cer[3,4], Jeffrey R. Kugelman[6], Kimberly A. Bishop-Lilly[3], Lindsay Morton[7], John Waitumbi[8] and Richard G. Jarman[1]

[1] Viral Diseases Branch, Walter Reed Army Institute of Research, Silver Spring, MD, United States, [2] Office of Genomics and Advanced Technologies National Institute of Allergy and Infectious Diseases, Bethesda, MD, United States, [3] Genomics & Bioinformatics Department, Biological Defense Research Directorate, Naval Medical Research Center-Frederick, Fort Detrick, MD, United States, [4] Leidos, Reston, VA, United States, [5] College of Public Health, University of Nebraska Medical Center, Omaha, NE, United States, [6] Center for Genomic Studies, United States Army Medical Research Institute for Infectious Diseases, Frederick, MD, United States, [7] Global Emerging Infections Surveillance, Armed Forces Health Surveillance Branch, Silver Spring, MD, United States, [8] Basic Science Laboratory, US Army Medical Research Directorate-Africa/Kenya Medical Research Institute, Kisumu, Kenya

Epidemics of emerging and re-emerging infectious diseases are a danger to civilian and military populations worldwide. Health security and mitigation of infectious disease threats is a priority of the United States Government and the Department of Defense (DoD). Next generation sequencing (NGS) and Bioinformatics (BI) enhances traditional biosurveillance by providing additional data to understand transmission, identify resistance and virulence factors, make predictions, and update risk assessments. As more and more laboratories adopt NGS and BI technologies they encounter challenges in building local capacity. In addition to choosing the right sequencing platform and approach, considerations must also be made for the complexity of bioinformatics analyses, data storage, as well as personnel and computational requirements. To address these needs, a comprehensive training program was developed covering wet lab and bioinformatics approaches to NGS. The program is meant to be modular and adaptive to meet both common and individualized needs of medical research and public health laboratories across the DoD. The training program was first deployed internationally to the Basic Science Laboratory of the US Army Medical Research Directorate-Africa in Kisumu, Kenya, which is an overseas Lab of the Walter Reed Army Institute of Research (WRAIR). A week-long workshop with intensive focus on targeted sequencing and the bioinformatics of genome assembly ($n = 24$ participants) was held. Post-workshop self-assessment (completed by 21 participants)

noted significant median gains in knowledge domains related to NGS targeted sequencing, bioinformatics for genome assembly, and sequence quality assessment. The participants also reported that the information on study design, sample preparation, sequencing quality control, data quality assessment, reporting, and basic and advanced bioinformatics analysis were the most useful information presented in the training. While longer-term evaluations are planned, the training resulted in significant short-term improvement of a laboratory's self-reported wet lab and bioinformatics capabilities. This framework can be used for future DoD laboratory development in the area of NGS and BI for infectious disease surveillance, ultimately enhancing this global DoD capability.

INTRODUCTION

Development of Next-Generation Sequencing (NGS), or High-Throughput Sequencing (HTS), has revolutionized life sciences, dramatically increasing the variety of questions that can be answered using genomic sequence data. With this continuously evolving and growing field, the need for adequate computational hardware resources, software, and expertise to analyze large and complex data is also increasing. The field of bioinformatics has thus experienced substantial growth and advancement in recent years, and the requirement for highly skilled and specialized personnel has surged.

Within the Department of Defense (DoD), NGS and bioinformatics are routinely used to answer many scientific and research questions that ultimately aid in protection of the armed forces, as well as the general population (Kijak et al., 2017; Colby et al., 2018; Ehrenberg et al., 2019; Waickman et al., 2019). Infectious diseases are one area where such research is of high importance. Like the general population, United States forces are vulnerable to many infections commonly occurring within the United States, such as influenza, coronavirus, adenovirus and antibiotic resistant bacterial infections including but not limited to infection by methicillin resistant *Staphylococcus aureus* (MRSA); pathogens that have the ability to negatively impact United States force readiness and mission goals (MacPherson et al., 1923; Beam et al., 1959; Earhart et al., 2001; Shanks and Hodge, 2011; Millar et al., 2017, 2019). In addition, global deployment of the United States forces also puts them at a higher risk for infections that occur more frequently outside the United States, such as Ebola, dengue, Zika, cholera, malaria, leishmaniasis, shigellosis, and many others (Riddle et al., 2011; Murray et al., 2015). The DoD Global Emerging Infections Surveillance (GEIS) program seeks to improve infectious disease surveillance, prevention, and response capability to better protect the health of the military force. Utilizing a global network of partner DoD medical research and public health laboratories, GEIS funds surveillance activities in over 70 countries to inform force health protection through timely and actionable infectious disease surveillance information (Chakhunashvili et al., 2017; Chang et al., 2018; Coleman et al., 2018; Koka et al., 2018; Anyamba et al., 2019; Guerra et al., 2019; Juma et al., 2019; Rivers et al., 2019; Rocha et al., 2019;

Sugiharto et al., 2019). Unsurprisingly, development of NGS and bioinformatics methods for infectious disease surveillance and control has enabled a rapid expansion of GEIS partner studies that utilize pathogen genomic information (Frey et al., 2016; Maljkovic Berry et al., 2016, 2019a; Lee et al., 2017; Mullins et al., 2017; Salje et al., 2017; Cowell et al., 2018; LaBreck et al., 2018; Srijan et al., 2018; Grubaugh et al., 2019; Kim et al., 2019; Mbala-Kingebeni et al., 2019; Millar et al., 2019; Pollett et al., 2019; Wiley et al., 2019). However, NGS and bioinformatics can generally be technically challenging, as it requires specific knowledge of complex wet lab and bioinformatics processes (Maljkovic Berry et al., 2019b). Therefore, and in spite of great interest in this technology, only a few partner laboratories have been adequately equipped to utilize these approaches to their full potential.

In 2017, GEIS created a Consortium to address the increasing needs and challenges associated with NGS and bioinformatics at DoD medical research and public health laboratories. The vision of the Consortium is to rapidly detect and characterize known, emerging, and novel infectious disease agents through establishment of a harmonized DoD laboratory NGS and bioinformatics capability to inform force health protection decision making. The Consortium today represents a network of DoD laboratories that use NGS and bioinformatics for infectious disease surveillance. A baseline assessment and initial training effort was led by GEIS and three DoD core sequencing and bioinformatics laboratories: WRAIR-VDB (Walter Reed Army Institute of Research-Viral Diseases Branch), NMRC-BDRD (Naval Medical Research Center-Biological Defense Research Directorate), and USAMRIID-CGS (United States Army Medical Research of Infectious Diseases-Center for Genome Science). The Consortium performed an assessment of the GEIS DoD laboratory partners with access to Illumina MiSeq or other NGS instrument(s), in order to evaluate existing laboratory capabilities in NGS and bioinformatics, and to map gaps and needs in laboratory utilization of these tools to meet their mission goals of infectious disease surveillance. Limited access to experienced and knowledgeable NGS and bioinformatics personnel was one of the main gaps, making basic and advanced bioinformatics analyses a common challenge across the network. Another challenge was the restrictive and limited informatics infrastructure, especially in some of the participating laboratories

located in low-and-middle income countries (LMICs). However, the challenge of finding personnel with sufficient training in NGS and bioinformatics was not only observed in laboratories located in LMICs, it was also apparent in domestic laboratories, thus highlighting the need to develop a structured NGS and bioinformatics training for the specific needs of DoD biosurveillance programs. Such training would have to be standardized across the Consortium network, as well as made agile enough to meet different levels of needs and computational resources of the participating DoD laboratories. Using the baseline information from the assessment, desired sequencing capabilities for DoD research and public health laboratories were divided into three tiers (**Figure 1**). Here we present the

deployment of NGS and bioinformatics training with our partner laboratory in Kenya, United States Army Medical Research Directorate – Africa (MRD-A). Future iterations of similar trainings and assessments will be used to further strengthen global infectious surveillance for DoD utilizing genomics and bioinformatics.

MATERIALS AND EQUIPMENT

Samples used for the NGS hands-on training included dengue virus 2 (DENV-2) and chikungunya (CHIKV) and were provided on-site. Controls for library preparation, MiSeq
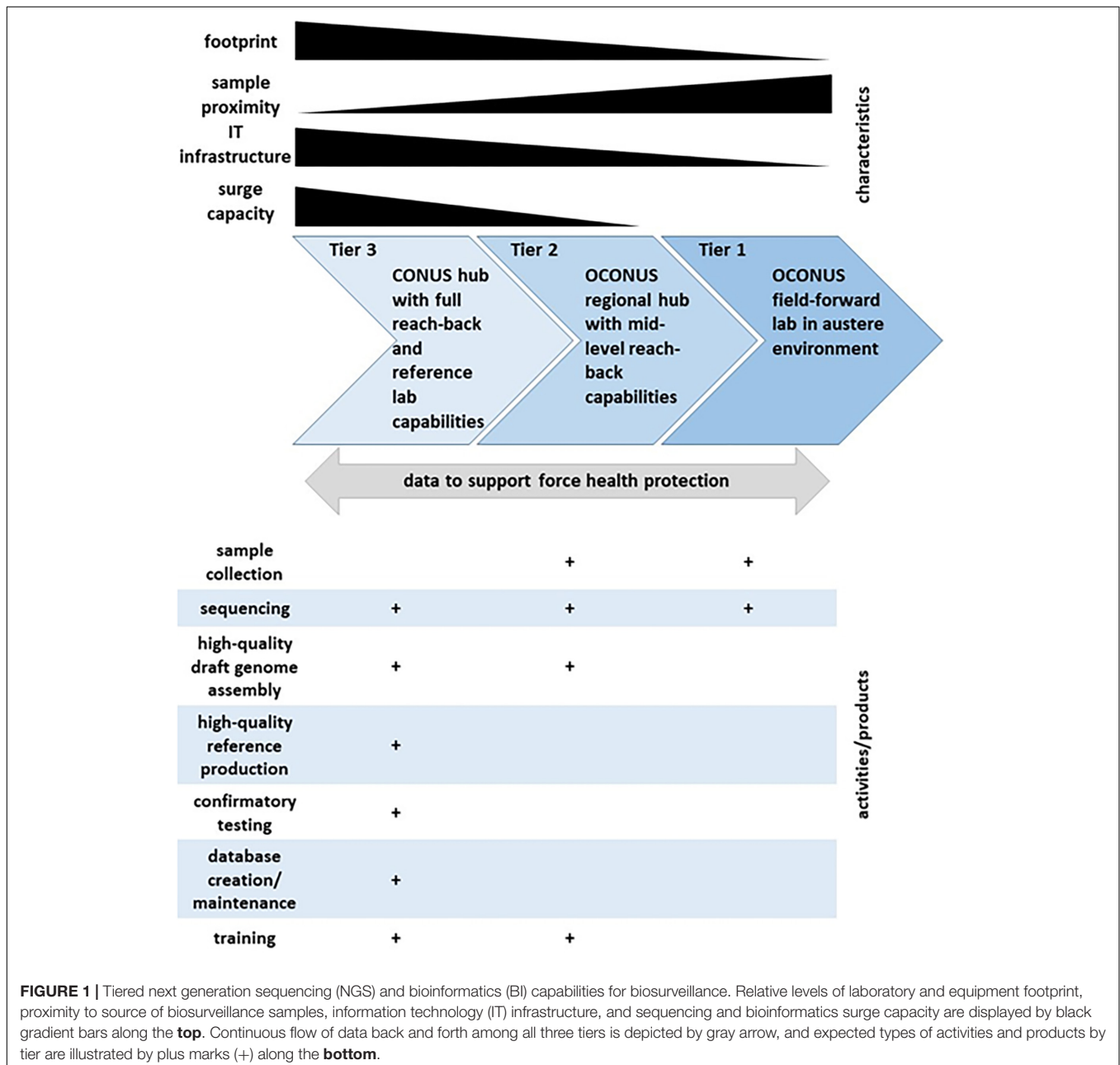
FIGURE 1 | Tiered next generation sequencing (NGS) and bioinformatics (BI) capabilities for biosurveillance. Relative levels of laboratory and equipment footprint, proximity to source of biosurveillance samples, information technology (IT) infrastructure, and sequencing and bioinformatics surge capacity are displayed by black gradient bars along the **top**. Continuous flow of data back and forth among all three tiers is depicted by gray arrow, and expected types of activities and products by tier are illustrated by plus marks (+) along the **bottom**.

sequencing and TapeStation for both DENV-2 and CHIKV were validated and prepared at VDB-WRAIR in the months prior to the planned NGS&BI training in Kenya. Prior to shipment of controls to Kenya, the control concentrations were measured and documented and the information was sent to MRD-A. Coordination of the reagent and control shipment from VDB-WRAIR to Kisumu, Kenya started a month prior to the training. Four Linux laptops and two Linux servers were prepared for hands-on bioinformatics training. A list of software was prepared by the Consortium and sent out to MRD-A Lab for installation onto the training computers. The software list included ngs_mapper, IGV, Geneious, MEGA7, EDGE (servers only) (Robinson et al., 2011; Kumar et al., 2016; Viral Diseases Branch WRAIR, 2016; Philipson et al., 2017). Three weeks prior to the training, a hands-on genome assembly training dataset was designed, consisting of dengue, chikungunya, and influenza raw fastq data, as well as hands-on performance instructions. The whole dataset was tested at VDB-WRAIR prior to training and saved onto the training computers.

METHODS

Day 1

Lectures and theory included: History of sequencing, overview of NGS, library preparation, quantification, validation and pooling. In detail: (i) List of library preparation kits used by core DoD for different projects and specimens were highlighted; (ii) Several topics on types of kits for viruses, bacteria and parasite work were heavily discussed throughout the lecture; (iii) Specific library preparation kits were highlighted including TruSeq, QIASeq Fx, Kappa, NexteraXT, RNA Access and DNAFlex; (iv) AmpureXP Beads clean up after PCR reactions and library preparation was emphasized as preferred method; (v) Different library validations, including qPCR, Qubit and TapeStation were highlighted as essentials for quality control (QC); (vi) Library pooling based on TapeStation and Qubit were introduced; (vii) Two exercises of how to calculate amount of each library for pooling were conducted. Preparations were made for the upcoming bioinformatics training.

Day 2

Hands-on training for NGS wet lab was performed with 24 participants. The participants were separated into two groups based on their NGS background and interests for hands-on performance. Group 1 prepared the NexteraXT library from the amplicons and assessed amplicons using both Qubit and TapeStation prior to NexteraXT library preparation. The NexteraXT libraries were validated using both Qubit and TapeStation. Group 2 validated the pooling based on the controls from the shipment and prepared sample sheets, the MiSeq instrument and PhiX controls. The libraries were loaded onto the Miseq. Bioinformatics training dataset was prepared on each computer. Server performance was tested for running the pipelines and tools needed for the training, and the training dataset analyses

were executed to test functionality prior to the hands-on bioinformatics training.

Day 3

Hands-on wet lab activities from Day 2 were summarized and any questions and concerns were addressed. Lectures on laboratory project experimental design (to include bioinformatics), bioinformatics data cleaning and pre-processing, and genome assembly through reference mapping were performed, as well as exercises in experimental design and genome consensus calling. For hands-on bioinformatics training, the 24 participants were divided into six different groups, each group utilizing one training computer or server. Ngs_mapper was used as the example of a reference mapping pipeline. The first training was performed on the DENV fastq dataset, including training on usage of different stages of the pipeline, setting a desired reference genome and running the pipeline. After ngs-mapper jobs were completed, interpretation of the output, how to utilize data quality scores and depth of coverage, how to assess the performance of the sequencing and the genome assembly were performed. Manual QC and genome curation were performed. The second training dataset consisted of CHIKV fastqs and was used for training on multiple reference usage and reference selection, in addition to repeating the above steps for dataset one.

Day 4

Bioinformatics hands-on training was continued by evaluation of the CHIKV runs for reference genome selection. Based on the best reference choice, the reference mapping run was repeated. The repetition was incorporated on purpose to ensure better knowledge retention. Following reference mapping, the output of CHIKV assembly was evaluated and its genome curated. The data that were used for this training were purposefully chosen to be of lower quality, so that different challenges of genome assembly curation were highlighted, as well as the importance of QC and what consequences a lack of QC might result in. The last reference mapping analysis was performed on CHIKV data but now the participants learned how to change different pipeline thresholds, picking their own requirements for minimum base quality, consensus type output and the like. In addition, lectures were conducted covering theory of *de novo* genome assembly, assembly of bacterial genomes, and troubleshooting and maintenance of the MiSeq platform.

Day 5

A summary of wet lab activities and library pooling to obtain optimal cluster density was presented. An exercise aimed at the evaluation of several MiSeq runs was performed. Management of sequencing libraries and data, and prevention of chimeric sequence data generation and mislabeling were discussed. Bioinformatics training on the influenza dataset was performed separately since influenza virus has a segmented genome and bioinformatically, full genome assembly is slightly more complicated. How to recognize presence of influenza reassortment was covered. A workshop survey was

distributed (**Supplementary Material**) and the workshop was concluded.

RESULTS

NGS and Bioinformatics Training Modules

A comprehensive training curriculum was constructed that consisted of standardized wet lab and bioinformatics theory modules (**Figure 2**) as well as hands-on training. The modules could be independently compiled into a set of theoretical lectures

that could be adjusted for the existing laboratory tiers and specific knowledge gaps. As they were designed to meet the particular DoD surveillance needs, the modules were divided into two main wet lab sequencing and two main bioinformatics analyses approaches. The wet lab lectures could thus be adjusted to cover: (i) the theory of targeted sequencing, which is mainly used in response to epidemics and outbreaks of known pathogens; and (ii) the theory of metagenomics, which is usually used for pathogen discovery and identification. The bioinformatics lectures focused on: (i) the genome assembly and curation analyses, an essential part of outbreak genomic surveillance; and (ii) the bioinformatics of pathogen discovery, usually the most
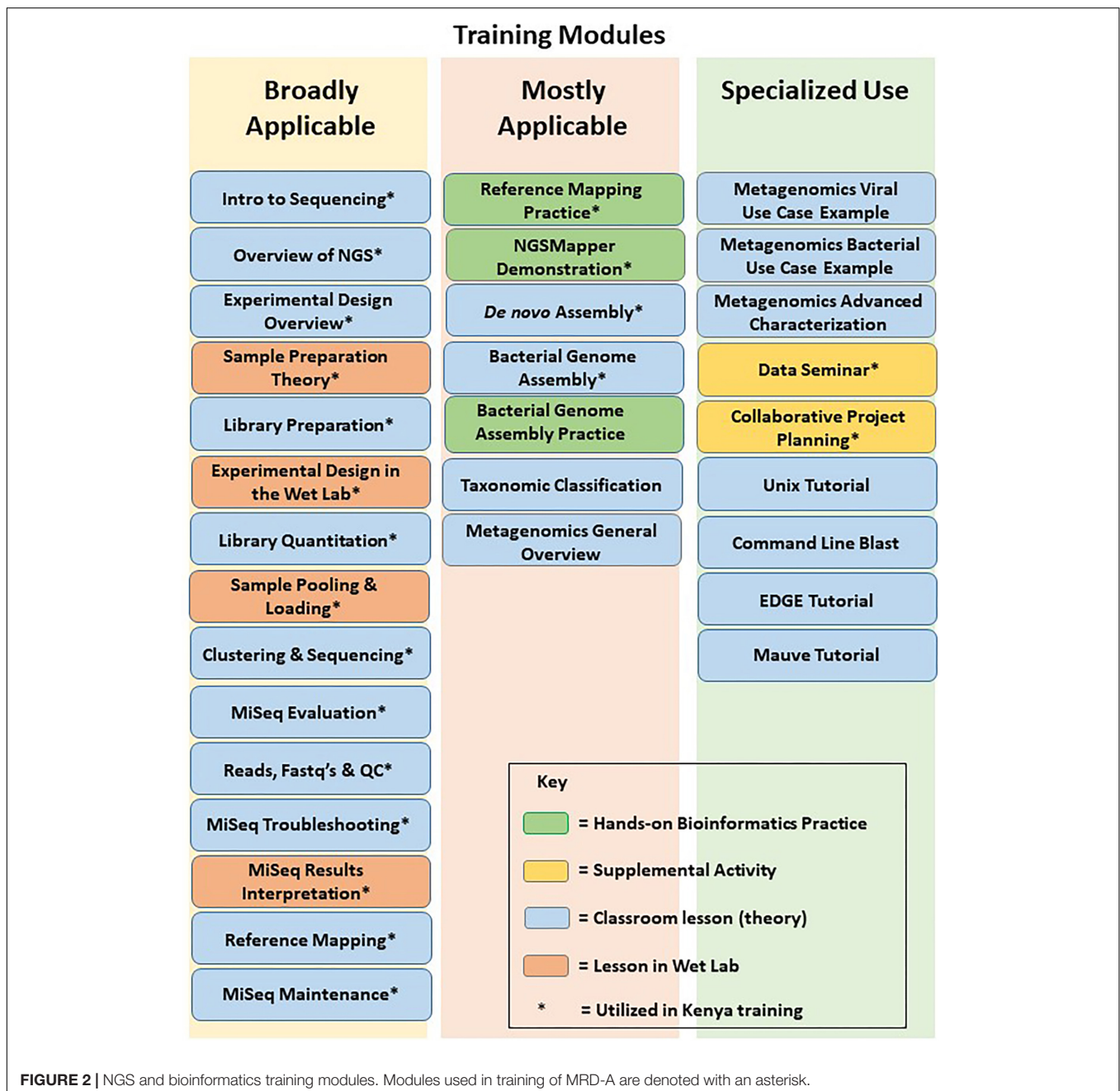
FIGURE 2 | NGS and bioinformatics training modules. Modules used in training of MRD-A are denoted with an asterisk.

challenging aspect of basic sequencing-based biosurveillance. In addition to these, modules covering other parts of NGS and bioinformatics were included, such as theory of experimental design, troubleshooting, and equipment maintenance. The theory modules were complemented with development of corresponding hands-on wet lab and bioinformatics training of the above approaches.

NGS and Bioinformatics Training Deployment

Based on the results of the initial laboratory assessment, training was recommended for the GEIS partner US Army Medical Research Directorate – Africa (MRD-A) laboratories in Kenya. For MRD-A's initial needs, which mainly cover sequencing and analyses of known pathogen outbreaks and epidemics in the region, a 1 week on-site workshop was constructed where the wet lab targeted sequencing was covered in both lectures (specific assembled modules) and hands-on practice, followed by bioinformatics theory (specific assembled modules) and hands-on practice of pathogen genome assembly and

curation (**Figure 2**). This approach was specifically designed based on the needs and gaps that were highlighted during the initial assessment of MRD-A capabilities. Participating in the training were representatives from various MRD-A and Kenya Medical Research Institute (KEMRI) laboratory divisions in Kenya: Basic Science, Viral Hemorrhagic Fevers, Entomology, Flu Lab, Antimicrobial Resistance, Sexually Transmitted Infections, Microbiology Hub-Kericho, Influenza, and KEMRI-Centers for Disease Control divisions (**Figure 3**). There was a total of 24 workshop participants.

We undertook a rapid evaluation of participants' self-reported baseline and post-workshop knowledge across ten skill domains related to genomic sequencing (**Supplementary Material**). We also determined individual-level gains in self-reported knowledge after completing the workshop. This was measured with a single hard-copy questionnaire administered after the workshop. This survey asked the participants to self-rate their knowledge in each skill domain on a customized scale of 1–10 (1 = "no prior knowledge", 10 = "high level of experience") before and after the workshop. Median baseline and post-workshop scores are presented in **Table 1**. While interpretation of
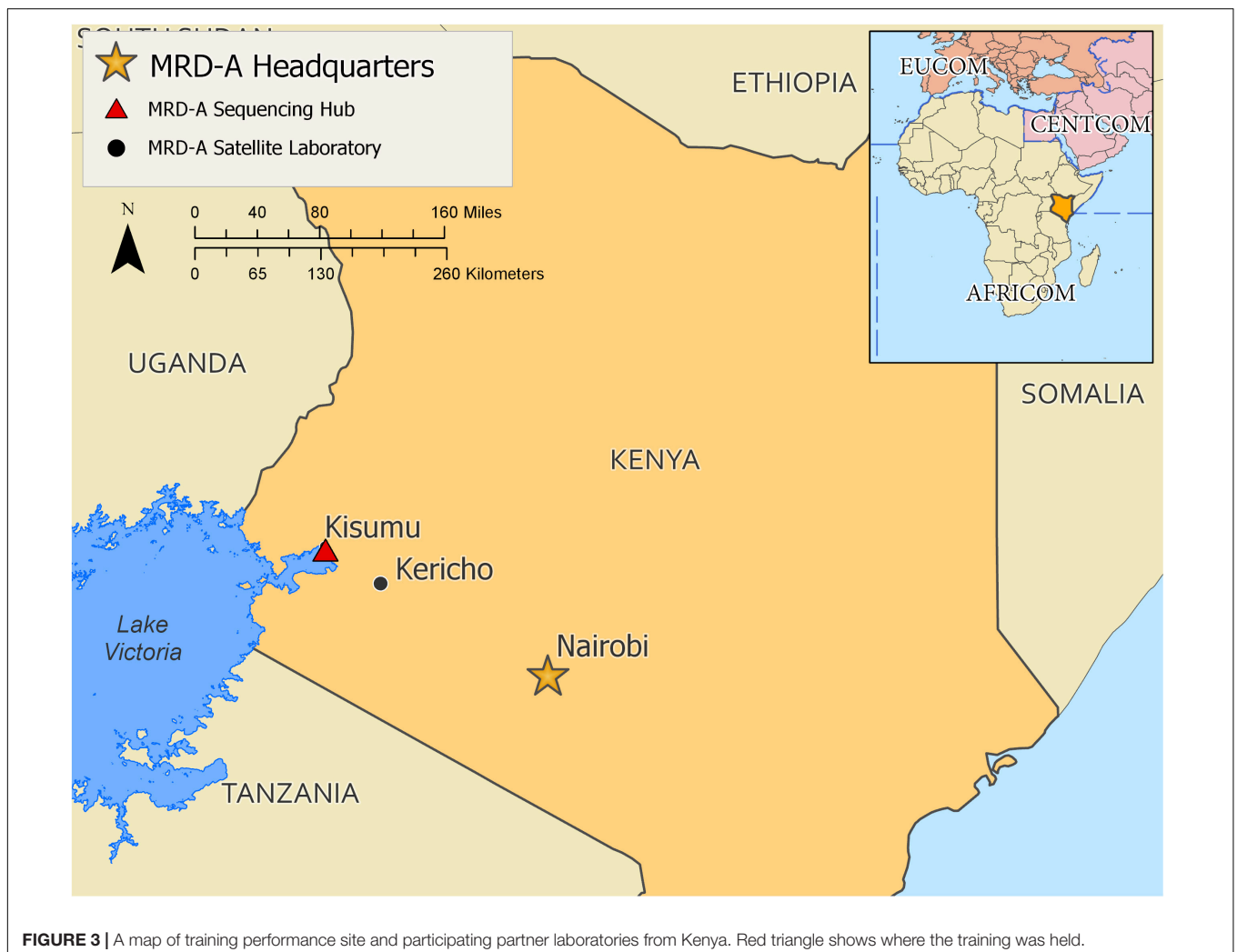
FIGURE 3 | A map of training performance site and participating partner laboratories from Kenya. Red triangle shows where the training was held.

TABLE 1 | Self-reported knowledge across skill domains of genomic sequencing (*n* = 21 respondents).

Knowledge domain	Pre-workshop score[a]	Post-workshop score[a]	Post-workshop gains	
	Median (IQR)	Median (IQR)	Median (IQR)	*p*-value[b]
NGS technology	4 (3,5)	7 (6,8)	3 (2,3)	<0.001
Illumina MiSeq sequencing chemistry	4 (2,5)	7 (6,8)	3 (2,4)	<0.001
NGS library preparation	5 (2,6)	8 (7,9)	3 (2,4)	<0.001
NGS library validation	2 (1,5)	8 (6,8)	4 (3,4)	<0.001
MiSeq run validation	2 (1,3)	6 (4,8)	3 (2,4)	<0.001
Experimental design for bioinformatics analysis	2 (1,4)	6 (5,8)	3 (2,4)	<0.001
FASTQ data cleaning and pre-processing	2 (1,5)	6 (5,8)	4 (2,4)	<0.001
Reference mapping	3 (1,5)	7 (6,9)	4 (1,5)	<0.001
Linux OS use and command line	2 (1,5)	5 (3,7)	1 (0,3)	<0.001
Consensus sequence calling and manual curation	2 (1,3)	6 (5,8)	4 (2,5)	<0.001

NGS, next generation sequencing; OS, operating system. [a]*Maximum possible score is 10.* [b]*Derived from Wilcoxon signed rank test.*

these metrics is limited due to the subjectivity of the self-reported knowledge measurements, particularly when measured at a single point in time, the IQR and range around the median reported knowledge scores did suggest that this sample of participants had varying expertise across each of these skill domains. Pre-training baseline scores suggested that the participants had, in particular, less self-reported expertise in NGS library validation, Illumina MiSeq run validation, experimental design for bioinformatics analysis, and FASTQ data cleaning and pre-processing.

There were substantial gains in self-reported knowledge across all skill domains (**Table 1**), with the notable exception of Linux OS and command line skills, suggesting that this is a particular area of residual training need. Indeed, Linux OS and command line skill had the lowest post-workshop self-reported knowledge scores. A module was later developed specifically to fill this gap (**Figure 2**). The questionnaire also measured the participants' perceptions on the most "useful" information learned during the NGS library and bioinformatics components of the workshop. This was measured by free-text open ended questions (**Table 2**).

The participants were also asked in which topics they felt they would like more training and experience (**Table 3**) and how to improve future iterations of this workshop (**Table 4**). The participant's responses all highlight the complexity and the diversity of considerations within NGS and bioinformatics. The many topics that can be covered and trained upon for the fields of infectious disease surveillance and control alone, and the associated time that it would take to train and educate the workforce, would indicate a large gap in the currently existing education programs.

TABLE 2 | Information reported by participants to be the most useful (*n* = 21 respondents)[a].

	n	%[b]
NGS library preparation	12	60
NGS library validation	7	35
Sample pooling	6	30
Tapestation use	5	25
Library normalization	5	25
Qubit use	2	10
QC for sequence reads	1	5
Experimental design considerations	1	5
Sequencing platform overview	1	5
MiSeq runs (hands-on experience)	1	5
Sample pre-processing	1	5
Nextera-XT protocol	1	5
MiSeq run troubleshooting	1	5
Genome assembly (reference mapping and curation)	15	71
Sequence read QC	8	38
De novo sequencing	4	19
Experimental considerations	3	14
Software (including NGS_mapper and IGV)	3	14
Mapping bacterial sequences	1	5
Output analysis	1	5

[a]*Derived from open questions: "What information was most useful to you that this NGS library provided?" and "What information was most useful to you that this bioinformatic workshop provided?".* [b]*Some participants indicated > 1 item of information in response. QC, quality control.*

DISCUSSION

The rapid growth and utility of NGS and bioinformatics for research and biosurveillance has resulted in the emergence of DoD requirements for implementation of sequencing and computational technologies, as well as access to highly trained and knowledgeable personnel in the fields of NGS and bioinformatics. Specifically the latter point remains one of the major challenges across the DoD, and even though bioinformatics programs have more recently gained larger momentum in academia, lack of workforce with early-on and/or specialized bioinformatics training is still palpable in the government settings, particularly in government labs outside the continental United States. Therefore, NGS and bioinformatics training programs for infectious disease surveillance have recently been developed by many government agencies or non-governmental organizations. Within the United States Government, Canada, and the European Union, there is movement towards training and coordinated promotion of standardized quality assurance and quality control practices for pathogen genome sequencing using NGS technologies (e.g., Illumina) (Cui et al., 2015; Gargis et al., 2016; Nadon et al., 2017). Some recent examples include the GenomeTrakr program at the Food and Drug Administration, Next Generation PulseNet at Centers for Disease Control, and the Global Microbial Identifier for food-borne pathogen surveillance (Moran-Gilad et al., 2015; Timme et al., 2018; Ribot et al., 2019). More recently, the SARS-CoV-2 Sequencing for Public Health Emergency Response,

TABLE 3 | Suggested topics for more training/experience, as reported by participants (n = 21 respondents)[a].

	n	%[b]
Phylogenetics and other advanced bioinformatics analysis	8	38
Metagenomics for pathogen discovery	4	19
De novo assembly	4	19
Linux OS/cluster	4	19
MiSeq loading and run evaluation	3	14
Bacterial genomics	3	14
Pipeline development (including open source bioinformatic tools)	3	14
Sequence assembly	2	10
Read QC	2	10
Library prep	1	5
16s and 18s molecular analysis	1	5
Sample pre-processing	1	5
Bioinformatic experimental design	1	5
SNP detection and variant calling	1	5
Sample sheet prep	1	5
Reference mapping	1	5
Plasmid sequencing	1	5
Recombination detection	1	5
Comparative genomics	1	5
Outbreak investigations	1	5

[a]Derived from open question: "What topic would you like more training/experience in (if any)?". [b]Some participants indicated more than one line item of information. QC, quality control.

TABLE 4 | Participants' suggestions for workshop improvements (n = 21 respondents)[a].

	n	%[b]
More time (longer workshop >1 week)	4	19
More hands on training/practical sessions (less theory/slides)	4	19
More time on bioinformatic data interpretation and analysis	3	14
Increased frequency of workshop with follow-up training	2	10
More wet lab time	2	10
More laptops	1	5
Split into beginner and advanced classes	1	5

[a]Derived from open question: "Suggestions for workshop improvement?". [b]Some participants had >1 suggestion.

Epidemiology, and Surveillance (SPHERES) national genomics consortium was set up by the Centers for Disease Control, to coordinate SARS-CoV-2 sequencing across the United States (Centers for Disease Control and Prevention, 2019). Within the DoD, the training designed and implemented by the GEIS Consortium aims to develop lasting and sustainable capabilities for pathogen genomic sequencing and bioinformatics at DoD medical research and public health laboratories in overseas locations.

Our experience in deploying a comprehensive yet customizable classroom and hands-on training in NGS and bioinformatics in Kenya was overall successful (see caveats of assessment below) and is a potential model for future training programs in similar environments. This training program consisted of foundational material in

sequencing theory and experimental design which formed a basis for more applied modules in targeted sequencing and metagenomics. Additionally, hands-on NGS wet lab and bioinformatics modules were further tailored to meet the needs of the laboratory participants using information obtained from a baseline landscape assessment. This training shows that a highly modular and deployable set of NGS and bioinformatics workshop components can be used within the DoD network of medical research and public health laboratories to improve sequencing wet lab capability, and analysis and interpretation of pathogen genomic data gathered using NGS and bioinformatics.

Embedded within this training workshop was a post-self-assessment questionnaire to gauge immediate improvements in knowledge gained from the workshop materials. It is important to note that this questionnaire has several limitations including a small sample size, the immediate nature of the assessment tool which does not allow one to measure long-term benefits, and the fact that the assessment was only delivered through written evaluation and self-report. Further, more objective measurements of knowledge and skill gains after workshops may not directly translate into effective implementation and retention of these skills. The latter requires medium and longer term evaluations in an implementation science framework (Nilsen, 2015). However, these data do suggest that the participants have perceived that this workshop offered productive training which has led to substantial gains in knowledge. In similar bioinformatics trainings in LMICs, technological limitations were identified as an impediment to knowledge acquisition and long-term improvements in bioinformatics capability (Pollett et al., 2016). This training attempted to overcome these barriers by (a) providing training laptops, (b) providing recommendations for IT upgrades, bioinformatics software, and computer networking, and (c) upgrading local IT equipment for bioinformatics during the workshop.

Following this workshop a mechanism to facilitate reach back support with embedded long-term training and mentorship has been instituted to overcome challenges associated with long-term sustainability of a sequencing capability at MRD-A. Included in this 5-year NGS and bioinformatics implementation plan for MRD-A are: (i) continuous contact and support by the core DoD sequencing laboratories, (ii) repetition of training with focus on real data and troubleshooting, (iii) additional hands-on training in other wet lab and bioinformatics approaches to achieve capability diversification, (iv) development of local computational infrastructure for bioinformatics, and (v) regular assessments of wet lab and bioinformatics knowledge retention. Laboratory-level assessments of proficiency and skill retention 1–2 years post-training have included external review of raw sequence data and consensus genomes generated from GEIS funded surveillance projects. We also anticipate deploying periodic blinded panel of samples or data files for follow-up assessments of knowledge retention and capability development. At the end of this period, the goal is to achieve a high quality diversified portfolio of NGS and bioinformatics capabilities at the site, which then may serve as a central DoD hub for sequencing

and advanced characterization of Force Health Protection (FHP) relevant pathogens in Africa.

The current COVID-19 pandemic has further highlighted the importance of access to the NGS and bioinformatics in laboratories throughout the world. This makes the need of workshops such as ours even greater. However, the pandemic has also made travel and in-person learning a challenge, and therefore, GEIS is planning on development of virtual versions of the workshops to continue development of this important DoD-wide capability. In addition, Oxford Nanopore's MinION platform has increasingly been used in pathogen outbreak studies for real-time in-field analyses throughout the world, including analyses of SARS-CoV-2 (Quick et al., 2016; Faria et al., 2018; Moore et al., 2020). Although training in the wet-lab and bioinformatics of this approach was not included in the workshop in Kenya to maintain simplicity and focus, the plan is to apply the modular approach for development and incorporation of a general DoD MinION-focused training for the GEIS partner laboratories. Currently, GEIS has established a separate MinION working group, and has been working in providing basic training in this technology to a subset of partner laboratories.

More broadly, the Consortium goal is the establishment of basic proficiencies and adopted norms in quality assurance and quality control in targeted (hybridization- or amplicon-based) and metagenomic sequencing for viral and bacterial pathogens leading to more reliable results which will ultimately improve DoD public health surveillance and response. An additional objective is the development and maintenance of advanced genomics and bioinformatics capabilities in the United States and priority overseas locations, in order to enhance global health surveillance and facilitate faster response to infectious disease outbreaks. Development of these capabilities with GEIS DoD laboratory partners will require sustained commitment and global coordination. The end results will be the ability to reliably and rapidly sequence, identify, and characterize pathogens of public health importance in order to improve biosurveillance efforts and inform FHP measures throughout the world.

DATA AVAILABILITY STATEMENT

All datasets presented in this study are included in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

IM, WR, KB-L, LM, and RJ designed the training modules and the workshop modules. IM, WR, LV, LM, KP, RJ, JW, SP, and RC prepared and instructed the workshop. IM, WR, LV, KP, SP, RC, JK, KB-L, LM, JW, and RJ performed workshop post-assessment and wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.577563/full#supplementary-material

REFERENCES

Anyamba, A., Chretien, J. P., Britch, S. C., Soebiyanto, R. P., Small, J. L., Jepsen, R., et al. (2019). Global disease outbreaks associated with the 2015-2016 El Nino Event. *Sci. Rep.* 9:1930. doi: 10.1038/s41598-018-38034-z

Beam W. E. Jr., Grayston, J. T., and Watten, R. H. (1959). Second Asian influenza epidemics occurring in vaccinated men aboard U.S. Navy vessels. *J. Infect. Dis.* 105, 38–44. doi: 10.1093/infdis/105.1.38

Centers for Disease Control and Prevention (2019). *SARS-CoV-2 Sequencing for Public Health Emergency Response, Epidemiology, and Surveillance 2020.* Available online at: https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/spheres.html (accessed August 6, 2020).

Chakhunashvili, G., Wagner, A. L., Machablishvili, A., Karseladze, I., Tarkhan-Mouravi, O., Zakhashvili, K., et al. (2017). Implementation of a sentinel surveillance system for influenza-like illness (ILI) and severe acute respiratory infection (SARI) in the country of Georgia, 2015-2016. *Int. J. Infect. Dis.* 65, 98–100. doi: 10.1016/j.ijid.2017.09.028

Chang, K. S., Kim, G. H., Ha, Y. R., Jeong, E. K., Kim, H. C., Klein, T. A., et al. (2018). Monitoring and control of *Aedes albopictus*, a vector of Zika Virus, near

residences of imported Zika Virus patients during 2016 in South Korea. *Am. J. Trop. Med. Hyg.* 98, 166–172. doi: 10.4269/ajtmh.17-0587

Colby, D. J., Trautmann, L., Pinyakorn, S., Leyre, L., Pagliuzza, A., Kroon, E., et al. (2018). Rapid HIV RNA rebound after antiretroviral treatment interruption in persons durably suppressed in Fiebig I acute HIV infection. *Nat. Med.* 24, 923–926. doi: 10.1038/s41591-018-0026-6

Coleman, R., Eick-Cost, A. A., Hawksworth, A. W., Hu, Z., Lynch, L., Myers, C. A., et al. (2018). Department of defense end-of-season influenza vaccine effectiveness estimates for the 2017-2018 season. *MSMR* 25, 16–20.

Cowell, A. N., Valdivia, H. O., Bishop, D. K., and Winzeler, E. A. (2018). Exploration of *Plasmodium vivax* transmission dynamics and recurrent infections in the Peruvian Amazon using whole genome sequencing. *Genome Med.* 10:52. doi: 10.1186/s13073-018-0563-0

Cui, H. H., Erkkila, T., Chain, P. S., and Vuyisich, M. (2015). Building international genomics collaboration for global health security. *Front. Public Health* 3:264. doi: 10.3389/fpubh.2015.00264

Earhart, K. C., Beadle, C., Miller, L. K., Pruss, M. W., Gray, G. C., Ledbetter, E. K., et al. (2001). Outbreak of influenza in highly vaccinated crew of U.S. Navy ship. *Emerg. Infect. Dis.* 7, 463–465. doi: 10.3201/eid0703.017320

Ehrenberg, P. K., Shangguan, S., Issac, B., Alter, G., Geretz, A., Izumi, T., et al. (2019). A vaccine-induced gene expression signature correlates with protection against SIV and HIV in multiple trials. *Sci. Transl. Med.* 11:507. doi: 10.1126/scitranslmed.aaw4236

Faria, N. R., Kraemer, M. U. G., Hill, S. C., Goes de Jesus, J., Aguiar, R. S., Iani, F. C. M., et al. (2018). Genomic and epidemiological monitoring of yellow fever virus transmission potential. *Science* 361, 894–899. doi: 10.1126/science.aat7115

Frey, K. G., Biser, T., Hamilton, T., Santos, C. J., Pimentel, G., Mokashi, V. P., et al. (2016). Bioinformatic characterization of mosquito Viromes within the Eastern United States and puerto rico: discovery of novel viruses. *Evol. Bioinform.* 12(Suppl 2), 1–12. doi: 10.4137/EBO.S38518

Gargis, A. S., Kalman, L., and Lubin, I. M. (2016). Assuring the quality of next-generation sequencing in clinical microbiology and public health laboratories. *J. Clin. Microbiol.* 54, 2857–2865. doi: 10.1128/jcm.00949-16

Grubaugh, N. D., Saraf, S., Gangavarapu, K., Watts, A., Tan, A. L., Oidtman, R. J., et al. (2019). Travel surveillance and genomics uncover a hidden Zika outbreak during the waning epidemic. *Cell* 178, 1057–1071.e11. doi: 10.1016/j.cell.2019.07.018

Guerra, R. I., Ore, M., Valdivia, H. O., Bishop, D. K., Ramos, M., Mores, C. N., et al. (2019). A cluster of the first reported *Plasmodium ovale* spp. infections in Peru occuring among returning UN peace-keepers, a review of epidemiology, prevention and diagnostic challenges in nonendemic regions. *Malar J.* 18:176. doi: 10.1186/s12936-019-2809-8

Juma, D. W., Muiruri, P., Yuhas, K., John-Stewart, G., Ottichilo, R., Waitumbi, J., et al. (2019). The prevalence and antifolate drug resistance profiles of *Plasmodium falciparum* in study participants randomized to discontinue or continue cotrimoxazole prophylaxis. *PLoS Negl. Trop. Dis.* 13:e0007223. doi: 10.1371/journal.pntd.0007223

Kijak, G. H., Sanders-Buell, E., Chenine, A. L., Eller, M. A., Goonetilleke, N., Thomas, R., et al. (2017). Rare HIV-1 transmitted/founder lineages identified by deep viral sequencing contribute to rapid shifts in dominant quasispecies during acute and early infection. *PLoS Pathog.* 13:e1006510. doi: 10.1371/journal.ppat.1006510

Kim, W. K., No, J. S., Lee, D., Jung, J., Park, H., Yi, Y., et al. (2019). Active targeted surveillance to identify sites of emergence of hantavirus. *Clin. Infect. Dis.* 70, 464–473. doi: 10.1093/cid/ciz234

Koka, H., Sang, R., Kutima, H. L., and Musila, L. (2018). *Coxiella burnetii* detected in tick samples from pastoral communities in Kenya. *Biomed. Res. Int.* 2018:8158102. doi: 10.1155/2018/8158102

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054

LaBreck, P. T., Rice, G. K., Paskey, A. C., Elassal, E. M., Cer, R. Z., Law, N. N., et al. (2018). Conjugative transfer of a novel staphylococcal plasmid encoding the biocide resistance gene, qacA. *Front. Microbiol.* 9:2664. doi: 10.3389/fmicb.2018.02664

Lee, S. H., Kim, W. K., No, J. S., Kim, J. A., Kim, J. I., Gu, S. H., et al. (2017). Dynamic circulation and genetic exchange of a shrew-borne hantavirus, Imjin virus, in the republic of Korea. *Sci. Rep.* 7:44369. doi: 10.1038/srep44369

MacPherson, W., Herringham, W., Elliott, T., and Balfour, A. (1923). *History of the Great War Based On Official Documents: Medical Services Diseases of the War.* London: HMSO.

Maljkovic Berry, I., Eyase, F., Pollett, S., Konongoi, S. L., Joyce, M. G., Figueroa, K., et al. (2019a). Global outbreaks and origins of a Chikungunya Virus variant carrying mutations which may increase fitness for *Aedes aegypti*: revelations from the 2016 Mandera, Kenya Outbreak. *Am. J. Trop. Med. Hyg.* 100, 1249–1257. doi: 10.4269/ajtmh.18-0980

Maljkovic Berry, I., Melendrez, M. C., Bishop-Lilly, K. A., Rutvisuttinunt, W., Pollett, S., Talundzic, E., et al. (2019b). Next generation sequencing and bioinformatics methodologies for infectious disease research and public health: approaches, applications, and considerations for development of laboratory capacity. *J. Infect. Dis.* 221(Suppl 3), S292–S307. doi: 10.1093/infdis/jiz286

Maljkovic Berry, I., Melendrez, M. C., Li, T., Hawksworth, A. W., Brice, G. T., Blair, P. J., et al. (2016). Frequency of influenza H3N2 intra-subtype reassortment: attributes and implications of reassortant spread. *BMC Biol.* 14:117. doi: 10.1186/s12915-016-0337-3

Mbala-Kingebeni, P., Aziza, A., Di Paola, N., Wiley, M. R., Makiala-Mandanda, S., Caviness, K., et al. (2019). Medical countermeasures during the 2018 Ebola virus disease outbreak in the North Kivu and Ituri provinces of the democratic republic of the Congo: a rapid genomic assessment. *Lancet Infect. Dis.* 19, 648–657. doi: 10.1016/S1473-3099(19)30118-5

Millar, E. V., Rice, G. K., Elassal, E. M., Schlett, C. D., Bennett, J. W., Redden, C. L., et al. (2017). Genomic characterization of USA300 methicillin-resistant *Staphylococcus aureus* (MRSA) to evaluate intraclass transmission and recurrence of skin and soft tissue infection (SSTI) among high-risk military trainees. *Clin. Infect. Dis.* 65, 461–468. doi: 10.1093/cid/cix327

Millar, E. V., Rice, G. K., Schlett, C. D., Elassal, E. M., Cer, R. Z., Frey, K. G., et al. (2019). Genomic epidemiology of MRSA infection and colonization isolates among military trainees with skin and soft tissue infection. *Infection* 47, 729–737. doi: 10.1007/s15010-019-01282-w

Moore, S., Penrice-Randal, R., Alruwaili, M., Dong, X., Pullan, S., Carter, D., et al. (2020). Amplicon based MinION sequencing of SARS-CoV-2 and metagenomic characterisation of nasopharyngeal swabs from patients with COVID-19. *medRxiv*[Preprint] doi: 10.1101/2020.03.05.20032011

Moran-Gilad, J., Sintchenko, V., Pedersen, S. K., Wolfgang, W. J., Pettengill, J., Strain, E., et al. (2015). Proficiency testing for bacterial whole genome sequencing: an end-user survey of current capabilities, requirements and priorities. *BMC Infect. Dis.* 15:174. doi: 10.1186/s12879-015-0902-3

Mullins, K. E., Hang, J., Clifford, R. J., Onmus-Leone, F., Yang, Y., Jiang, J., et al. (2017). Whole-genome analysis of *Bartonella ancashensis*, a novel pathogen causing verruga peruana, rural ancash region, Peru. *Emerg. Infect. Dis.* 23, 430–438. doi: 10.3201/eid2303.161476

Murray, C. K., Yun, H. C., Markelz, A. E., Okulicz, J. F., Vento, T. J., Burgess, T. H., et al. (2015). Operation united assistance: infectious disease threats to deployed military personnel. *Mil. Med.* 180, 626–651. doi: 10.7205/milmed-d-14-00691

Nadon, C., Van Walle, I., Gerner-Smidt, P., Campos, J., Chinen, I., Concepcion-Acevedo, J., et al. (2017). PulseNet International: vision for the implementation of whole genome sequencing (WGS) for global food-borne disease surveillance. *Euro Surveill.* 22:30544. doi: 10.2807/1560-7917.ES.2017.22.23.30544

Nilsen, P. (2015). Making sense of implementation theories, models and frameworks. *Implement Sci.* 10:53. doi: 10.1007/978-3-030-03874-8_3

Philipson, C., Davenport, K., Voegtly, L., Lo, C. C., Li, P. E., Xu, J., et al. (2017). Brief protocol for EDGE bioinformatics: analyzing microbial and metagenomic NGS data. *Bio Protoc.* 7:e2622. doi: 10.21769/BioProtoc.2622

Pollett, S., Fauver, J. R., Maljkovic, B, I., Melendrez, M., Morrison, A., Gillis, L. D., et al. (2019). Genomic epidemiology as a public health tool to combat mosquito-borne virus outbreaks. *J. Infect. Dis.* 221(Suppl. 3), S308–S318.

Pollett, S., Leguia, M., Nelson, M. I., Maljkovic Berry, I., Rutherford, G., Bausch, D. G., et al. (2016). Feasibility and effectiveness of a brief, intensive phylogenetics workshop in a middle-income country. *Int. J. Infect. Dis.* 42, 24–27. doi: 10.1016/j.ijid.2015.11.001

Quick, J., Loman, N. J., Duraffour, S., Simpson, J. T., Severi, E., Cowley, L., et al. (2016). Real-time, portable genome sequencing for Ebola surveillance. *Nature* 530, 228–232.

Ribot, E. M., Freeman, M., Hise, K. B., and Gerner-Smidt, P. (2019). PulseNet: entering the age of next-generation sequencing. *Foodborne Pathog. Dis.* 16, 451–456. doi: 10.1089/fpd.2019.2634

Riddle, M. S., Kaminski, R. W., Williams, C., Porter, C., Baqar, S., Kordis, A., et al. (2011). Safety and immunogenicity of an intranasal *Shigella* flexneri 2a Invaplex 50 vaccine. *Vaccine* 29, 7009–7019. doi: 10.1016/j.vaccine.2011.07.033

Rivers, C., Chretien, J. P., Riley, S., Pavlin, J. A., Woodward, A., Brett-Major, D., et al. (2019). Using "outbreak science" to strengthen the use of models during epidemics. *Nat. Commun.* 10:3102.

Robinson, J. T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., et al. (2011). Integrative genomics viewer. *Nat. Biotechnol.* 29, 24–26. doi: 10.1038/nbt.1754

Rocha, C., Bernal, M., Canal, E., Rios, P., Meza, R., Lopez, M., et al. (2019). First report of New Delhi metallo-beta-lactamase carbapenemase-producing *Acinetobacter baumannii* in Peru. *Am. J. Trop. Med. Hyg.* 100, 529–531. doi: 10.4269/ajtmh.18-0802

Salje, H., Lessler, J., Maljkovic Berry, I., Melendrez, M. C., Endy, T., Kalayanarooj, S., et al. (2017). Dengue diversity across spatial and temporal scales: local structure and the effect of host population size. *Science* 355, 1302–1306. doi: 10.1126/science.aaj9384

Shanks, G. D., and Hodge, J. (2011). The ability of seasonal and pandemic influenza to disrupt military operations. *J. Mil. Veterans Health* 19, 13–18.

Srijan, A., Margulieux, K. R., Ruekit, S., Snesrud, E., Maybank, R., Serichantalergs, O., et al. (2018). Genomic characterization of nonclonal MCR-1-positive multidrug-resistant *Klebsiella pneumoniae* from clinical samples in Thailand. *Microb. Drug Resist.* 24, 403–410. doi: 10.1089/mdr.2017. 0400

Sugiharto, V. A., Widjaja, S., Hartman, L. J., Williams, M., Myers, T. E., and Simons, M. P. (2019). Zika virus surveillance in active duty U.S. military and dependents through the Naval Infectious Diseases Diagnostic Laboratory. *MSMR* 26, 18–23.

Timme, R. E., Rand, H., Sanchez Leon, M., Hoffmann, M., Strain, E., Allard, M., et al. (2018). GenomeTrakr proficiency testing for foodborne pathogen surveillance: an exercise from 2015. *Microb. Genom.* 4:e000185.

Viral Diseases Branch WRAIR (2016). *ngs mapper*. Available online at: https://github.com/VDBWRAIR/ngs_mapper (accessed August 6, 2020).

Waickman, A. T., Victor, K., Li, T., Hatch, K., Rutvisuttinunt, W., Medin, C., et al. (2019). Dissecting the heterogeneity of DENV vaccine-elicited cellular immunity using single-cell RNA sequencing and metabolic profiling. *Nat Commun.* 10:3666.

Wiley, M. R., Fakoli, L., Letizia, A. G., Welch, S. R., Ladner, J. T., Prieto, K., et al. (2019). Lassa virus circulating in Liberia: a retrospective genomic characterisation. *Lancet Infect. Dis.* 19, 1371–1378. doi: 10.1016/s1473-3099(19)30486-4