Whole Genome sequencing in infectious diseases: Examples from the Ebola outbreaks.

Abstract:

Whole genome sequencing has transformed the public health microbiology field, through the increase in diagnostics precision and monitoring of pathogens. This essay uses the Ebola virus as a case study to explain the virus's natural reservoir and the difficulty in determining it, with a brief overview of its characteristics. Then it explores the role of genetics and WGS in pathogen detection, demonstrating how these technologies provide rapid and accurate diagnostic capabilities. Additionally, it investigates the role of WGS in outbreak detection and surveillance by looking into the different techniques and technologies available, highlighting its effectiveness in tracking the transmission of EBOV during outbreaks. Finally, it reviews the future application of such technologies, with a discussion regarding its limitations and challenges that hinder its use in public health microbiology.

Introduction:

Infectious diseases remain one of the leading causes of mortality worldwide. An estimated 13.7 million deaths per year are attributed to infectious diseases, with respiratory and bloodstream infections having the highest mortality rate (1). The past 2 decades have witnessed a revolution in medical care and access to the public health system to reduce the burden of infectious diseases on the human population. However, challenges emerged and continuous efforts to improve outcomes are at the forefront of science today, especially after the SARS-COV2 pandemic in 2019 that highlighted the need for a global effort to tackle public health issues. Since over 62% of human infections are zoonotic in origin(2), it is important to define the natural reservoir of a pathogen to understand its implication on human health and be able to track it. The advances in genetics and the introduction of newer technologies for Whole genome sequencing (WGS) allowed for a faster and deeper approach to studying and researching infections, by understanding the genetic host factors and pathogen genomic structure that influences them. WGS revolutionized public health microbiology by facilitating the detection and tracking of transmission between their reservoir and humans which can be essential for pandemic/endemic detection and surveillance. Using Ebola Virus (EBOV) as an example, after describing its natural reservoir and characteristics, this essay will explore how the knowledge acquired in the field of genetics and WGS can facilitate pathogen detection, how WGS can be used to enhance outbreak detection and surveillance, and finally, inspect the future applications of such technologies while addressing its limitations and challenges that face WGS use in public health microbiology.

Discussion:

1- The Natural Reservoir and Characteristics:

Ebola Viruses are a group of viruses that belong to the genus *Ebolavirus* and the family *Filoviridae*. Six different species of ebolavirus are well described and named after the region where they were discovered: Bundibugyo virus (BDBV), Ebola virus/Zaire Ebolavirus (EBOV), Reston virus (RESTV), Sudan virus (SUDV), Taï Forest virus (TAFV), and most recently Bombali Virus (BOMV). However, according to the World Health Organization, ebolavirus disease is caused by only EBOV (3). The different species within ebolavirus cause hemorrhagic fever, with symptoms including fever, diarrhea, vomiting, and pain.

The natural reservoir for EBOV is not clearly understood yet with different hypotheses explaining it. Ever since the discovery of ebolaviruses, there has been only one confirmed source of EVD, which occurred from great apes(4), however, the high mortality rate excludes them from being a natural reservoir since pathogen reservoirs require the host to be tolerant and support the replication and survival of the infective organism without being pathogenic. (5) argues that the widely accepted hypothesis for the EBOV natural reservoir is fruit bats in which they created a model to link the habitat of 9 bat species that are associated with EBOV infections. However, recently it has been speculated that insectivorous microbats (*Mops condylurus*) might be a more suitable host (6). This is based on *filovirus* entry mechanism which depends on Niemann-Pick C1 (NPC1) integral protein (7). Since *Mops condylurus* bats have a low level of NCP1, this influences the host-virus interaction affecting virulence and replication, which in turn supports the persistence of the infection in the host. This is a key factor in determining the natural reservoir of a pathogen. (8), were able to determine a link between NPC1 level and EBOV replication rate with lower levels attributed to high tolerance to EBOV infections, and repeated establishment of persistent infection. Continuous efforts are made to accurately determine the natural reservoir of Ebola, with technological advancement supporting such research.

EBOV characteristics, under electron microscopy, appear to have a filament-like shape (Figure 1) that can appear as long, unbranched, short, branched, or forming “6” and “U” configurations. Also, the Ebola virus has a single-strand RNA genome, which is negative sense(9). This means that the RNA

must be transcribed into a positive sense RNA before translation to protein, which is usually done by an RNA-Polymerase enzyme(10).

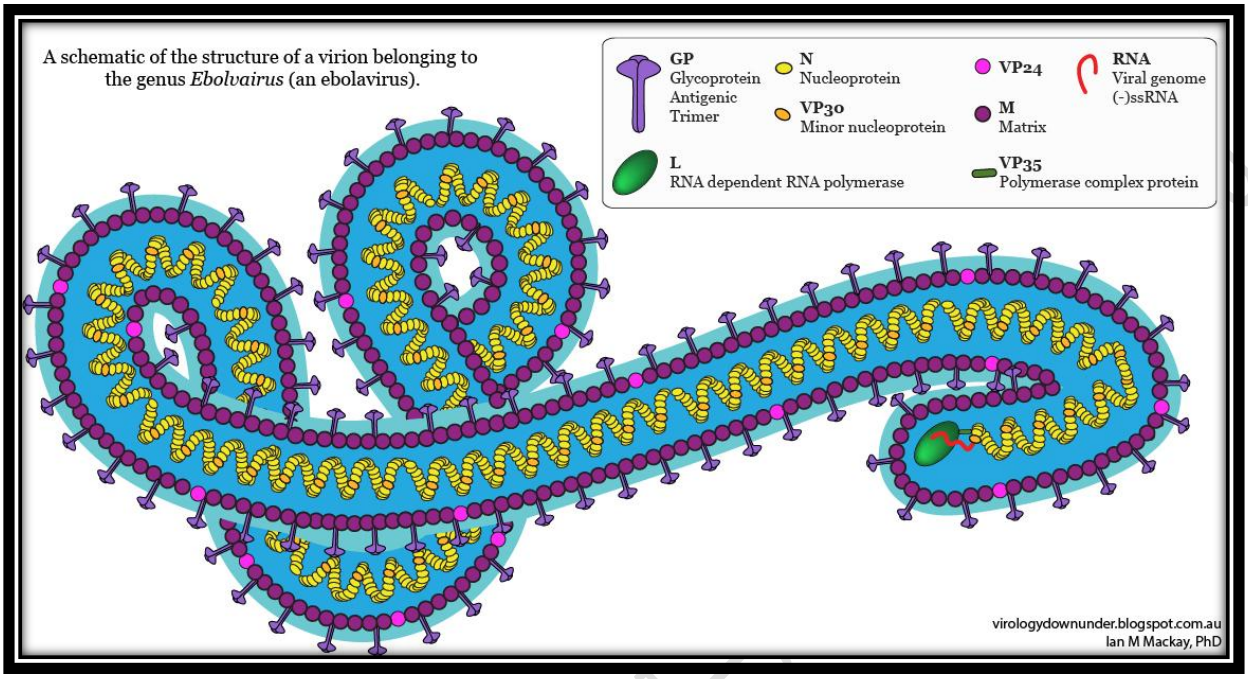

Figure 1: Schematic structure of EBOLAVIRUS(11)

The Ebola virus genome encodes for 7 genes: nucleoprotein (NP), viral protein 35 (VP35), VP40, glycoprotein (GP), VP30, VP24, and RNA polymerase (L) each having its specific function. NP is essential for transcription, with VP35 contributing to immune system evasion and GP for host cell interaction and entry (12). In addition, the virus has an incubation period of 2- 21 days, with signs of Ebola including fever and symptoms like severe headache, fatigue, muscle pain, vomiting, diarrhea, stomach pain, or bleeding or bruising according to the Centre of Disease Control (CDC). The Ebola virus is considered one of the most severe and complicated viral infections in humans, with mortality rates varying between 20 to 90 % depending on the infecting species, with EBOV having the highest rate of over 60% (13). Multiple outbreaks in the past 30 years warrant an extensive need for research, to enhance our understanding and therefore treatment of such diseases, using technological advances, especially in genetics and WGS.

2- Genetic and WGS in enhancing pathogen detection:

The first step in solving any infection management is accurate diagnosis. In the case of the Ebola virus, traditional methods for detecting the virus vary between hemagglutination inhibition tests enzyme-linked immunosorbent assays (ELISAs), and culture-based methods. Despite their effectiveness and simple handling methods, they still have many disadvantages, culture methods require long times for results and have a lower specificity and sensitivity associated with antibody tests(14).

With the fast-growing advancement in genetic and genomic techniques, molecular methods focused on detecting the nucleic acids of pathogens are ever-increasing in use for detecting infections in multiple settings. Genotyping and phenotyping methods, using genetic and WGS, allow for the detection of genes, genetic sequences, and virulence factors associated with a specific pathogen. Such methods include targeted sequencing, which employs the selection or enrichment of a specific genomic region in an organism. It is performed through either Polymerase chain reaction (PCR) or probe hybridization and Whole genome sequencing (WGS), where the entirety of a pathogen's genome is determined (15). According to (16), PCR genotyping using multiple-locus variable-number tandem-repeat analysis and Multi-locus sequence typing (MLST), can differentiate between strains, which aids in identifying transmission patterns and therefore enhance pathogen detection. In addition, it can offer a faster and more direct approach, where it is feasible to detect the virus directly from clinical samples. During the Ebola outbreak in West Africa, sequencing allowed for the detection of EBOV directly from 3 British healthcare worker samples, that were in Sierra Leone(17). This can help in identification, by eliminating the need for culture wait time without affecting the quality of the results. As a bonus, such methods can catch viruses that are unaccounted for, thus enhancing detection. This was seen during the 2014 outbreak in Guinea, West Africa while screening for EVD, 39 patients were found to be negative for the Ebola virus but detected infection with other organisms such as HIV-1, enterovirus A, and malaria and Lassa virus (18). Thus WGS can bypass multiple steps in pathogen detection, by looking directly into the genome.

Moreover, WGS via the use of metagenomics, can drastically change the game of pathogen identification. Metagenomics is defined as the sequencing technique where all the DNA/RNA materials are sequenced from a sample, and then the bioinformatics pipeline will filter out the host genome to detect a specific pathogen.(19) states that metagenomics is a powerful tool in pathogen identification, with a superior ability to detect novel viruses, and therefore pathogen emergence prediction. For example, metagenomics was used in 2014, during the Congo outbreak of EBOV, to

detect the presence of EBOV or other pathogens, and It was able to detect EBOV in multiple samples (20). Also, sequencing uncovered the virus in 22 of the original 70 samples, when it was done in a genomic lab in the US, despite some degradation of the RNA samples. This proves that WGS is a powerful tool in microbiology, and the subsequent effect of using such technologies can drastically change this field, by offering a rapid, culture-free, and robust identification method.

3- WGS in outbreak detection and surveillance:

Human history has witnessed many infectious disease outbreaks that devastated whole communities and sometimes countries, such as smallpox, Spanish flu, the ongoing HIV/AIDS pandemic, COVID-19, and the multiple Ebola outbreaks in the past 10 years. Therefore, it is crucial to implement outbreak detection and surveillance strategies that can stop or even predict global endemic/pandemics. Genetic sequencing through the use of WGS offers a new approach to detecting and surveilling infection. This is done via comparative genomic analyses with the help of phylogenetic models. Basically, phylogenetic models look into the viral sequences of pathogens collected from different samples and then examine the nucleotide substitution, which can infer a relationship between different clusters (21). In the case of Ebola, during the 2014 outbreak in West Africa, massive parallel sequencing revealed 341 fixed substitutions between all previously published EBOV sequences and the 2014 EBOV (35 nonsynonymous, 173 synonymous, and 133 noncoding)(18). This allowed us to understand whether the outbreak was a continuous one, or a new spillover from its natural reservoir, and it was found that this particular outbreak was likely originating from central Africa within the past decade. In another example, during the same period, WGS was able to differentiate between sexual transmission of EBVD and normal human-to-human transmission(22).In addition to outbreak detection, surveillance can also be inferred from whole genetic sequencing through the use of portable devices that eliminate the need for big machines. For example, during the EBOV outbreak in Guinea, WGS was used through the new technology of Nanopore sequencing, using a device that can be connected to a USB power outlet (Figure 2)(23) bypassing the logistical hurdles in remote areas where laboratories equipment and infrastructure can be limited. This technique allowed for

the sequencing of 142 Ebola virus genomes on site, and results were shared and sent to the cloud servers for analysis, coming back the next day(24).

Figure 2: MinION Portable Sequencing device.

As a result of such methods, genomic sequencing allows for the generation of data that can be shared among laboratories to help in outbreak surveillance. This data can be standardized and then integrated with epidemiological studies that can enhance public health measures in detecting infections. Such an example would be global health partnerships like the World Health Organization's (WHO) influenza surveillance system (25). (22) argues that to enhance the use of genomic data during outbreaks for surveillance, we can generate genomic-epidemiological reports from WGS and release them publicly for interactive analysis. This facilitates the surveillance of pathogens by providing rapid real-time sequencing directly from clinical samples and then establishing coordination with different disease control centers such as the CDC and WHO. This was seen during the 2015 Ebola Outbreak in Sierra Leone, where " The Ebola Outbreak Sequencing Support (EOSS) " was established to facilitate the transformation of raw data into a phylogenetic context to track the transmission of the virus (26). In addition, the integration of this data with metadata such as date, location, laboratory, and sources of samples, with publicly available databases and laboratories for rapid and robust monitoring services; and then linking it to epidemiological information provides an important tool to accurately

track the transmission. This provides the necessary knowledge that can be implemented to limit the spread and provide preventive measures against growing infectious threats.

4- The anticipated future application, limitations, and challenges associated with the use of WGS in public health microbiology:

In the last decade, WGS was not only used in research settings but also clinical practice and public health issues. This is mainly due to the outstanding efforts in advancing these techniques, that resulted in a continuous decrease of sequencing costs (27). Genomic sequencing is ever-changing and the use of WGS in this field has not reached a peak yet. With this rapidity in sequencing technologies advancement, the applications in public health microbiology are still under development and future applications are promising. Some of these applications, include but are not limited to enhanced surveillance, antimicrobial resistance (AMR) monitoring, and vaccine development.

According to (28), the potential for genomic data in public health, lies in its ability to forecast and rapidly monitor outbreaks through real-time sequencing and source attribution. This was seen during the 2016 Ebola outbreak and the 2018 Lassa virus (29, 30), where sequencing allowed for the rapid surveillance of Ebola using the portable sequencer from Oxford Nanopore Technologies, in addition to the prediction of the sources of rapidly increasing Lassa virus that was the result of a change in rodent reservoir population. More generally, AMR mitigation in public health via genomics is another future application. AMR is usually made via culture-based antimicrobial susceptibility testing (AST) in bacterial cases but not viral. Thus, the need for a different approach to viral diseases is necessary. WGS offers the ability to detect the genes responsible for resistance quickly and reliably (31), which aids in policy decisions and public health interventions to counter AMR. For example, the National Institute for Health and Care Research (NIHR) Global Health Research Unit aims to provide strategies to better implement the use of WGS in AMR surveillance, especially in low-income and middle-income countries. Lastly, genomic analysis when linked to clinical data, can inform about vaccine efficacy influencing vaccine design. According to (32), phylogenetic and photodynamic studies can alter vaccine design by looking into the mutations affecting its efficacy. This was seen during the SARS-COV-2 pandemic, where a mutation in a spike protein of SARS-COV-2 elucidated a change in the immune response to the vaccine, with the Omicron variant having the highest escape rate(33).

On the other hand, despite the robustness and importance of data generated from WGS in impacting public health policies, some challenges and limitations remain. Concerning the Data generated from WGS, handling them can be challenging. They require a high level of expertise in bioinformatics and data management to understand, store, and translate the information gathered into understandable language (Figure 3). According to (34), there is a limited presence of standardized guidelines for reporting genomic data, and the analysis reports might differ according to the diversity of end-user needs. Also,(35) state that Genomic-informed pathogen surveillance in Africa requires networking between different labs to implement standardized tools to enhance surveillance systems. Subsequently, quality assurance is limited, which affects the public health decisions that are extracted from such data. This is mainly due to the variation in DNA extraction methods, reagents used and technologies and bioinformatics pipeline used(36).

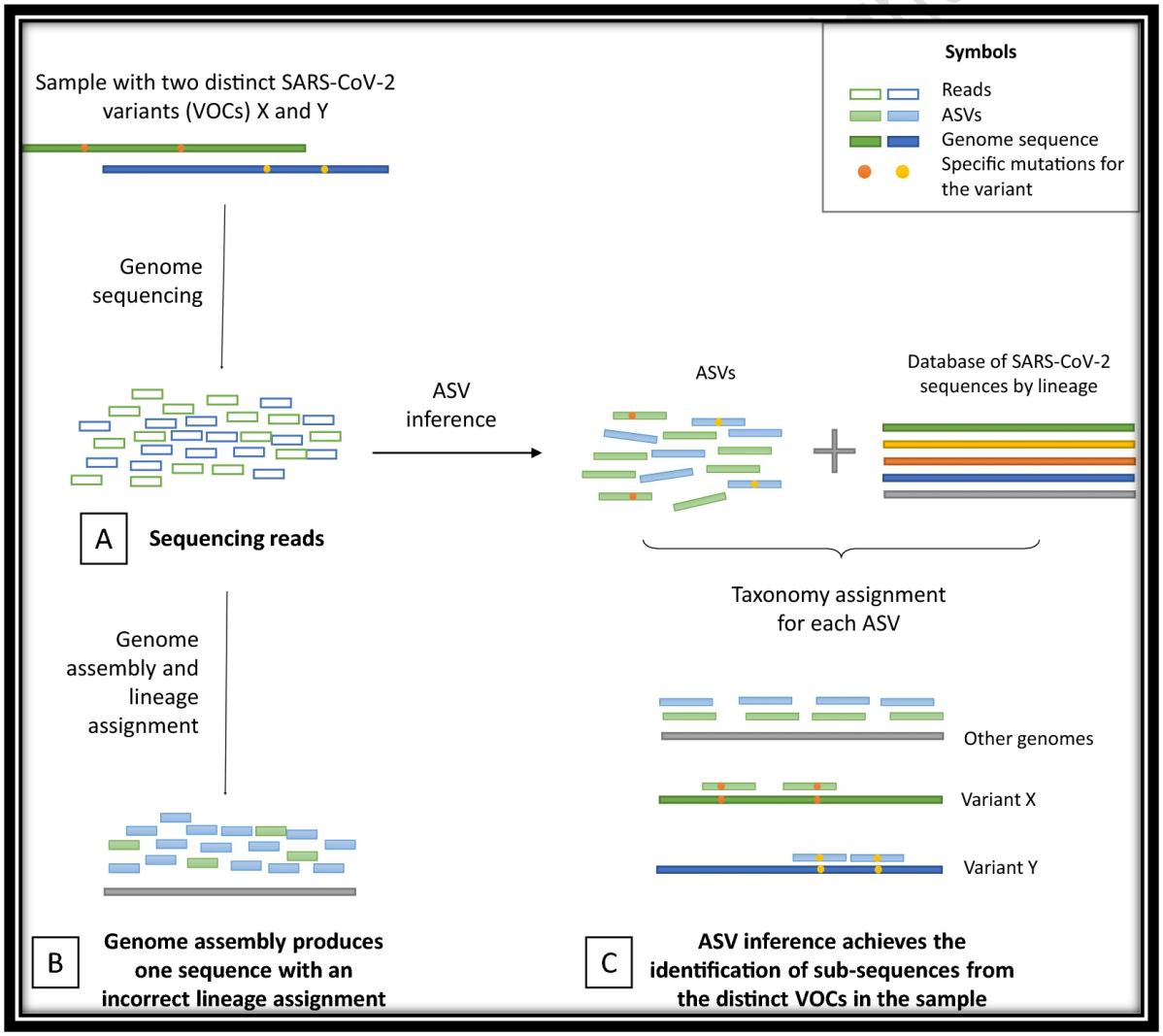

Figure 3: Bioinformatics Pipeline.(37)

In addition, some challenges that face WGS in public health microbiology are mainly logistical and financial in nature. This is especially the case in the global south countries. For example, in Africa, an examination of the landscape of sequencing capacity revealed that the majority of labs are concentrated in a few countries (South Africa Kenya Morocco, and Egypt)(35). This is mainly due to the high cost of setting up such equipment, since according to WHO the capital for such investment requires about US$ 100,000–700,000 (38). This hinders the ability to implement WGS in the public health system due to a shortage of required infrastructure. Subsequently, storing genomic data can be challenging, since they require huge databases and therefore necessitate the presence of specific hardware and software requirements that can be costly. Also, cost estimation for the implementation of WGS in public health is complicated(31). According to (39), producing a good cost-effectiveness analysis is challenging, despite the continuous decrease in the cost of sequencing. This will decrease investment by specific funding parties, due to a lack of clear financial reports which ultimately will negatively affect the implementation of WGS in public health.

Conclusion:

In conclusion, infectious diseases continue to pose a major risk to human health especially with the growing resistance to antibiotics, and the continuous emergence of new pandemics/endemics. Continuous investment in public health microbiology has proved to be a priority, especially since the SARS-COV-2 pandemic in 2019 that took the lives of millions and froze the economy. WGS has dramatically transformed public health microbiology by enabling rapid detection, tracking, and surveillance of infectious agents. WGS, with its capacity to analyze entire genomes, has facilitated a deeper understanding of pathogen genetics and host interactions. This was seen during the Ebola outbreak, where sequencing made surveillance easier, via the portability of sequencing devices. Future applications for WGS in public health microbiology are promising, especially in outbreak prediction, AMR surveillance, and vaccine development. However, these technologies are not free of limitations and challenges, particularly in terms of data management, standardization, and the financial and logistical barriers to widespread implementation, especially in low-income regions.

Addressing these challenges requires a continuous investment in infrastructure and training, that maybe can only be achieved through global and unanimous efforts.

References:

(11)

1. Gray A, Sharara F. Global and regional sepsis and infectious syndrome mortality in 2019: a systematic analysis. The Lancet Global Health. 2022;10:S2.
2. Taylor LH, Latham SM, Woolhouse ME. Risk factors for human disease emergence. Philosophical Transactions of the Royal Society of London Series B: Biological Sciences. 2001;356(1411):983-9.
3. Kuhn JH, Adachi T, Adhikari NK, Arribas JR, Bah IE, Bausch DG, et al. New filovirus disease classification and nomenclature. Nature Reviews Microbiology. 2019;17(5):261-3.
4. Leroy EM, Rouquet P, Formenty P, Souquiere S, Kilbourne A, Froment J-M, et al. Multiple Ebola virus transmission events and rapid decline of central African wildlife. Science. 2004;303(5656):387-90.
5. Koch LK, Cunze S, Kochmann J, Klimpel S. Bats as putative Zaire ebolavirus reservoir hosts and their habitat suitability in Africa. Scientific Reports. 2020;10(1):14268.
6. Marí Saéz A, Weiss S, Nowak K, Lapeyre V, Zimmermann F, Düx A, et al. Investigating the zoonotic origin of the West African Ebola epidemic. EMBO molecular medicine. 2015;7(1):17-23.
7. Hofmann-Winkler H, Kaup F, Pöhlmann S. Host cell factors in filovirus entry: novel players, new insights. Viruses. 2012;4(12):3336-62.
8. Bokelmann M, Vogel U, Debeljak F, Düx A, Riesle-Sbarbaro S, Lander A, et al. Tolerance and Persistence of Ebola Virus in Primary Cells from Mops condylurus, a Potential Ebola Virus Reservoir. Viruses. 2021;13(11):2186.
9. Reece R, Smit MA, Flanigan TP. Ebola Virus. In: Ratcliffe MJH, editor. Encyclopedia of Immunobiology. Oxford: Academic Press; 2016. p. 355-62.
10. Yu DS, Weng TH, Wu XX, Wang FXC, Lu XY, Wu HB, et al. The lifecycle of the Ebola virus in host cells. Oncotarget. 2017;8(33):55750-9.
11. Dr Ian M Mackay, virologydownunder.blogspot.com.au.
12. Jain S, Martynova E, Rizvanov A, Khaiboullina S, Baranwal M. Structural and Functional Aspects of Ebola Virus Proteins. Pathogens. 2021;10(10):1330.
13. Feldmann H, Geisbert TW. Ebola haemorrhagic fever. The Lancet. 2011;377(9768):849-62.
14. Liu Q, Jin X, Cheng J, Zhou H, Zhang Y, Dai Y. Advances in the application of molecular diagnostic techniques for the detection of infectious disease pathogens (Review). Mol Med Rep. 2023;27(5):104.
15. Satam H, Joshi K, Mangrolia U, Waghoo S, Zaidi G, Rawool S, et al. Next-Generation Sequencing Technology: Current Trends and Advancements. Biology (Basel). 2023;12(7).
16. Gilchrist CA, Turner SD, Riley MF, Petri WA, Jr., Hewlett EL. Whole-genome sequencing in outbreak analysis. Clin Microbiol Rev. 2015;28(3):541-63.
17. Bell A, Lewandowski K, Myers R, Wooldridge D, Aarons E, Simpson A, et al. Genome sequence analysis of Ebola virus in clinical samples from three British healthcare workers, August 2014 to March 2015. Euro Surveill. 2015;20(20).
18. Gire SK, Goba A, Andersen KG, Sealfon RSG, Park DJ, Kanneh L, et al. Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. Science. 2014;345(6202):1369-72.
19. Palacios G, Druce J, Du L, Tran T, Birch C, Briese T, et al. A new arenavirus in a cluster of fatal transplant-associated diseases. New England journal of medicine. 2008;358(10):991-8.
20. Li T, Mbala-Kingebeni P, Naccache SN, Thézé J, Bouquet J, Federman S, et al. Metagenomic Next-Generation Sequencing of the 2014 Ebola Virus Disease Outbreak in the Democratic Republic of the Congo. J Clin Microbiol. 2019;57(9).
21. Wohl S, Schaffner SF, Sabeti PC. Genomic Analysis of Viral Outbreaks. Annu Rev Virol. 2016;3(1):173-95.
22. Kinganda-Lusamaki E, Black A, Mukadi DB, Hadfield J, Mbala-Kingebeni P, Pratt CB, et al. Integration of genomic sequencing into the response to the Ebola virus outbreak in Nord Kivu, Democratic Republic of the Congo. Nature Medicine. 2021;27(4):710-6.

23. Quick J, Loman NJ, Duraffour S, Simpson JT, Severi E, Cowley L, et al. Real-time, portable genome sequencing for Ebola surveillance. Nature. 2016;530(7589):228-32.
24. Gwinn M, MacCannell D, Armstrong GL. Next-Generation Sequencing of Infectious Pathogens. JAMA. 2019;321(9):893-4.
25. Gwinn M, MacCannell DR, Khabbaz RF. Integrating advanced molecular technologies into public health. Journal of clinical microbiology. 2017;55(3):703-14.
26. Arias A, Watson SJ, Asogun D, Tobin EA, Lu J, Phan MVT, et al. Rapid outbreak sequencing of Ebola virus in Sierra Leone identifies transmission chains linked to sporadic cases. Virus Evolution. 2016;2(1).
27. Price V, Ngwira LG, Lewis JM, Baker KS, Peacock SJ, Jauneikaite E, et al. A systematic review of economic evaluations of whole-genome sequencing for the surveillance of bacterial pathogens. Microbial genomics. 2023;9(2):000947.
28. Stockdale JE, Liu P, Colijn C. The potential of genomics for infectious disease forecasting. Nature Microbiology. 2022;7(11):1736-43.
29. Siddle KJ, Eromon P, Barnes KG, Oguzie JU, Mehta S, Odia I, et al. Genomic analysis of Lassa virus from the 2018 surge in Nigeria. The New England Journal of Medicine. 2018;379(18):1745.
30. Quick J, Loman NJ, Duraffour S, Simpson JT, Severi E, Cowley L, et al. Real-time, portable genome sequencing for Ebola surveillance. Nature. 2016;530(7589):228-32.
31. Waddington C, Carey ME, Boinett CJ, Higginson E, Veeraraghavan B, Baker S. Exploiting genomics to mitigate the public health impact of antimicrobial resistance. Genome Medicine. 2022;14(1):15.
32. Struelens MJ, Ludden C, Werner G, Sintchenko V, Jokelainen P, Ip M. Real-time genomic surveillance for enhanced control of infectious diseases and antimicrobial resistance. Frontiers in Science. 2024;2.
33. McLean G, Kamil J, Lee B, Moore P, Schulz TF, Muik A, et al. The Impact of Evolving SARS-CoV-2 Mutations and Variants on COVID-19 Vaccines. mBio. 2022;13(2):e02979-21.
34. Ferdinand AS, Kelaher M, Lane CR, da Silva AG, Sherry NL, Ballard SA, et al. An implementation science approach to evaluating pathogen whole genome sequencing in public health. Genome Medicine. 2021;13(1):121.
35. Inzaule SC, Tessema SK, Kebede Y, Ogwell Ouma AE, Nkengasong JN. Genomic-informed pathogen surveillance in Africa: opportunities and challenges. The Lancet Infectious Diseases. 2021;21(9):e281-e9.
36. Van Goethem N, Descamps T, Devleesschauwer B, Roosens NHC, Boon NAM, Van Oyen H, et al. Status and potential of bacterial genomics for public health practice: a scoping review. Implementation Science. 2019;14(1):79.
37. Molina-Mora JA, Cordero-Laurent E, Calderón-Osorno M, Chacón-Ramírez E, Duarte-Martínez F. Metagenomic pipeline for identifying co-infections among distinct SARS-CoV-2 variants of concern: study cases from Alpha to Omicron. Scientific Reports. 2022;12(1):9377.
38. WHO. Whole genome sequencing for foodborne disease surveillance: landscape paper. World Health Organization Geneva; 2018.
39. Rossen JW, Friedrich AW, Moran-Gilad J. Practical issues in implementing whole-genome-sequencing in routine diagnostic microbiology. Clinical microbiology and infection. 2018;24(4):355-60.