BioMedInformatics

MDPI

Review

Genomics for Emerging Pathogen Identification and Monitoring: Prospects and Obstacles

Vishakha Vashisht [1], Ashutosh Vashisht [1], Ashis K. Mondal [1], Jaspreet Farmaha [1], Ahmet Alptekin [1], Harmanpreet Singh [1], Pankaj Ahluwalia [1], Anaka Srinivas [2] and Ravindra Kolhe [1,*]

[1] Department of Pathology, Augusta University, Augusta, GA 30912, USA; vvashisht@augusta.edu (V.V.); avashisht@augusta.edu (A.V.); amondal@augusta.edu (A.K.M.); jfarmaha@augusta.edu (J.F.); aalptekin@augusta.edu (A.A.); hsingh1@augusta.edu (H.S.); pahluwalia@augusta.edu (P.A.)
[2] Northwestern University, Evanston, IL 60208, USA; anakasrinivas2024@u.northwestern.edu
* Correspondence: rkolhe@augusta.edu

Abstract: Emerging infectious diseases (EIDs) pose an increasingly significant global burden, driven by urbanization, population explosion, global travel, changes in human behavior, and inadequate public health systems. The recent SARS-CoV-2 pandemic highlights the urgent need for innovative and robust technologies to effectively monitor newly emerging pathogens. Rapid identification, epidemiological surveillance, and transmission mitigation are crucial challenges for ensuring public health safety. Genomics has emerged as a pivotal tool in public health during pandemics, enabling the diagnosis, management, and prediction of infections, as well as the analysis and identification of cross-species interactions and the categorization of infectious agents. Recent advancements in high-throughput DNA sequencing tools have facilitated rapid and precise identification and characterization of emerging pathogens. This review article provides insights into the latest advances in various genomic techniques for pathogen detection and tracking and their applications in global outbreak surveillance. We assess methods that leverage pathogen sequences and explore the role of genomic analysis in understanding the epidemiology of newly emerged infectious diseases. Additionally, we address technical challenges and limitations, ethical and legal considerations, and highlight opportunities for integrating genomics with other surveillance approaches. By delving into the prospects and obstacles of genomics, we can gain valuable insights into its role in mitigating the threats posed by emerging pathogens and improving global preparedness in the face of future outbreaks.

Keywords: pathogen surveillance; emerging infectious diseases; genomics; high-throughput DNA sequencing; emergent pathogen

Citation: Vashisht, V.; Vashisht, A.; Mondal, A.K.; Farmaha, J.; Alptekin, A.; Singh, H.; Ahluwalia, P.; Srinivas, A.; Kolhe, R. Genomics for Emerging Pathogen Identification and Monitoring: Prospects and Obstacles. *BioMedInformatics* **2023**, *3*, 1145–1177. https://doi.org/10.3390/biomedinformatics3040069

Academic Editor: Alexandre G. De Brevern

Received: 21 September 2023
Revised: 12 October 2023
Accepted: 30 November 2023
Published: 7 December 2023

1. Introduction

In recent years, emerging infectious diseases (EID), originating from animals or through the increased prevalence of existing human diseases, have posed significant challenges to public health and global security [1]. Zoonotic viruses, transmitted from animals to humans, have been responsible for a significant percentage of these EID, including well-known cases like SARS, MERS, Ebola, HIV/AIDS, and Lyme disease [2,3]. The rise in these outbreaks can be attributed to various factors, including cross-species transmission and anthropogenic and environmental factors such as expanded agriculture, global travel, international trade, population growth, changing diets, and climate change [4–6]. The likelihood of inter-species transmission of pathogens has escalated remarkably due to interactions between humans and animals, encompassing activities such as hunting, animal husbandry, trade involving animal-derived edibles, wet markets, and the trade of exotic pets [7]. The dynamics of climate alterations also wield a considerable impact on the transmission patterns of pathogens, exemplified by diseases like Dengue, Chikungunya,

Zika, Japanese encephalitis, West Nile viruses, and *Borrelia burgdorferi* [8]. This is attributed to the expansion of suitable habitats for various prevalent vectors responsible for transmitting these zoonotic diseases, notably exemplified by the Aedes albopictus mosquitoes and ticks. The intrusion of vector-borne pathogens into regions where they were not initially endemic often culminates in explosive epidemic outbreaks. Alterations in land use driven by increasing human interference similarly influence the geographical dispersion of these vector-borne pathogens [9]. Moreover, the proliferation of numerous infectious diseases, such as tuberculosis, malaria, and cholera, across extended geographical boundaries has given rise to significant health apprehensions among substantial population segments [10]. These diseases exhibit an expanded range attributed to the acquisition of drug resistance, developing mosquito vectors' resistance to insecticides, suboptimal sanitation practices, climate fluctuations, and augmented human travel activities [11]. Additionally, cholera outbreaks in regions struck by natural catastrophes, like earthquakes and floods, have been well-documented [12].

Despite their efficacy, conventional pathogen identification and monitoring methods are frequently slow, labor intensive, and resource demanding. For instance, the established approach to characterize influenza strains commences with multiple culture passages, potentially leading to virus adaptations [13]. Genomics has revolutionized our ability to rapidly detect and characterize emerging pathogens, offering promising prospects for early identification and efficient monitoring of these novel pathogens. Since the 1990s, genomic approaches such as multiple DNA fingerprinting technologies have proven invaluable for characterizing different subtypes of *Mycobacterium tuberculosis* [14]. These technologies aid health department experts in identifying clusters of closely related strains, enabling the detection of cases potentially connected to recent transmission.

The advent of high-throughput DNA sequencing tools has markedly expedited identifying and characterizing emerging pathogens, making it faster, more cost effective, and more accurate [15,16]. The role of genomics and bioinformatics has been irreplaceable during the most recent COVID-19 pandemic, supplying invaluable insights into the architecture and functionality of genetic sequences and contributing to the diagnosis, management, and prediction of infections [17,18]. Nevertheless, challenges and limitations linked to genomic methods in pathogen detection and tracing remain to be addressed. A critical hurdle involves the development of robust and efficient sequencing technologies. Optimizing the cost and time associated with genome sequencing is imperative to enable widespread adoption, particularly in resource-constrained environments.

Moreover, the analysis and interpretation of extensive genomic datasets necessitate sophisticated bioinformatics tools and specialized expertise, which may not be universally accessible. Sharing genomic data also raises ethical and legal concerns centered on privacy and data protection. Striking a balance between the open sharing of data for global collaboration and protecting individual privacy is pivotal for effectively harnessing genomics for pathogen surveillance. Integrating genomics with other surveillance and diagnostic methodologies presents an opportunity for comprehensive pathogen detection and tracking. Continuous innovation and enhancement of genomics-based techniques harbor the potential to amplify our capability to detect and trace emerging pathogens [3,16] effectively.

This comprehensive review article offers an encompassing exploration of recent breakthroughs in genomic methodologies geared toward detecting and tracking emerging pathogens. It highlights the manifold applications of virus genomics in the realm of emerging pathogen detection, delving into technical impediments and limitations, addressing ethical and legal considerations, and illuminating avenues for synergizing genomics with other surveillance and diagnostic modalities.

2. Genomic Techniques for Pathogen Detection and Tracking

Genomic pathogen detection and tracking techniques involve the analysis of pathogens' genetic material to identify and trace the spread of diseases. These techniques rely on high-throughput DNA sequencing technologies to quickly sequence and analyze pathogen

genomes. By comparing the genetic profiles of different strains, researchers can determine the source of an outbreak, track its transmission patterns, and monitor the pathogen's evolution over time. Figure 1 shows the methodology of different genomic techniques used for pathogen detection.

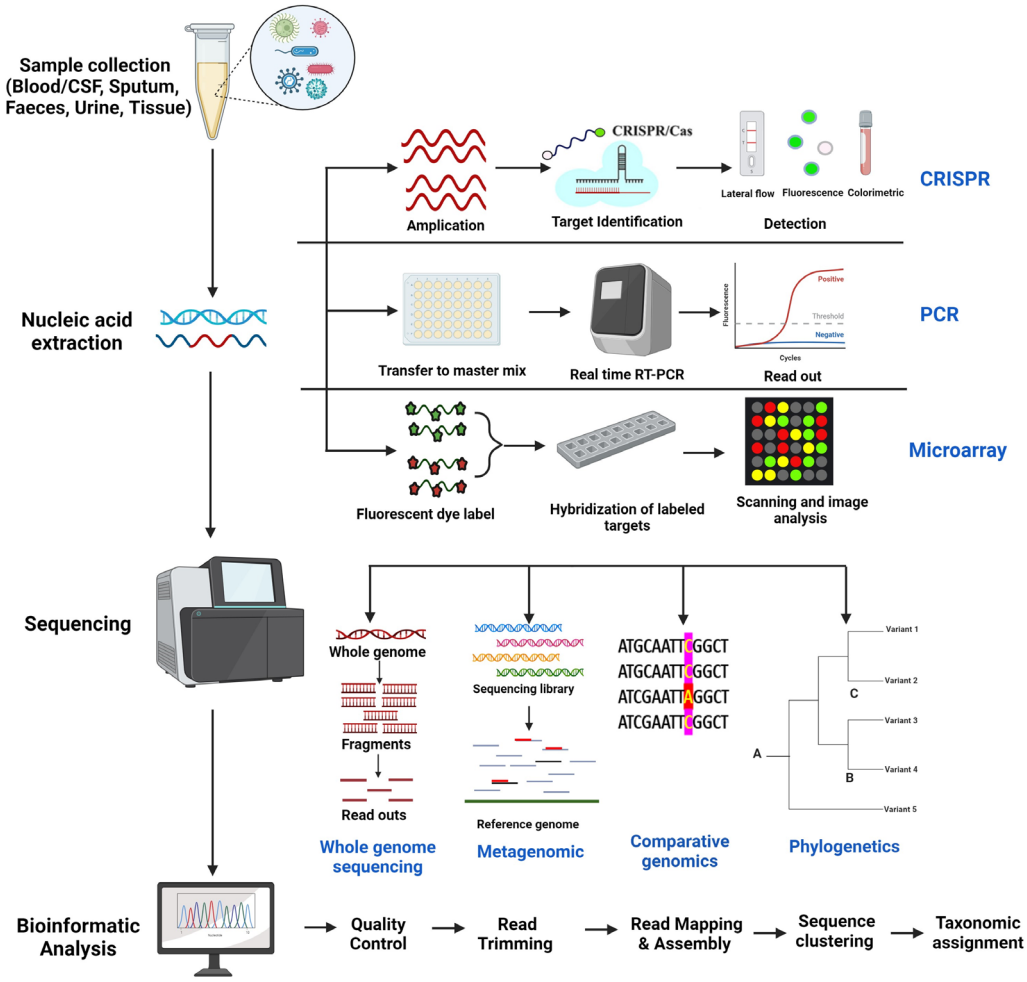

Figure 1. Showing schematic workflow for pathogen isolation and genomic analysis, the procedure integrates cutting-edge genomic techniques, including CRISPR, PCR, microarrays, sequencing, and pertinent bioinformatics pipelines.

2.1. Whole Genome Sequencing

Whole-genome sequencing (WGS) has emerged as a powerful tool in viral epidemiology research, particularly for analyzing pathogen outbreaks, by leveraging next-generation sequencing (NGS) technology. The comparison of different sequencing techniques for pathogen identification is given in Table 1. WGS can also be applied to cultural isolates of bacteria and fungi, enabling rapid pathogen identification, susceptibility testing to antimicrobials, outbreak investigation, and surveillance [17,18]. By tracking resistance mechanisms, including motifs on mobile genetic elements such as plasmids, and elucidating gene transfer mechanisms, WGS can provide insights into the spread of infection [19,20]. With the advent of NGS technologies, WGS has become an increasingly accessible and affordable tool for public health officials and researchers to investigate and control the spread of infectious diseases.

WGS has been used successfully in a wide range of public health investigations, from tracking the spread of the Ebola virus during the 2014–2016 outbreak in West Africa [21] to identifying the source of a food-borne illness outbreak caused by *E. coli* O157:H7 in England [22]. Faria et al. (2016) [23] used WGS to investigate the early spread of the

Zika virus in Brazil and identify a mutation associated with increased virulence and transmissibility in the Americas. WGS can provide valuable insights into pathogens' diversity, habitats, physiological traits, and virulence. One example is a study conducted by Forsythe et al. (2014) [24], which aimed to formally recognize the Cronobacter genus, a group of seven species known to cause neonatal infections. Through WGS, they discovered that only the *C. Sakazakii* clonal complex 4 (CC4) was predominantly associated with neonatal meningitis, highlighting the importance of genomics in identifying virulence factors and providing accurate pathogen classification. Koser et al. (2012) [25] used rapid WGS to investigate a hospital's neonatal methicillin-resistant *Staphylococcus aureus* (MRSA) outbreak. They found a close genetic connection between MRSA isolates in the neonatal and adult units, indicating cross-transmission. By sequencing environmental samples, they identified genetically identical MRSA, confirming the hospital environment as the outbreak source. The study demonstrated the power of WGS in tracing transmission and guiding control strategies for hospital-acquired MRSA infections. In a similar study, the researchers sequenced the genomes of 45 isolates of *Legionella pneumophila* from patients, the hospital water system, and the local environment. They found that all the patient isolates were closely related, indicating that they had a common source of infection [26]. The WGS analysis also revealed that the outbreak strain of *L. pneumophila* was most closely related to strains isolated from the hospital water system in previous years, suggesting that a persistent environmental source of the bacteria caused the outbreak. Grad et al.'s (2012) [27] study exemplifies the importance of multiple genome sequences in outbreak investigations. Using multiplatform WGS, they analyzed *Escherichia coli* O104:H4 isolates from outbreaks in Germany and France in 2011. Interestingly, the German isolates showed little diversity, whereas the French isolates showed greater diversity and formed a clade with the German isolates. The findings suggest possible factors influencing the observed diversity, such as a bottleneck effect, mutation rate variation, or differences in the initial seed populations leading to each outbreak.

WGS also plays a crucial role in understanding and responding to EIDs, including identifying and surveilling novel variants to understand their transmissibility, resistance to treatments and vaccines, and potential impact on clinical outcomes. A study by Frampton et al. (2021) utilized WGS during the emergence and spread of the B.1.1.7 variant of SARS-CoV-2 in the south of England, providing valuable insights into its impact on hospitalized patients. The findings indicate that B.1.1.7 is associated with higher viral loads than non-B.1.1.7 infections, as evidenced by PCR Ct values and genomic read depths. However, no evidence suggests that the B.1.1.7 variant is associated with increased disease severity or mortality. The study emphasizes the importance of rapid genomic surveillance and clinical data analysis to monitor and comprehend the impact of emerging variants on COVID-19 outcomes [28].

WGS has also been used to track the evolution and epidemiology of significant infections, including those with multidrug-resistant organisms [20]. In a study by Sharma et al., 2016, WGS confirms the clonal nature of *Candidia auris* strains in Indian settings, suggesting a common origin or recent differentiation. The strains analyzed showed low genetic diversity despite being isolated from different patients and hospitals. Furthermore, the WGS analysis revealed that *C. auris* has a divergent relationship with other *Candida* species but may be closely related to *C. haemulonii*. The presence of the MATα mating locus and the presence of ABC and MFS transporters may contribute to multidrug resistance in *C. auris*, suggesting that the emergence of *C. auris* as a successful multidrug-resistant pathogen may be linked to the indiscriminate use of antifungals [29]. In another example, Researchers used WGS and mycobacteriophage (Φ^2GFP10) in vitro to detect low-frequency *Mycobacterium tuberculosis* drug resistance and successfully detected resistant *M. tuberculosis* isolates at a low frequency of 1:100,000. Notably, they identified emergent drug resistance in one patient at three weeks, which conventional testing failed to detect [30].

Also, in 2021, WGS was used to track and analyze the Omicron variant's genetic makeup diligently, applied to over 90% of SARS-CoV-2 isolates, allowing for early detection

and comprehensive monitoring of Omicron's spread within the population [31]. Denmark's proactive use of WGS has proven instrumental in shaping effective public health responses to the Omicron variant. A study by Fonager et al., 2022, shed light on the utility of WGS for the prevalence and characteristics of two sublineages of the Omicron variant, BA.1 and BA.2, among SARS-CoV-2-positive cases. WGS played a pivotal role in this endeavor, enabling the precise differentiation and genetic analysis of these sublineages, uncovering lineage-specific mutations, especially in the spike protein, revealing their distinct features and the relatively comparable risk of hospitalization between the two sublineages [32].

One of the critical advantages of WGS over traditional sequencing methods is its ability to provide high-resolution genomic data that can be used for molecular epidemiology studies. By analyzing the entire genome of a pathogen, WGS can identify even subtle differences between isolates, allowing for accurate strain typing and tracking of outbreaks [8]. Whaley et al. (2018) [33] examined 15 *Neisseria meningitidis* outbreaks in the United States (2009–2015) and sporadic cases. WGS effectively identified genetically linked outbreak isolates when conventional methods failed. Ten outbreaks were caused by a single strain, whereas five involved multiple strains. Two sporadic isolates were also linked to outbreaks. By analyzing outbreak timing, researchers estimated the origin and emergence time of the most recent common ancestor. The findings suggest that U.S. meningococcal outbreaks were primarily caused by a single clonal strain in organizational settings or divergent strains in community outbreaks.

WGS has certain limitations, including the need for high-quality sequencing data, precise genome assembly, and specialized bioinformatics expertise, and barriers to adoption, especially in low- and middle-income countries, due to resource and human capital requirements [5]. Although WGS provides detailed genomic information, it does not directly translate to gene expression and transcription knowledge, necessitating RNA sequencing to detect gene expression patterns and enzyme activity [34]. One of the significant limitations of WGS in clinical laboratories is the lack of validation and utility comparisons, highlighting the need for standardized evaluation of WGS pipelines. While the analytical validity of WGS approaches is notably high and continuously improving, the clinical validity of the resultant data is considerably more intricate than commonly perceived. Moreover, its utility has often been evaluated in relatively small cohorts. Notably, a significant portion of genomic variations remains challenging to interpret; some are likely benign, and others may only be pathogenic under specific circumstances, potentially influenced by as-yet-undiscovered epistatic factors.

To address these complexities, there has been a growing inclination toward using gene panels or analyzing selected genetic segments. This may appear somewhat unconventional in the progression of whole-genome approaches. However, if we intend to utilize WGS for clinical inquiries, some form of filtering or focus on the sequence output becomes imperative. While targeted approaches have traditionally involved investigations, in WGS, the targeting occurs during the analysis phase. The pivotal challenge lies in translating genomic sequences into medically actionable information. Standardized protocols for sample acquisition and comprehensive epidemiological information are essential challenges to address [35]. Initiatives like the Global Biosurveillance Technology Initiative which is a part of the WHO Global Influenza Surveillance and Response System (GISRS) seek to establish standard NGS protocols worldwide [36]. The challenge of evaluating sequence quality and bioinformatic methods for variant identification can be overcome by developing standard metrics and using artificial DNA samples for result calibration across different sequencing platforms. Lastly, improving data analysis procedures, including accurate base calling and enhanced analytical techniques, is crucial to effectively handle the large volumes of data generated by various NGS platforms [34].

To strengthen and expand genomic surveillance for future, WHO developed a comprehensive strategy focusing on five key objectives: improving geographical representation, strengthening the workforce, enhancing data sharing, maximizing connectivity, and maintaining readiness. Multisectoral partnerships are crucial for successful implementation,

as exemplified by networks formed during the pandemic. The strategy, emphasizing the importance of resilient laboratory and surveillance systems, aims to detect pandemic and epidemic potential pathogens earlier, thus mitigating outbreaks and saving lives, and it underscores the significance of access to WGS for effective public health action [37].

Table 1. Comparison of different sequencing techniques for pathogen identification.

Sequencing Technique	Advantage	Disadvantage	Pathogen [Reference]
454 pyrosequencing	First developed HTS method—preferred sequencing method for metabarcoding projects for a while but now discontinued.	Lower throughput and subsequently higher sequencing cost per base	Ebola [38]
Illumina	A more popular choice for both metabarcoding and shotgun metagenomics studies. Best bases/cost ratio—Short (150–300 bps) but high-quality (99.9% accuracy) paired-end (P.E.) sequences—Very high level of accuracy	Does not perform well with low-quality material	SARS-CoV-2 [39]
			Mycobacterium tuberculosis [40]
			Candida auris [41]
IonTorrent	Bidirectional amplification and greater read length (400–450 bp)	Robust species inference	Zika [42]
PacBio	Produces long reads of 30–100 kb—Better average contig length and a higher number of large contigs in shotgun metagenomics studies. Allows for sequencing of longer PCR fragments such as the full ITS1-5.8S-ITS2 in metabarcoding studies.	Does not perform well with low-quality material. Higher rates of sequencing errors. Higher cost.	Influenza A (H1N1) [43]
Oxford Nanopore Technologies	Faster than Illumina or PacBio—Enables users to detect pathogens within minutes of the start of sequencing—small size and ability to be operated from a simple laptop.	Less accuracy (around 95% consensus)—Requires more DNA—More susceptib le to library construction or sequencing inhibitors.	Zika Virus in Brazil and America [44]
			Ebola Virus [45]

2.2. Metagenomics

Metagenomics is a rapidly evolving field that has become increasingly important in the context of pathogen surveillance. The COVID-19 pandemic highlights the need for improved global microbial surveillance through metagenomics. Metagenomics involves sequencing and analysis of genetic material from microbial communities, providing a comprehensive understanding without the need for culture or targeted assays [46]. Metagenomic workflows are tailored to specific surveillance objectives, resource availability, feasibility, and technical compatibility. The vast amount of sequencing data generated can be subsequently utilized to discover virulence genes, predict antibiotic resistance phenotypes, and notify enhanced outbreak investigation [47]. Many pathogens are challenging to culture in labs, leading to delays in diagnosis and treatment. Metagenomics overcomes this limitation by detecting all genetic material in a sample, including non-culturable microorganisms, by detecting pathogens that traditional culturing methods fail to identify. This is particularly valuable for emergent pathogens, which may be novel or possess unique characteristics that evade traditional detection methods. Metagenomic studies can now be performed from a single cell, expanding research possibilities. Notably, genomes can be assembled from organisms with low nucleic acid abundances, as low as 0.1%, enabling the detection of pathogens present in small quantities. Moreover, multiplex testing allows for the simultaneous detection of multiple pathogens, enabling the inference of interactions and a better understanding of microbial dynamics [46,47].

Public health metagenomics encompasses two approaches: deep amplicon sequencing and shotgun metagenomics. Deep amplicon sequencing focuses on specific taxonomic markers and amplifies them before sequencing, targeting a highly conserved region of the 16S rRNA gene to identify and classify bacterial and archaeal species in a sample. Kingrey et al. (2020) [48] highlighted the effectiveness of a high-throughput 16S V1-V2 metagenomics approach in detecting tick-borne bacterial pathogens in clinical specimens.

The method showed comparable accuracy to real-time PCR and successfully identified tick-borne pathogens at both genus and species levels. The study detected known tick-borne pathogens, such as *Anaplasma*, *Borrelia*, *Ehrlichia*, *Francisella*, and *Rickettsia*, and previously unrecognized tick-borne bacterial species. The results demonstrate the potential of this approach as a comprehensive diagnostic test for tick-borne infections, simplifying the diagnostic process and improving patient care. Shotgun metagenomics involves randomly fragmenting all DNA in a sample, sequencing the fragments, and then assembling them to reconstruct the genomes of the microorganisms in the sample. In shotgun metagenomics, the sequencing can be performed immediately or after an enrichment step, such as capture-based methods or subtraction, to enhance the detection of relevant sequences [46]. In a study by Balière et al., 2023 [49], the whole-genome sequences of a monkeypox virus (MPXV) from a French patient were obtained by shotgun metagenomics using Oxford Nanopore Technologies sequencing. The genomic sequences of the clinical sample and the isolate were successfully obtained even with a low-viral load sample. These techniques offer comprehensive insights into microbial communities, enabling the identification of emerging pathogens and providing valuable information for risk assessment and disease association studies [47]. Some of the examples of metagenomics for various emerging pathogen detection and surveillance in recent years are given in Table 2.

Metagenomic NGS (mNGS) analysis proves valuable in swiftly identifying and understanding novel viral strains, aiding in effective monitoring and response strategies during current and future pandemics. Gauthier et al. (2021) [50] used nanopore-based sequence-independent single primer amplification (SISPA) as a rapid and comprehensive diagnostic strategy for detecting and characterizing SARS-CoV-2 and its variants. The assay demonstrated high specificity (100%) and sensitivity (95.2%) for samples with a low RT-PCR cycle threshold value. The assay successfully distinguished between different variants, such as the alpha and gamma, which was impossible with the standard PCR method. Kugelman et al., 2015 [51], demonstrated the potential of this technology for rapid and accurate diagnosis of infectious diseases by detecting the presence of the Ebola virus in patient samples within hours of collection during the 2014 Ebola outbreak in West Africa.

Similarly, a study of the 2016 Zika virus outbreak in Brazil used metagenomic sequencing to identify viral sequences in patient samples, providing a more comprehensive view of the viral diversity present in the outbreak [23]. Claro et al., 2022 [52], used shotgun metagenomics to detect the monkeypox virus in Brazil. Greninger et al., 2010 [53], used metagenomics to detect and characterize the 2009 H1N1 flu virus, demonstrating its potential to replace multiple conventional diagnostic tests in investigating outbreaks. By combining virochip and deep sequencing, the researchers successfully identified the virus, analyzed its genetic information, and gained insights into host responses and other microorganisms present in the respiratory system.

In addition to identifying emergent pathogens, metagenomics can also be used to monitor the spread and evolution of these pathogens. For example, a study of the 2013–2014 MERS outbreak in Saudi Arabia used metagenomics to track the evolution of the virus, providing insight into its origins and spread [54]. Metagenomics can also provide a more comprehensive view of the microbial populations present in a given environment, allowing for the identification of potential sources of infection or transmission. For example, a study of the gut microbiome in patients with *Clostridium difficile* infection identified a specific bacterial species associated with increased susceptibility to the infection, highlighting the potential of metagenomics for identifying microbial risk factors [55]. Sheahan et al., 2019 [56], used rapid metagenomics for monitoring pathogen load from EMS (emergency medical service) vehicles. More than 68 species in ambulances were identified using the nanopore DNA sequencing technology, and some of these were linked to illnesses common in healthcare settings, including *Clostridium* spp. and *Staphylococcus* spp.

Table 2. Metagenomics for emerging pathogen detection and surveillance.

Pathogen	Metagenomics Implications	Platform Used	References
Zika virus	Detected in Aedes mosquitoes during the epidemic. Arbovirus detection can be a useful tool for identifying epidemic-causing arboviruses.	Illumina MiSeq, Illumina HiSeq	[57]
Ebola virus	Broad-based pathogen detection and outbreak surveillance	Ion Torrent PGM	[58]
MERS-CoV	Rapid sequencing for genotype information and co-infections enables identification of genotype changes, including insertions, deletions, and minor variants, while also providing insights into the background microbiome.	Amplicon-based approach coupled to Oxford Nanopore long read length sequencing	[59]
SARS-CoV-2	It successfully assembled complete or near-complete genomes and accurately classified phylogenetic lineages, including the identification of Variant of Concern (VOC) strains. The assay's capability to distinguish between different SARS-CoV-2 variants, such as Alpha and Gamma, surpassed the standard VOC PCR method.	Nanopore-based Sequence-Independent Single Primer Amplification (SISPA)	[50]
Chikungunya Dengue, Zika virus	Viral metagenomics was found to be a potent method for the identification of emerging arboviruses.	Illumina NextSeq 2000	[60]
Avian influenza virus (H7N9)	Viral infection surveillance in poultry farms.	Ion Torrent PGM	[61]
Influenza virus	Diagnostic test, insights on transmission, evolution, and drug resistance.	Oxford Nanopore	[62]
Shiga-toxigenic *Escherichia coli* (STEC) O104:H4	Identification and characterization of bacterial strains during diarrheal disease outbreaks, including the STEC outbreak strain, as well as detection of other pathogens.	Illumina HiSeq 2500 Illumina MiSeq2500	[63]

Metagenomic surveillance for pathogen detection faces challenges in analyzing large datasets, distinguishing between pathogens and non-pathogens, and a lack of standardization in data analysis. Optimization, standardization, and automation efforts can improve assay reproducibility and facilitate adoption. Advancements in sequencing and bioinformatics enhance the feasibility of metagenomic surveillance, but further research and data standardization are needed for comprehensive pathogen risk assessment [15,47]. By creating a unified One Health network and using high-resolution transmission mapping, we can identify sources of concern and develop targeted mitigation strategies. This data-driven approach can boost infectious disease research and inform evidence-backed public policies, helping us prevent future pandemics [15].

2.3. Comparative Genomics

Comparative genomics is a valuable field that investigates the genetic makeup of diverse organisms, aiming to comprehend their similarities and differences. This approach enables the exploration of evolutionary relationships, the identification of disease-related genetic variations, and a deeper understanding of gene functions and other genomic elements [64]. In the context of emergent pathogens, comparative genomics plays a crucial role in their detection and characterization. Sequencing the genomes of various pathogen isolates and comparing them to each other and known pathogens can identify specific genetic markers associated with virulence or pathogenicity [65,66]. Comparative genomics is the foundation for epidemiological research and the molecular evolution of related pathogens. These insights contribute to developing effective public health strategies, including nationwide surveillance programs and the creation of vaccines, to combat potential outbreaks [64–66].

In 2017, Zhang et al. [67] performed a comparative genomic analysis of human adenovirus type 14 (HAdV-B14) strains associated with outbreaks in the USA, Canada, UK, Ireland, and China. They observed that the three Chinese strains were similar to the original prototype but different from the U.S. strain. There were also slight differences between the strains found in southern and northern China. These findings suggest the possibility of multiple lineages of HAdV-B14 and either independent introductions from abroad or subsequent divergence from a single lineage. Understanding these emerging strains is essential for developing vaccines and public health strategies to control future outbreaks with potentially high fatality rates. Faria et al., 2017 [44], used comparative genomics to analyze the genetic diversity of the Zika virus during the 2015–2016 outbreak in the Americas. The researchers found that the virus had been introduced to Brazil from multiple sources, and different strains had circulated in different regions. During the hemolytic-uremic syndrome outbreak caused by the rare Shiga-toxin-producing *E. coli* O104:H4 strain, the genomes of the German outbreak strain and related strains from Africa were analyzed. They were found to belong to the enteroaggregative pathotype and had distinct genetic characteristics compared to other O104:H4 strains. The German strain carried a prophage encoding Shiga toxin 2 and exhibited unique virulence factors and antibiotic resistance genes. Horizontal genetic exchange played a role in the emergence of this highly virulent strain, emphasizing the adaptability of bacterial genomes in the development of new and potentially harmful pathogens [68]. Comparative genomics can also study emergent pathogens' mutations, transmission, and spread. Ahammad et al., 2021, analyzed the comparative genomic data of SARS-CoV-2 during the first and second waves of the COVID-19 pandemic in Bangladesh [69]. The second wave showed a higher mutation rate, particularly in the deletion and transversion events, linked to altered antigenicity and antibody escape mechanisms. The predominant clade also shifted from G.R. to G.H., and the most common variants changed from B.1.1.25 to B.1.351.3. Asrani et al., 2020, analyzed the genetic modifications of SARS-CoV-2 and other CoV strains using comparative genomics [70]. They found variations in the number of nsp and accessory proteins between different CoV strains, which could lead to differences in their pathogenicity. For instance, MERS-CoV had a higher fatality rate than SARS-CoV and SARS-CoV-2, while SARS-CoV-2 has a higher transmission rate than the other CoVs, possibly due to differences in their genetic structures. Khan et al. (2020) analyzed the complete genome sequences of 13 SARS-CoV-2 isolates from different countries and compared them with SARS-CoV [71]. They found that the SARS-CoV-2 isolates showed high identity with each other (>99%), but only 82% identity with SARS-CoV. The researchers focused on three major proteins of the virus—Mpro, RdRp, and spike proteins—which play important roles in the virus's life cycle and interaction with host cells. They found only one point mutation (R60C) in the Mpro of the Vietnam strain of SARS-CoV-2 (Figure S1a) and one mutation (A406V) in the RdRp of the Indian SARS-CoV-2 isolate (Figure S1b). The spike proteins showed five amino acid mutations at various positions within the investigated SARS-CoV-2 isolates, including India, Finland, Australia, South Korea, and Sweden. Additionally, one amino acid change occurred in each envelope protein of the South Korea SARS-CoV-2 isolate and the nucleocapsid protein of the Japan SARS-CoV-2 isolate (Figure S1c). Finally, the researchers found that all 13 SARS-CoV-2 isolates had a deletion of glycine and serine at positions 70 and 8 in the envelope and nucleocapsid proteins compared to SARS-CoV (Figure S1d,e).

Comparative genomics can be used to develop effective vaccines against emergent pathogens. Cheng et al., 2018 conducted comparative genomics to study the HAdV-B55 strain and its genetic changes over time [72]. The study found a high degree of genome identity between the 2011 and 2016 strains, with only a few substitutions and indels. Non-synonymous substitutions in protein pVI were associated with the population distribution of HAdV-B55, but no changes were observed in the major capsid proteins responsible for type-specific neutralizing antibodies. These findings will be helpful for the development of highly effective vaccines that specifically target the re-emergent HAdV-B55 strains linked

to severe CAP in adults. Comparative genomics can also identify the genetic changes that enable pathogens to adapt to new hosts or environments. For example, Jun et al., 2015, used comparative genomics to analyze the evolution of the Ebola virus during the 2014 outbreak in West Africa [73]. The researchers found that the virus had acquired several genetic mutations that enabled it to infect human cells more efficiently and evade the human immune system. By understanding the genetic changes that allowed the virus to adapt to human hosts, researchers may be able to develop more effective treatments and vaccines and predict and prevent future outbreaks.

While comparative genomics has many potential applications in emergent pathogen surveillance, its use has several challenges and limitations. One of the main challenges is the need for high-quality genome sequences from a diverse range of pathogens and their host organisms and vectors. Obtaining and analyzing these sequences can be time-consuming and expensive, particularly for newly emergent pathogens or complex genomes. In addition, comparative genomics can be limited because genetic changes may have different effects in different host species or environments, making it difficult to predict the behavior of emergent pathogens in the real world [64,66]. Despite these challenges, comparative genomics has proven to be a valuable tool for identifying and characterizing emergent pathogens. It will likely be increasingly important in future infectious disease surveillance and control.

2.4. Phylogenetic Analysis

Phylogenetic analysis uses molecular data to track the transmission of pathogens and prevent future epidemics by identifying and monitoring the emergence and spread of new and existing pathogens. The relationships are inferred or estimated using phylogenetic analysis, which produces treelike diagrams representing the evolutionary history or pedigree of the inherited relationships among these entities. These diagrams are sometimes called "gene trees" or "phylogenetic trees". These trees are built based on shared ancestry, with closely related organisms sharing a more recent common ancestor and having more similar genetic or morphological characteristics. Key methods in phylogenetics include molecular techniques like DNA sequencing and computational algorithms like Maximum Likelihood or Bayesian inference. The phylogenetic study can help analyze the evolution and the similarities among diseases and viruses and further help prescribe vaccines against them [74]. Phylodynamic analysis has proven valuable in recent pandemic events such as SARS, avian influenza H5N1, Ebola, and Zika [74,75]. Figure 2 illustrates the interconnected pathways of evolution, highlighting the identification and isolation of genomic sequences that may contribute to future pandemics. Additionally, it outlines technical approaches to enhance the analysis of emerging viral diseases [75]. The use of phylogenetic analysis in emergent pathogen surveillance offers numerous advantages. The phylogenetic analysis enables researchers to identify the pathogen's origin and its transmission pathways by comparing the genetic sequences of pathogens from different individuals or populations. Hodcraft et al., 2020 [76], used phylogenetic analysis to track a SARS-CoV-2 variant, 20E (EU1), spread across Europe in the summer of 2020. Based on phylogenetic analysis, it was suggested that Spain likely originated most introductions of this variant into other countries. The variant initially spread among agricultural workers in Spain and then moved into the local population, eventually reaching other regions of the country. The phylogenetic tree shows that the variant has diverse genotypes observed in multiple European countries, indicating that it was introduced into different countries multiple times. The variant spread was primarily driven by travel between countries, especially during the summer when travel restrictions were eased. The large number of introductions can explain the rise in the variant frequency through travel rather than intrinsic differences in transmission advantage.

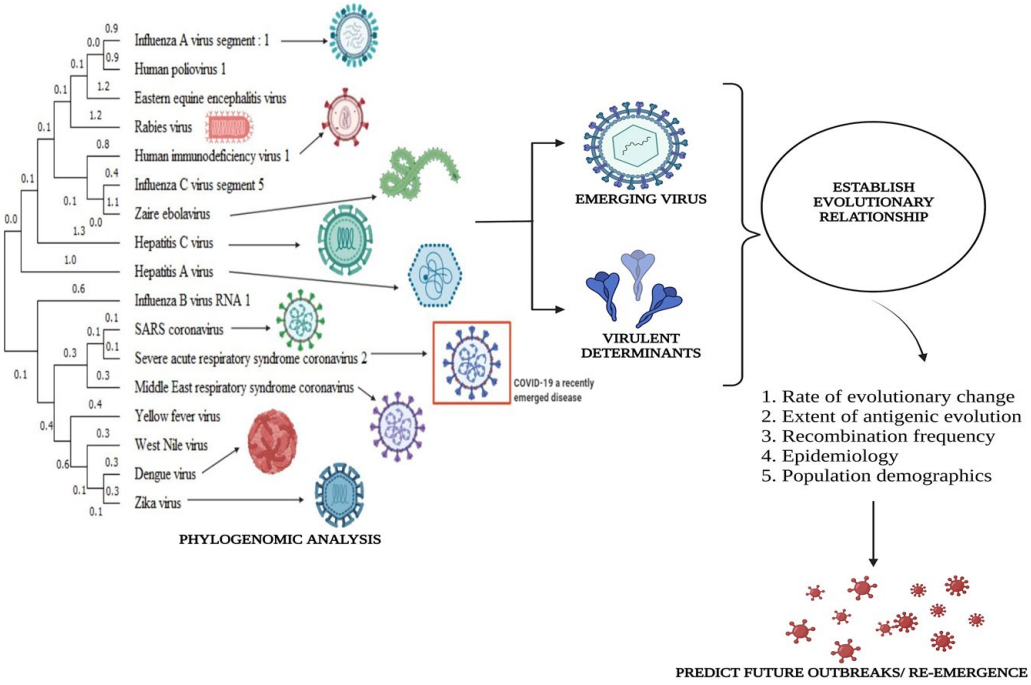

Figure 2. Phylogenetic tree shows the similarity between various single-stranded RNA viruses that have caused a pandemic over time and technical strategies to further analyze emerging viral diseases. Reused with permission from Elsevier Behl et al., 2022 [75].

Zhou et al. conducted a comprehensive phylogenetic analysis of RdRp gene sequences of a bat coronavirus (BatCoV RaTG13) and spike (S) protein, which revealed that RaTG13 is the closest known relative to 2019-nCoV, and they together form a distinct lineage that is distinct from other SARS-CoVs [77]. The close phylogenetic affinity with RaTG13 strongly supports the hypothesis that 2019-nCoV originated from bats. Forster et al., 2020, analyzed 160 SARS-CoV-2 genomes and found three central variants (A, B, and C) distinguished by amino acid changes, with A being the ancestral type. A and C types are common in Europeans and Americans, whereas B is most common in East Asia [78]. Dudas et al., 2017, conducted phylodynamic analysis on existing MERS-CoV sequences to investigate its two main hosts: humans and camels. Using coalescent models, they demonstrated that long-term MERS-CoV evolution occurs exclusively in camels, whereas humans act as transient and ultimately terminal hosts. By analyzing the distribution of human outbreak cluster sizes and zoonotic introduction times, the researchers also revealed that human outbreaks in the Arabian Peninsula are driven by the seasonally varying zoonotic transfer of viruses from camels [79]. Phylogenetic analysis can also provide insights into the transmission dynamics of pathogens. By analyzing the genetic sequences of pathogens from different individuals or populations, researchers can determine how the pathogen is spreading and identify factors that may be contributing to its spread. For example, phylodynamic and phylogeographic analyses Bayesian Evolutionary Analysis Sampling Trees (BEAST) was used to study the transmission of HIV in a high-prevalence population in Uganda and the subsequent spread in Kampala and its extensive interconnection with adjacent countries. By analyzing the genetic sequences of HIV from different individuals, researchers could identify transmission patterns and suggest that a significant proportion of new infections were occurring within stable partnerships [80]. Benvenuto et al., 2019, used Bayesian phylogenetics and phylodynamic analysis to study the genetic diversity and transmission dynamics of the Madariaga virus (MADV), a member of the eastern equine encephalitis virus (EEEV) complex, which is being watched as an emerging pathogen threat in South America. Two main clusters were identified in the phylogenetic trees, with Lineage II showing an epizootic infection in monkeys and Lineage III showing an epizootic infection in humans in Haiti and Venezuela. The tree's root was estimated to date back

to year 346, with probable origin in Brazil. Gene flow analysis revealed viral exchanges between different South American countries. The virus shows high conservation, possibly remaining concealed under the Dengue umbrella. The findings underscore the need for improved molecular epidemiological surveillance to prevent potential outbreaks and better to understand the transmission patterns of emerging pathogens like MADV [81].

Chen et al. (2018) conducted a study investigating the phylogenetic analysis of *Vittaforma cornea*, a microsporidian parasite found in aquatic environments that can cause ocular infection. Phylogenetic analysis revealed 12 clusters of clade IV microsporidia in the samples, with Cluster A1 representing a diverse group of *Vittaforma*-like fungi associated with symptomatic diarrhea. Cluster B was identified as ocular microsporidia, and Cluster C showed potential for zoonotic transfer. A novel microsporidia group (Cluster F) was also found in aquatic environments. The close phylogenetic relation between taxa infecting humans and animals suggests the potential for zoonotic transmission [82]. Bowers et al., 2018, attempted to clarify the strain nomenclature of CC8, classifying the major strain types based on WGS phylogenetics using both methicillin-resistant *S. aureus* (MRSA) and methicillin-susceptible *S. aureus* (MSSA) genomes. The study found that isolates of the Archaic and Iberian clones from decades ago make up the most basal clade of the main CC8 lineages and that at least one successful lineage of CC8, made up mostly of MSSA, diverged before the other well-known strain types USA500 and USA300. They also showed the significant benefit of using more stable genomic markers based on evolutionary lineages over traditional *S. aureus* typing techniques. This more accurate and accessible *S. aureus* typing system may improve surveillance and better inform the epidemiology of this significant pathogen [83].

Phylogenetic analysis, while valuable for understanding virus evolution, has several limitations. It often requires a large amount of genetic data, which may not be available in low-resource settings, and the data's quality can impact the analysis's accuracy. Care must be taken in interpreting phylogenetic clusters since they are based on partially sampled transmission chains, potentially leading to erroneous conclusions. Bias can be introduced through genetic markers, sample selection, and analysis methods. Recombination or horizontal transfer events in the pathogen's evolutionary history can compromise accuracy [84]. Uneven geographic coverage and limited understanding of genetic determinants and virulence factors pose challenges for surveillance. The complexity of viral quasi-species and limitations in detecting minor genetic variants also present obstacles. Nevertheless, phylogenetic analysis offers valuable insights into virus evolution that can inform public health measures, vaccine design, and control strategies [85].

3. Other Genomic Techniques

CRISPR-based genomic and PCR-based techniques are commonly used for pathogen detection and tracking due to their high sensitivity and specificity. These techniques involve amplifying and targeting specific genetic sequences to identify and characterize pathogens.

3.1. CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) Based Methods

CRISPR-based diagnostic technologies, such as DETECTR (DNA Endonuclease Targeted CRISPR Trans Reporter) and SHERLOCK (Specific High-sensitivity Enzymatic Reporter un-LOCKing), have shown great promise in revolutionizing molecular diagnostics. The main component of a surveillance CRISPR/Cas system is a ribonucleoprotein formed by combining a Cas effector complex with a single guide RNA (sgRNA). The sgRNA consists of a non-coding CRISPR RNA (crRNA) and a trans-activating crRNA (tracrRNA). This complex acts as a molecular scissor for targeted surveillance [86]. These technologies offer portable, highly sensitive tools for rapidly diagnosing infectious and noninfectious diseases. They can potentially improve population screening, control infectious outbreaks, and provide affordable diagnostics in resource-limited settings [87]. CRISPR demonstrated versatility in detecting various infectious agents (Table 3).

Chen et al., 2018, reported a breakthrough in nucleic acid detection by developing DETECTER. By combining Cas12a's non-specific ssDNA cleavage activity (which is the ability of the enzyme to cut or break single-stranded DNA molecules into smaller pieces) with isothermal amplification (i.e., under constant temperature conditions unlike PCR), they achieved highly sensitive detection of DNA, specifically demonstrating its ability to detect human papillomavirus (HPV) in patient samples. The primary advantage of DETECTR is its high sensitivity, as it can detect a single molecule of viral particle within a microliter of sample. This study reveals a new aspect of Cas12a proteins, highlighting their ability to unleash non-specific ssDNA cleavage when bound to a guide RNA and complementary ssDNA. This discovery has implications for understanding bacterial immune systems and developing improved molecular diagnostic tools [88]. A lateral flow assay based on OR-DETECTR was also developed for convenient COVID-19 detection. Sun et al., 2021 [89], developed OR-DETECTR, a one-tube detection platform for rapid and accurate detection of SARS-CoV-2. It combines RT-RPA and DETECTR technologies, providing results in approximately 50 min. The platform showed high specificity and sensitivity, detecting SARS-CoV-2 from samples of human coronaviruses and Influenza A (H1N1). Broughton et al., 2020, developed CRISPR–Cas12-based assay for the rapid detection (40 min) of SARS-CoV-2 from respiratory swab RNA extracts with 95% positive and 100% negative predictive agreement [90]. In the SHERLOCK approach, RNA detection is achieved through the CRISPR-Cas13a system, which targets specific RNA sequences via complementary CRISPR RNAs (crRNAs). Patchsung et al., 2020, clinically validated the SHERLOCK assay for detecting SARS-CoV-2. The Cas13a enzyme from Leptotrichia wadei was found to be 100% specific and 100% sensitive with a fluorescence readout and 100% specific and 97% sensitive with a lateral-flow readout within a detection limit of 42 RNA copies per reaction [91]. Gootenberg et al., 2017 [92], conducted experiments using lentiviruses containing genome fragments of Zika virus (ZIKV) or the related flavivirus Dengue virus (DENV). SHERLOCK successfully detected viral particles at concentrations as low as 2 attomolar (2 aM) and demonstrated the ability to distinguish between ZIKV and DENV.

Agarwal et al., 2021, developed a microfluidic platform called DISCoVER (Diagnostics with Coronavirus Enzymatic Reporting), which combines CRISPR-based molecular diagnostics with optimized DNA and RNA amplification mechanisms, along with Cas13 detection for specificity. The test demonstrated high sensitivity, detecting SARS-CoV-2 in unextracted saliva samples with as few as 40 copies/μL within 35 min [93]. In a study by Wang et al. (2020), the CRISPR/Cas system was integrated into a lateral flow assay to create the CRISPR/Cas9-mediated lateral flow nucleic acid assay (CASLFA). The study demonstrated the effectiveness of CASLFA in detecting Listeria monocytogenes, genetically modified organisms (GMOs), and African swine fever virus (ASFV) with 100% accuracy compared to real-time PCR (RT-PCR) assay [94]. Pardee et al., 2016, used novel CRISPR/Cas9 to discriminate Zika viral strains with single-base resolution [95]. Nguyen et al., 2020, developed CRISPR-ENHANCE (ENHanced Analysis of Nucleic acids with CrRNA Extensions). This technology combines CRISPR-Cas12a with engineered crRNAs and DNA extensions for highly sensitive and specific detection of nucleic acids. They optimized existing CRISPR-based assays and utilized ENHANCE to detect clinically relevant targets such as PCA3, HIV, HCV, and SARS-CoV-2. By incorporating an isothermal amplification step, the ENHANCE system exhibited improved detection of SARS-CoV-2 genomic RNA, surpassing the performance of the wild-type CRISPR-Cas12a system in both fluorescence-based and paper-based lateral flow assays. The findings highlight the potential of ENHANCE in enhancing nucleic acid detection and its applicability for sensitive diagnostics [96]. Miao et al., 2023, highlighted the effectiveness of a one-pot RPA-CRISPR/Cas13a assay for rapid NiV (Nipah Virus) detection, offering cost-efficiency, quick results, and minimal contamination risks. While it may have a higher limit of detection than qRT-PCR, it complements existing diagnostic methods, aiding early NiV infection detection, especially in resource-limited regions. This assay signifies a significant advancement in NiV diagnostics for healthcare systems with limited infrastructure. These breakthroughs show-

case the potential of CRISPR-based approaches to transform disease diagnosis, offering sensitive and rapid nucleic acid screening for viral infections [97].

There are limitations of CRISPR-based approaches, such as the need for specialized expertise in protein purification and RNA biology, the lack of commercially available assays, limitations in precise target quantification, dependence on known DNA sequences, and challenges in achieving high sensitivity levels. The extraction method varies depending on the sample origin, adding complexity to the process. Eliminating the need for nucleic acid extraction while maintaining accuracy remains a challenge. Furthermore, the availability and manufacturing of CRISPR proteins, such as Cas13 and Cas14, present obstacles in developing CRISPR-based testing. Currently, labs express and purify these proteins, leading to variations between laboratories. Widespread manufacturing of CRISPR proteins using good manufacturing practice (GMP) standards would accelerate the development and availability of CRISPR-based detection assays [98].

Table 3. CRISPR-Cas diagnostics for emergent pathogen.

Pathogen	Sample Type	Detection Technique	Sensitivity	Specificity	References
Helicobacter pylori	Stool	CRISPR-Cas12, a system-based method			[99]
Zika	Plasma of a viremic macaque.	NSBA CRISPR-Cas9			[91]
SARS-CoV-2	Nasopharyngeal swab samples	Isotachophoresis (ITP)-enhanced CRISPR-Cas12a			[100]
Pneumocystis jirovecii	Bronchial alveolar lavage fluid	Transcription-mediated amplification/ CRISPR-Cas13a/ (Fluorescence plate reader)	78.9%	97.7%	[101]
SARS-CoV-2	Nasopharyngeal swabs	Loop-mediated Isothermal Amplification (LAMP)/SHERLOCK/(lateral flow)	-	100%	[102]
Zika and Dengue	Human Serum/Saliva/Urine	RT-RPA–HUDSON/ Cas13-based SHERLOCK/ (Fluorescence/calorimetric)	-	100%	[103]
Mycobacterium tuberculosis	Clinical sputum samples	LACD (loop-mediated isothermal amplification)/CRISPR-Cas12a/ (Later flow/real-time fluorescence)	~10 copies/ test	100%	[104]
Mycobacterium tuberculosis	Sputum	RPA/CRISPR-Cas12a/(Fluorescent detection)	79%	98%	[105]
Hepatitis B virus	Plasma	RT-LAMP/CRISPR–Cas12a/ (Lateral flow strips or fluorescence detector)	96%	100%	[106]
Monkey pox	Synthetically produced Congo Basin clade D14L and West African clade ATI and cloned into the pUC57 vector	Loop-mediated isothermal amplification (LAMP)/CRISPR-Cas12b/ real-time fluorescence and a gold nanoparticle-based lateral flow biosensor (AuNP-LFB)	10 copies/ reaction	-	[107]
Ebola virus	Urine, Saliva	HUDSON/CRISPR-Cas13a-based (SHERLOCK)/ (Fluorescent and lateral flow readouts)	91%	100%	[108]
Influenza (H1N1)	Synthetic DNA strands	CRISPR/Cas13a/hybridization chain reaction (HCR)/ Colorimetric biosensor	0.152 pM	-	[109]

3.2. PCR-Based Methods

PCR-based methods have been crucial for emergent pathogen surveillance, providing fast and accurate diagnosis of viral infections. It offers several advantages in detecting, quantifying, and typing DNA or RNA sequences in various matrices. Its key advantages include fast and high-throughput analysis, simultaneous amplification and visualization of amplicons, avoidance of cross-contamination, wide dynamic range for quantification, and the ability to multiplex amplification of multiple targets in a single reaction [110,111]. During the 2009 H1N1 influenza pandemic, real-time PCR was used to detect and quantify viral RNA in clinical samples, enabling timely diagnosis and outbreak monitoring [112]. Similarly, in response to the current COVID-19 pandemic caused by the novel coronavirus SARS-CoV-2, PCR-based assays have been developed and widely used for diagnosis and surveillance. These assays target specific viral genome regions, such as the N and E gene variants, which allow for sensitive and specific virus detection [113,114]. PCR-based surveillance methods have also been used for other emergent viral pathogens Ebola, Zika Dengue, and HIV. For example, in the 2014–2016 Ebola outbreak in West Africa, real-time PCR was used to diagnose and track the virus rapidly [115].

Similarly, during the 2015–2016 Zika virus outbreak, real-time PCR was used to detect the virus in clinical samples, aiding in the timely diagnosis and control of the outbreak [116]. In the 2003 Ebola outbreak (Zaïre ebolavirus) in Mbomo, the Ministry of Health of the Republic of Congo, in collaboration with partners, established a small field laboratory under the leadership of the WHO. This laboratory used antigen capture and qRT-PCR techniques to diagnose acute cases of Ebola [117] accurately. Seki et al., 2015, conducted active screening using PCR for surveillance of methicillin-resistant *Staphylococcus aureus* (MRSA) and compared the backgrounds of PCR and/or culture method-positive patients [118]. Kim and Hwang, 2020 [119] established a real-time RT-PCR method for simultaneously detecting and typing all four serotypes of Dengue virus (DENV). The method demonstrated high sensitivity, with 5–10 copies/reaction detection limits for each serotype. The developed method holds enormous potential for accurately assessing the circulating DENV serotypes, predicting disease severity, and efficiently analyzing seroprevalence, all of which contribute to a better understanding of DENV epidemiology and effective disease management.

PCR-based surveillance methods have been utilized for tick-borne bacterial pathogens like *Borrelia burgdorferi* (causing Lyme disease) and other bacterial pathogens such as *Escherichia coli*, *Salmonella*, and *Staphylococcus aureus*, aiding in detection and monitoring of outbreaks and antimicrobial resistance. Tokarz et al. (2017) [120] developed a multiplex one-step RT-PCR assay to detect major pathogens transmitted by Ixodes scapularis ticks. The assay successfully identified *Anaplasma phagocytophilum*, *Babesia microti*, *Borrelia miyamotoi*, *Borrelia burgdorferi*, and *Powassan virus* in ticks from the northeastern U.S. *B. burgdorferi* was the most frequently detected agent in ticks. The study emphasized the need for monitoring co-infections and their impact on public health. Li et al., 2017 [121], developed a multiplex real-time PCR assay to detect *E. coli* O157:H7 and screen for non-O157 STEC strains. The assay targeted the Z3276 gene and Shiga toxin genes (stx1 and stx2). It exhibited high sensitivity and specificity, detecting all *E. coli* O157:H7 and STEC strains tested without cross-reactivity with other pathogens. The assay's limit of detection was 200 femtograms of bacterial DNA. Simultaneous detection of *E. coli* O157:H7, *Staphylococcus aureus*, and *Salmonella* was developed by practical multiplex PCR technique, making it valuable for microbial epidemiology and food safety investigations [122].

Yadav et al. (2022) employed real-time RT-PCR and ELISA to confirm Nipah virus (NiV) infection, identifying the NiV genotype and its close relation to Indian sequences. They discovered NiV presence in local bat populations, with 21% of *Pteropus medius* and 37.73% of *Rousettus leschenaultia* testing positive for NiV antibodies, underscoring the need for stringent surveillance and awareness campaigns to mitigate human–bat interactions and reduce the risk of NiV outbreaks. There was a successful containment of a NiV outbreak in Kerala, India, through prompt diagnosis, contact tracing, and a better understanding of

the virus's presence in local bat populations. This study highlights the need for ongoing vigilance and preventive measures to reduce the risk of future outbreaks [123].

These results suggest that PCR-based active surveillance is more rapid and sensitive than conventional culture methods and will continue to play a critical role in global health security. Some recent examples of PCR-based methods for rapid diagnosis and tracking of viral outbreaks are given in Table 4.

Table 4. PCR-based methods for various emergent pathogens detection.

Pathogen	Sample Type	Method	Sensitivity	Specificity	References
SARS-CoV-2	Sputum as well as nose and throat swabs	Real-time RT-PCR	95%	-	[124]
Zika	Serum	RealStar ZIKV rRT-PCR test kit	91%	97%	[125]
Zaire Ebolavirus (ZEBOV)	Cell lines	TaqMan RT-PCR	109 copies to 103 copies/reaction	-	[117]
Monkeypox	Lesion swabs	Non-variola orthopoxvirus PCR test	100 copies/mL and 100% agreement.	100%	[126]
Candida auris	Axilla-groin composite surveillance swabs	SYBR green qPCR	0.93	0.96	[127]
Influenza C virus (FLUCV)	Nasopharyngeal samples (nasal and oropharyngeal swabs)	Multiplex RT-PCR	-	-	[128]
Noroviruses (NoV)	Stool specimens	TaqMan RT-PCR assay	<10 copies of viral genome per reaction.	-	[129]
African swine fever virus (ASFV)	EDTA blood and serum samples from pig	Real-time PCR/UPL PCR	4–8 DNA copies	10-fold high for the different ASFV isolates tested, representing p72 genotypes I (the one mostly distributed in Sardinia and West Africa), VIII (the most divergent p72 genotype) and IX (representative of East Africa), in comparison with the OIE reference TaqMan PCR	[130]
MERS CoV	Environmental samples (air and surface swab)	RT-PCR	-	-	[131]

PCR-based methods have limitations, including specific temperatures, time-consuming procedures, sophisticated equipment, and a lack of standardized protocols, and they depend on well-equipped laboratories and skilled technicians. The complexity and infrastructure requirements make conducting PCR testing in remote or resource-limited areas challenging, often necessitating samples being sent to centralized facilities [101]. While syndromic PCR panels exist for respiratory pathogens, they have limitations in addressing emerging threats and require regular updates and validation. Moreover, the COVID-19 pandemic has highlighted the challenges of redesigning PCR-based assays, especially considering the virus's continued mutagenesis [132]. However, advancements in diagnostic methods have paved the way for overcoming these limitations. Anderson and Malderlii, 2018 [133], presented a method for quantifying total HIV DNA using (ddPCR). The data obtained from this protocol helped determine the viral DNA burden and enabled examining changes in different proviruses over time. The ddPCR (droplet digital PCR) platform allows for absolute quantification, making it possible to observe small changes accurately. Overall, this research suggests that ddPCR holds promise as an alternative to bulk qPCR, especially in scenarios requiring quantification of low viral loads without extensive sample processing. However, challenges related to scalability and instrumentation need to be addressed for wider clinical adoption. Fowler et al., 2021, introduced a rapid and versatile reverse-transcription loop-mediated isothermal amplification (RT-LAMP) assay,

effective in detecting high SARS-CoV-2 viral loads directly from swabs, bypassing RNA extraction bottlenecks. It underscored the importance of considering viral load for result interpretation and highlighted the potential for near-patient testing in diverse healthcare settings [134].

3.3. Microarray Analysis

In addition to DNA sequencing and polymerase chain reaction (PCR), oligonucleotide microarray technology is one of the most robust techniques for detecting and exploring microorganisms utilizing nucleic acid samples [135]. While DNA sequencing offers unparalleled comprehensive insights and unbiased information, enabling the identification of entirely novel organisms as previously elucidated, its implementation can be economically and temporally demanding for specific applications, particularly when factoring in the resources essential for data processing and analysis. Conversely, PCR assays at the opposite end of the cost spectrum demonstrate rapidity and sensitivity, albeit with restricted capacity for multiplexing (simultaneous detection of multiple targets). In instances necessitating the concurrent assessment of numerous organisms, the requirement for numerous PCR reactions can nullify the anticipated cost advantages.

Oligonucleotide microarrays occupy an intermediary position concerning cost, processing duration, sensitivity, specificity, and the capability to detect emerging organisms. Notably, most microarray designs derive their probes from fully sequenced genomes housed within databases [136]. Despite this, the phenomenon of cross-hybridization between probes and closely related yet non-identical sequences enables the detection of novel species, provided their genetic makeup bears proximity to those organisms employed for probe development.

Commencing a pivotal opportunity in pathogen detection, the ViroChip emerged as the foremost microarray tailored to discern a comprehensive spectrum of pathogens [137]. The inaugural version of ViroChip encompassed 1600 probes, extrapolated from 140 intact viral genomes cataloged in GenBank at the array's inception. Subsequent versions of the platform were ingeniously devised to encompass an expanding panorama of viruses, synchronously with the proliferation of published genomes [138]. Fabricated through the precision of mechanized spotting, ViroChip materialized as an ensemble of synthesized oligonucleotides intricately arrayed upon a glass substrate. Constituting 70-mer oligomers, these elements were carefully curated to align with sequences against a specific taxonomic family, demarcating their distinctiveness from other lineages. By its design predicated upon conserved motifs, ViroChip emerged as a potent instrument in identifying nascent viruses within lineages, coextensive with known and sequenced viral counterparts. Notably, this application was utilized in the characterization of the 2003 SARS outbreak-inciting SARS-CoV variant [139].

Another novel methodology for pathogen detection involving microarray, termed resequencing microarrays, was employed in various studies [140,141]. These microarrays featured concise 25- or 29-mer probes strategically positioned along specific genes within the target pathogenic species. For each genomic locus in the target gene, four probes were meticulously designed: one exhibiting a perfectly matched base at the probe's central position, and three others corresponding to each of the alternative bases. Subsequent hybridization and analysis of these microarrays culminated in the determination of sequences corresponding to each homologous target gene present within the specimen. Through sequence comparison utilizing a basic alignment search tool (BLAST) [142] against a comprehensive database, the ascertained sequences were correlated with distinct species and strains. The first edition of this approach, the Respiratory Pathogen Microarray (RPM v1.1), was expertly crafted as a tailored Affymetrix array [140]. It encompassed probes targeting an array of prevalent human respiratory viruses and bacteria, encompassing influenza, adenovirus, coronavirus, rhinoviruses, as well as bacterial species such as *Bordetella pertussis* and *Streptococcus pneumoniae*. Further evolutionary iterations of this strategy encompassed the integration of virulent genes from tropical infectious agents, alongside its application

for the high-throughput molecular diagnosis of diverse pathogens linked to central nervous system infections [143].

Furthermore, a similar technology, called GreeneChip arrays, emerged as a more expansive and versatile approach in pathogen detection [144,145]. The high-density arrays utilized cutting-edge Agilent inketsystem for the genome profiling. The initial version of GreeneChip encompassed a total of 9477 probes to target viral targets. The latest versions now incorporate a plethora of probes meticulously curated to identify myriad pathogenic bacteria, fungi, and protozoa, culminating in a staggering compilation, totaling 29,495 60-mer oligonucleotides. As technology advances and databases expand, these microarray-based approaches continue to evolve, enhancing our capacity to tackle infectious disease challenges efficiently and with increasing precision.

3.4. Bioinformatics Tools

The use of bioinformatics tools to analyze the growing data from fields like molecular biology, genomics, transcriptomics, and proteomics is getting has achieved significant success [146]. In combination with bioinformatics, the utilization of NGS technology has evolved into a formidable instrument for characterizing and scrutinizing human pathogens. This amalgamation presents a potent approach for enhancing the precision of screening for resistance mutations/genes, identifying vaccine-evading variants, probing virulence and pathogenicity attributes [147–149]. Some of the currently available bioinformatics tools offering comprehensive sequence assembly and analysis are Lasergene, Geneious, Mauve, DECIPHER, ChimeraSlayer, and mothur etc. [150–155].

The proliferation of genomic data, arising from numerous advanced technologies, has brought about a need for sophisticated solutions in data storage, along with the utilization of bioinformatics tools, pipelines, and analyses, owing to the intricate nature of these datasets. Several prominent entities are available within the realm of microbial pathogen genomics resources. First, the NCBI pathogens database commands significance due to its affiliation with the NCBI pathogen detection initiative [156]. This repository database encompasses genomic sequence data from pathogenic bacteria and fungi, sourced from clinical-, environmental-, and food-related contexts. A staggering compilation of over a million isolates spanning 49 bacterial groups, comprising 33 genera, is housed within this repository. Impressively, 1.18% (12,573) of these isolates are meticulously annotated as complete genomes. A second notable resource is the Pathosystems Resource Integration Centre (PATRIC) [157]. This repository comprises 548,000 bacterial genomes, largely drawn from the NCBI GenBank and RefSeq databases. Nearly 7% (around 36,000) of this total represents complete genomes. Of particular interest are roughly 5800 genomes that are designated as "High Quality", presenting associated disease information. This subset includes notable infections such as MRSA-positive infections (methicillin-resistant *S. aureus*) with approximately 780 genomes, shigellosis with around 580 genomes, and typhoid fever with roughly 530 genomes. A third important resource in this domain is the Pathogenwatch initiative at the Wellcome Trust Sanger Institute "https://www.sanger.ac.uk/tool/pathogenwatch/ (accessed on 15 August 2023)". This platform boasts the capability to rapidly process genomes, enabling the execution of tasks like multi-locus sequence typing (MLST), the identification of genes and single-nucleotide polymorphisms (SNPs) implicated in antimicrobial resistance, and the inference of susceptibility to antibiotics. Notably, Pathogenwatch also furnishes information about the closest phylogenetic neighbors and their geographic origins when such data are available. Several other comprehensive reference databases have also been developed to facilitate accurate microbial pathogen identification. The Greengenes database contains 1,049,116.

In addition to established reference tools, various programs have emerged to unveil the underlying pathogenic nature of a given test pathogen. One prominent bioinformatics solution for evaluating the pathogenic potential of newly identified bacterial pathogens is the PathogenFinder 1.1 [158]. This web-based tool serves as a predictive resource for bacterial pathogenicity assessment, leveraging proteomic, genomic, or raw sequencing data.

The pathogenicity of bacterial strains hinges on specific clusters of proteins recognized for their involvement in pathogenic processes. PathogenFinder employs a diverse set of proteins deliberately selected due to their lack of annotated functionality or established pathogenic roles. Remarkably, the tool demonstrates an impressive 88.6% accuracy in predicting pathogenicity across the spectrum of bacterial taxonomic groups. The program holds promise in uncovering novel factors contributing to pathogenicity.

A more recent advancement in pathogenicity prediction comes in the form of PaPrBaG (Pathogenicity Prediction for Bacterial Genomes), accessible through its GitHub repository [159]. Built upon machine learning principles, PaPrBaG is encapsulated within an R package. The methodology behind PaPrBaG centers on extensive training using a substantial collection of confirmed pathogenic species juxtaposed with nonpathogenic counterparts. PaPrBaG is particularly well-suited for handling Next-Generation Sequencing (NGS) data characterized by limited genomic coverage.

New sequencing technologies and computer methods have helped pathogen genomics progress quickly. These tools handle the challenges of lots of genomic data, making research in microbial pathogen genomics effective. But, like any fast-growing area, there are problems and chances to deal with them. How quickly bacterial pathogen genomics transforms forensics, food safety, and clinical microbiology in different countries depends on new technology breakthroughs.

4. Advances in Genomic Epidemiology and Pathogen Surveillance

As previously addressed, genomics emerges as a potent tool not only for the detection of pathogens but also for the progression of our comprehension regarding epidemiology. Throughout recent decades, the sequencing of pathogens has held a pivotal role in comprehending the patterns of transmission during viral outbreaks. Nevertheless, most investigations with a focus on transmission have been retrospective in nature, with only a fraction occurring in real time, coinciding with the diagnosis of cases [160–163]. In transmission-oriented studies, genetic variations come into play to identify instances of person-to-person transmission like happened during the recent COVID outbreak (WHO GLOBL SURVEILLANCE REFERENCE). This identification process can involve either a manual analysis of shared variants among cases within an outbreak [14], or the adoption of model-based methodologies [164], culminating in the formation of a transmission network. Contrarily, when it comes to epidemic investigations, only a subset of cases within the epidemic undergo sequencing. As such, the advancement of pathogen surveillance aims to utilize the population structure of the pathogen as a means to grasp the overall dynamics of the epidemic.

Moreover, when considering the deployment of genomics for the purposes of surveillance, diagnostics, and epidemiological inquiries, a pivotal query emerges: where should these efforts be concentrated? This question is especially significant given that numerous regions lack the necessary diagnostic laboratory capacity to conduct even fundamental surveillance activities. Maintaining continuous genomic surveillance across all these diverse settings is an impractical endeavor. Various initiatives have endeavored to outline the reservoirs of geographical focal points and potential pathogenic sources that could give rise to future epidemics or pandemics. Woolhouse et al. meticulously documented a total of 1399 human pathogens, among which 87—primarily of viral origin—have emerged since the year 1980 [165]. Extending this perspective, Jones et al. have broadened the scope to encompass 335 new instances of EIDs since the year 1940 [1]. Their analysis reveals a progressive increase in the frequency of such events over successive decades, predominantly clustered in geographic hot spots characterized by specific environmental, ecological, and socio-economic attributes. The identification of these crucial elements plays a central role in forecasting and averting spillover incidents. A notable observation is the zoonotic origin of the majority of these EIDs, with regions harboring high biodiversity and undergoing recent demographic shifts and/or increased agricultural activity presenting the highest risk of spillover [1]. Supporting this notion, a global biogeographic assessment of human

infectious diseases reinforces the utility of biodiversity as an indicator for identifying EID hot spots [166].

To address these challenges, one of the approaches being utilized on a higher level is the One Health approach, adopted by the United States, which is a comprehensive strategy that has been put into practice through initiatives like the PREDICT project [167]. This project is an integral component of the Emerging Pandemic Threats (EPT) program established by the U.S. Agency for International Development (USAID). The core focus of PREDICT revolves around investigating the potential transmission of select viral zoonoses originating from specific wildlife taxa. Early endeavors have concentrated on devising non-invasive sampling methodologies for wildlife [168], estimating the extent of viral diversity within nine viral families and a minimum of 320,000 previously undiscovered species within the mammalian realm [169], and demonstrating that the diversity of viral communities is influenced, at least in part, by deterministic factors. This observation implies that predicting alterations in community dynamics—potentially signaling an impending spillover event—is within the realm of possibility [170].

Nevertheless, despite notable progress in the previous decades, recommendations from various expert collectives have consistently underscored the imperative for enhancing surveillance capabilities [171]. This encompasses the implementation of syndromic surveillance—an approach less focused on specific pathogens and more centered on the prompt identification of emerging diseases [172,173]. Syndromic surveillance systems have the potential to harness distinct data sources, such as instances of student or worker absenteeism, purchases of certain items at grocery stores or medications from pharmacies, and calls to clinic/hospital hotlines. These indicators can serve as markers of illness within a population. This novel surveillance paradigm recognized as digital epidemiology is also termed digital disease detection [174]. In the realm of digital epidemiology, data are initially aggregated from diverse origins, including digital media, news sources, official releases, and crowd-sourced inputs. Subsequently, this information is translated, processed—encompassing the extraction of disease-related events and the elimination of duplicated reports—analyzed for trends, and finally disseminated to the public through various media channels such as websites, email distributions, and mobile alerts [175–177]. Presently, there are over 50 operational digital epidemiology platforms [178]. Their adaptable characteristics and cost-efficient, real-time reporting render them potent tools for acquiring epidemic intelligence, especially in regions lacking conventional disease surveillance systems.

The integration of this digital pathogen surveillance system with initiatives like PREDICT holds immense potential in the field of epidemiology. This system could bring about significant changes, especially in regions where laboratory and surveillance capabilities are traditionally limited. Regular surveillance would involve the pooling of samples, achieved through targeted diagnostic methods or potentially metagenomics, provided that the challenge of analytical sensitivity can be addressed. A potential pathogenic signal arises, and individual samples would undergo comprehensive genomic analysis. Simultaneously, pre-existing online platforms like HealthMap [179] and emerging local participatory epidemiology endeavors would gather data to identify potential hotspot regions and detect EID events. This collective data accumulation would enable proactive and swift deployment of supplementary sequencers. The resultant genomic sequencing data, accompanied by comprehensive metadata, would be instantly shared on web-based platforms such as Virological "http://www.virological.org (accessed on 17 August 2023)", fostering collaborative analysis. Additionally, platforms like Nextstrain [180], successfully utilized in responses to Ebola and Zika outbreaks, would facilitate analysis and visualization.

Furthermore, Artificial Intelligence (AI), often associated with machine learning, rests on the fundamental premise that human thought and reasoning can be mechanized. Remarkably, the COVID-19 pandemic showcased the pivotal role of AI in various aspects, such as genome sequencing, drug and vaccine development, disease outbreak detection, disease spread monitoring, and the tracking of viral variants [181]. In 2022, Sundermann and a

team of researchers introduced an innovative solution known as the Enhanced Detection System for Healthcare-Associated Transmission (EDS-HAT) [182]. EDS-HAT seamlessly combines the power of infection surveillance through WGS with the intelligent analysis of patient electronic health records (EHR) data. This synergy enables the identification of disease outbreaks and transmission pathways that might otherwise elude detection.

Over two years spanning from November 2016 to November 2018, this novel AI-driven approach was applied at nine hospitals affiliated with the University of Pittsburgh. The outcome was the identification of healthcare-associated bacterial pathogens, thereby preventing multiple outbreaks and tracing the transmission routes. Notably, this AI-assisted learning underscores the utilization of electronic medical records to steer clinical algorithms that effectively link infections [182,183].

These platforms would act as the central hub for a global network of stakeholders keen on contributing to real-time phylodynamic and epidemiological investigations. The collective effort would focus on identifying signals of pathogen spillover, expansion of pathogen populations, and sustained human-to-human transmission. Outcomes from these analyses would be promptly disseminated to the forefront of projects like One Health, comprising epidemiologists, veterinarians, and community health workers. Armed with evidence-based insights, these professionals would then implement interventions to curtail the further spread of the diseases effectively.

5. Challenges and Future Directions

As genomics becomes integrated into clinical and public health practices, the accessibility of genomic surveillance data faces several obstacles. Such issues include fragmented metadata across multiple databases, limited interoperability, insufficient workforce expertise, privacy concerns, variable data quality in resource-limited settings, third-party data reuse concerns, and the absence of standardized frameworks to capture multisite data [184]. Addressing these issues, such as establishing data-sharing networks and promoting research transparency, is crucial for effective pandemic preparedness. This section explores these challenges and provides insights into future perspectives.

5.1. Lack of Trained Workforce and Networking Infrastructure

The lack of a trained workforce hinders the effective utilization of cutting-edge research and technologies. Additionally, inadequate networking infrastructure impedes data sharing and collaboration critical for advancements in these fields.

5.1.1. Shortage of Skilled Genomic Experts

The successful implementation of genomics in pandemic situations requires a proficient workforce with expertise in genomic technologies, bioinformatics, and data analysis. Unfortunately, there is a global shortage of trained professionals in these fields [185]. The scarcity of skilled personnel hampers the timely and accurate analysis of genomic data, hindering the rapid identification and characterization of emerging pathogens.

5.1.2. Training Gaps

The dynamic nature of genomics necessitates continuous training and skill development. However, the existing educational curricula often lack comprehensive genomic training programs tailored to infectious disease surveillance and outbreak response [186]. As a result, many healthcare professionals and researchers may not possess the necessary competencies to utilize genomic technologies in pandemic situations effectively.

5.1.3. Data Sharing Challenges

The success of genomics-based approaches heavily relies on seamless data sharing among researchers, public health agencies, and international collaborators. However, challenges such as data privacy concerns, regulatory hurdles, and lack of standardized protocols hinder the effective sharing of genomic data across borders [187]. The absence

of a robust networking infrastructure limits the timely exchange of critical information, hindering global surveillance and response efforts during a pandemic.

5.1.4. Limited Collaboration Platforms

Genomics research often involves interdisciplinary collaboration, necessitating efficient networking platforms for knowledge exchange and resource sharing. Unfortunately, the lack of centralized platforms and coordination mechanisms hinders effective collaboration among researchers, clinicians, and policymakers [188]. This fragmentation of efforts limits the real-time application of genomics in pandemic response.

5.2. Technical Considerations of Genomics in Pathogen Detection and Tracking

The application of genomics to pathogen detection and tracking faces several technical challenges and limitations. Some of these include:

- Time-consuming and resource-intensive sequencing procedures, particularly for complex microbial communities, such as those found in the gut or soil [148,189].
- Low sensitivity of sequencing techniques in detecting low abundance or low copy number pathogens, such as those found in cerebrospinal fluid [190].
- Genomic sequencing generates vast amounts of data, which can overwhelm computational resources and expertise. Analyzing and interpreting this data require advanced bioinformatics tools and skilled personnel. This can lead to delays in obtaining actionable results, especially in resource-limited settings.
- Data analysis and interpretation complexity, including the need for bioinformatics expertise and computational resources.
- Technical variability and errors in sequencing data, particularly in the presence of genetic polymorphisms, genomic rearrangements, or repetitive regions [191].
- The accuracy of genomics-based pathogen identification relies on the quality and representativeness of the sample collected. Biases can be introduced if the sampling process is not well designed or if the pathogen is in low quantities. Additionally, the genomic material of interest might be mixed with host DNA, affecting the quality of the sequencing data.
- Pathogens can rapidly evolve through mutation and recombination, leading to genetic diversity within a single species. This diversity can make it challenging to design universal genomic markers for identification. Furthermore, the identification of novel strains or variants might require frequent updates to reference databases.
- Genomic sequencing often requires fresh or well-preserved samples. The logistics of storing and transporting samples to sequencing facilities without compromising their integrity can be challenging, particularly in remote or disaster-affected areas.
- While the cost of genomic sequencing has decreased significantly over the years, it can still be expensive, especially for large-scale surveillance or in low-resource environments. The cost of equipment, reagents, and skilled personnel can be a significant barrier to widespread adoption. Emerging initiatives, such as the Human Heredity and Health in Africa (H3Africa), the Qatar Genome Project, and the Mexico National Institute of Genomic Medicine (INMEGEN), are playing a pivotal role in bolstering the genomic research capabilities of Low- and Middle-Income Countries (LMICs). They achieve this by providing funding for locally driven research endeavors and empowering indigenous researchers to assume leadership roles in genomics projects [192]. Noteworthy accomplishments in this domain are exemplified by projects like the African Genome Variation Project [193] and the Mexico Genomic Variation Project [194]. These initiatives are dedicated to elucidating the intricate genetic structures within diverse ethnic groups, aiming to advance genomic medicine in Africa and Mexico [194].
- Effectively translating the genome sequence into actionable medical insights presents a significant hurdle. One major challenge is accurately anticipating the functional impact of genetic variations that disrupt protein-coding sequences. These variations

can manifest in various ways, such as affecting transcription factor binding sites, interfering with microRNA target sites, influencing RNA splicing and stability, or even leading to protein truncation. Moreover, the intricacy of linkage disequilibrium, where seemingly benign genetic variations are situated near disease-predisposing variants, further complicates the interpretation of recurrent risk factors. Considering these complexities, there is a growing reliance on the use of in silico tools for inferring the functional consequences of mutations. As mentioned in the aforementioned sections, computational algorithms can play a crucial role in predicting the potential pathogenicity of genetic variants.

Addressing these challenges will require developing new sequencing technologies, improved data analysis methods, and a better understanding of microbial ecology and evolution. For example, single-cell or nanopore sequencing advances can increase sequencing sensitivity and accuracy, whereas machine learning algorithms and cloud-based computing can streamline data analysis and interpretation [195,196].

5.3. Ethical and Legal Considerations of Genomic Data Sharing and Privacy

Using genomics in pathogen detection and tracking raises essential ethical and legal considerations related to data sharing and privacy. In particular, the potential for the misuse or misinterpretation of genomic data, the risk of re-identification of individuals, and the issues of informed consent and data ownership are essential concerns that must be addressed [197]. HIV phylogenetic data are inherently relational, as information from one individual can affect others by identifying them as possible sources of infection. Additionally, the increasing richness of sequence data resulting from next-generation sequencing (NGS) presents challenges for maintaining the anonymity of viral sequence data. Genome sequencing also presents risks to communities and groups. Numerous authors have highlighted the potential for certain groups to be unfairly stigmatized as high-risk or disease carriers, as observed in cases involving geographically defined populations, sexual or gender minorities, and individuals categorized by factors such as ethnicity, nationality, or migration status [198]. This extends to situations where data on the transmission patterns of diseases such as multidrug-resistant tuberculosis (TB) could be misused to discriminate against specific ethnic groups, potentially leading to challenges related to immigration [199]. Reidentifying an individual utilizing a virus isolated from a different time-point or study is possible. Concerns have also arisen regarding the imposition of mandatory testing on certain groups, such as healthcare workers [200,201]. It is important to acknowledge the individual professional interests of researchers and practitioners in terms of data ownership and utilization [202]. Consequently, it is vital to establish ethical frameworks to address these issues and ensure responsible handling of such data.

The guidelines and regulations have been developed to ensure the responsible and ethical use of genomic data, such as the Global Alliance for Genomics and Health (GA4GH) and the European General Data Protection Regulation (GDPR) [203,204]. These guidelines emphasize the importance of data privacy, informed consent, and transparency in genomic research and provide frameworks for data sharing and collaboration among researchers and stakeholders.

5.4. Integration of Genomics with Other Surveillance and Diagnostic Methods

Integrating genomics with other surveillance and diagnostic methods is essential for improving pathogen detection and tracking. For example, combining genomics with epidemiological data, such as patient travel history or contact tracing, can help identify sources of infection and potential transmission routes [205]. In addition, integrating genomics with traditional microbiological methods, such as culturing or antigen detection, can enhance the sensitivity and specificity of pathogen detection [149].

5.5. Innovations and Opportunities for Improving Genomics-Based Pathogen Detection and Tracking

The rapid development of portable and robust sequencing technologies that can be deployed in field settings or resource-limited areas should be explored. Also, efforts should be made to integrate genomics with host immune profiling, such as transcriptomics or proteomics, to understand host–pathogen interactions better and identify novel biomarkers for diagnosis or treatment [206,207].

Overall, the challenges and future directions in genomics for pathogen detection and tracking are complex and multifaceted. Addressing these challenges will require interdisciplinary collaborations, innovative approaches, and responsible use of genomic data to improve public health outcomes.

6. Conclusions

In conclusion, genomics has transformed the way we understand and investigate infectious diseases, especially in the context of outbreak investigations and surveillance. The use of high-resolution molecular data provided by genomic sequencing has enabled the identification of closely related strains and tracing the source and transmission of infectious diseases. However, technical challenges and limitations still need to be addressed, such as resource-intensive sequencing procedures, low sensitivity in detecting low-abundance pathogens, and data analysis and interpretation complexity. Additionally, ethical and legal considerations concerning genomic data sharing and privacy must be streamlined. Advances in sequencing technologies and data analysis methods, coupled with an improved understanding of microbial ecology and the establishment of multi-disciplinary collaborations, can help resolve these challenges and enhance the use of genomics in pathogen detection and tracking in the foreseeable future.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/biomedinformatics3040069/s1, Figure S1: Alignment of SARS-CoV-2 major proteins (a) Mpro, (b) RdRp, (c) spike proteins, (d) envelope proteins, and (e) nucleocapsid proteins from different countries {Reprinted with permission Khan et al., 2020 [71].

Author Contributions: Conceptualization, V.V. and R.K.; methodology, V.V. and A.V.; software, V.V., A.V. and A.K.M.; validation, V.V. and A.V.; formal analysis, V.V; investigation, V.V.; resources, R.K.; data curation, V.V., A.K.M., A.V., H.S., A.A., P.A. and A.S.; writing—original draft preparation, V.V. and A.V.; writing—review and editing, J.F. and R.K.; visualization, V.V.; supervision, R.K.; project administration, R.K.; funding acquisition, R.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Jones, K.E.; Patel, N.G.; Levy, M.A.; Storeygard, A.; Balk, D.; Gittleman, J.L.; Daszak, P. Global trends in emerging infectious diseases. *Nature* **2008**, *451*, 990–993. [CrossRef]
2. Grubaugh, N.D.; Ladner, J.T.; Lemey, P.; Pybus, O.G.; Rambaut, A.; Holmes, E.C.; Andersen, K.G. Tracking virus outbreaks in the twenty-first century. *Nat. Microbiol.* **2019**, *4*, 10–19. [CrossRef] [PubMed]
3. Firth, C.; Lipkin, W.I. The genomics of emerging pathogens. *Annu. Rev. Genom. Hum. Genet.* **2013**, *14*, 281–300. [CrossRef] [PubMed]
4. Hui, E.K.W. Reasons for the increase in emerging and re-emerging viral infectious diseases. *Microbes Infect.* **2006**, *8*, 905–916. [CrossRef]
5. Parrish, C.R.; Holmes, E.C.; Morens, D.M.; Park, E.C.; Burke, D.S.; Calisher, C.H.; Laughlin, C.A.; Saif, L.J.; Daszak, P. Cross-species virus transmission and the emergence of new epidemic diseases. *Microbiol. Mol. Biol. Rev.* **2008**, *72*, 457–470. [CrossRef] [PubMed]
6. Patz, J.A.; Reisen, W.K. Immunology, climate change and vector-borne diseases. *Trends Immunol.* **2001**, *22*, 171–172. [CrossRef]
7. Bengis, R.G.; Leighton, F.A.; Fischer, J.R.; Artois, M.; Morner, T.; Tate, C.M. The role of wildlife in emerging and re-emerging zoonoses. *Rev. Sci. Tech.-Off. Int. Epizoot.* **2004**, *23*, 497–512.
8. Caminade, C.; McIntyre, K.M.; Jones, A.E. Impact of recent and future climate change on vector-borne diseases. *Ann. N. Y. Acad. Sci.* **2019**, *1436*, 157–173. [CrossRef]

9. Kilpatrick, A.M.; Randolph, S.E. Drivers, dynamics, and control of emerging vector-borne zoonotic diseases. *Lancet* **2012**, *380*, 1946–1955. [CrossRef]
10. Morens, D.M.; Folkers, G.K.; Fauci, A.S. The challenge of emerging and re-emerging infectious diseases. *Nature* **2004**, *430*, 242–249. [CrossRef]
11. Cutler, S.J.; Fooks, A.R.; Van Der Poel, W.H. Public health threat of new, reemerging, and neglected zoonoses in the industrialized world. *Emerg. Infect. Dis.* **2010**, *16*, 1–7. [CrossRef] [PubMed]
12. Watson, J.T.; Gayer, M.; Connolly, M.A. Epidemics after natural disasters. *Emerg. Infect. Dis.* **2007**, *13*, 1–5. [CrossRef] [PubMed]
13. Allen, J.D.; Ross, T.M. H3N2 influenza viruses in humans: Viral mechanisms, evolution, and evaluation. *Hum. Vaccines Immunother.* **2018**, *14*, 1840–1847. [CrossRef] [PubMed]
14. Guthrie, J.L.; Gardy, J.L. A brief primer on genomic epidemiology: Lessons learned from *Mycobacterium tuberculosis*. *Ann. N. Y. Acad. Sci.* **2017**, *1388*, 59–77. [CrossRef]
15. Eyre, D.W. Infection prevention and control insights from a decade of pathogen whole-genome sequencing. *J. Hosp. Infect.* **2022**, *122*, 180–186. [CrossRef]
16. Gardy, J.L.; Johnston, J.C.; Sui, S.J.H.; Cook, V.J.; Shah, L.; Brodkin, E.; Rempel, S.; Moore, R.; Zhao, Y.; Holt, R.; et al. Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. *N. J. Med.* **2011**, *364*, 730–739. [CrossRef]
17. Ko, K.K.; Chng, K.R.; Nagarajan, N. Metagenomics-enabled microbial surveillance. *Nat. Microbiol.* **2022**, *7*, 486–496. [CrossRef]
18. Nwadiugwu, M.C.; Monteiro, N. Applied genomics for identification of virulent biothreats and for disease outbreak surveillance. *Postgrad. Med. J.* **2022**, *99*, 403–410. [CrossRef]
19. Cameron, A.; Bohrhunter, J.L.; Taffner, S.; Malek, A.; Pecora, N.D. Clinical Pathogen Genomics. *Clin. Lab. Med.* **2020**, *40*, 447–458. [CrossRef]
20. Goldberg, B.; Sichtig, H.; Geyer, C.; Ledeboer, N.; Weinstock, G.M. Making the leap from research laboratory to clinic: Challenges and opportunities for next-generation sequencing in infectious disease diagnostics. *MBio* **2015**, *6*, e01888-15. [CrossRef]
21. Gire, S.K.; Goba, A.; Andersen, K.G.; Sealfon, R.S.; Park, D.J.; Kanneh, L.; Jalloh, S.; Momoh, M.; Fullah, M.; Dudas, G.; et al. Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. *Science* **2014**, *345*, 1369–1372. [CrossRef] [PubMed]
22. Jenkins, C.; Dallman, T.J.; Grant, K.A. Impact of whole genome sequencing on the investigation of food-borne outbreaks of Shiga toxin-producing *Escherichia coli* serogroup O157: H7, England, 2013 to 2017. *Eurosurveillance* **2019**, *24*, 1800346. [CrossRef] [PubMed]
23. Faria, N.R.; Azevedo, R.D.S.D.S.; Kraemer, M.U.; Souza, R.; Cunha, M.S.; Hill, S.C.; Thézé, J.; Bonsall, M.B.; Bowden, T.A.; Rissanen, I.; et al. Zika virus in the Americas: Early epidemiological and genetic findings. *Science* **2016**, *352*, 345–349. [CrossRef] [PubMed]
24. Forsythe, S.J.; Dickins, B.; Jolley, K.A. Cronobacter, the emergent bacterial pathogen *Enterobacter sakazakii* comes of age; MLST and whole genome sequence analysis. *BMC Genom.* **2014**, *15*, 1121. [CrossRef]
25. Koser, C.U.; Holden, M.T.; Ellington, M.J.; Cartwright, E.J.; Brown, N.M.; Ogilvy-Stuart, A.L.; Hsu, L.Y.; Chewapreecha, C.; Croucher, N.J.; Harris, S.R.; et al. Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak. *N. Engl. J. Med.* **2012**, *366*, 2267–2275. [CrossRef]
26. Bartley, P.B.; Ben Zakour, N.L.; Stanton-Cook, M.; Muguli, R.; Prado, L.; Garnys, V.; Taylor, K.; Barnett, T.C.; Pinna, G.; Robson, J.; et al. Hospital-wide eradication of a nosocomial *Legionella pneumophila* serogroup 1 outbreak. *Clin. Infect. Dis.* **2016**, *62*, 273–279. [CrossRef]
27. Grad, Y.H.; Lipsitch, M.; Feldgarden, M.; Arachchi, H.M.; Cerqueira, G.C.; FitzGerald, M.; Godfrey, P.; Haas, B.J.; Murphy, C.I.; Russ, C.; et al. Genomic epidemiology of the *Escherichia coli* O104: H4 outbreaks in Europe, 2011. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 3065–3070. [CrossRef]
28. Frampton, D.; Rampling, T.; Cross, A.; Bailey, H.; Heaney, J.; Byott, M.; Scott, R.; Sconza, R.; Price, J.; Margaritis, M.; et al. Genomic characteristics and clinical effect of the emergent SARS-CoV-2 B. 1.1. 7 lineage in London, UK: A whole-genome sequencing and hospital-based cohort study. *Lancet Infect. Dis.* **2021**, *21*, 1246–1256. [CrossRef]
29. Sharma, C.; Kumar, N.; Pandey, R.; Meis, J.F.; Chowdhary, A. Whole genome sequencing of emerging multidrug resistant *Candida auris* isolates in India demonstrates low genetic variation. *New Microbes New Infect.* **2016**, *13*, 77–82. [CrossRef]
30. O'Donnell, M.R.; Larsen, M.H.; Brown, T.S.; Jain, P.; Munsamy, V.; Wolf, A.; Uccellini, L.; Karim, F.; de Oliveira, T.; Mathema, B.; et al. Early detection of emergent extensively drug-resistant tuberculosis by flow cytometry-based phenotyping and whole-genome sequencing. *Antimicrob. Agents Chemother.* **2019**, *63*, e01834-18. [CrossRef]
31. Espenhain, L.; Funk, T.; Overvad, M.; Edslev, S.M.; Fonager, J.; Ingham, A.C.; Rasmussen, M.; Madsen, S.L.; Espersen, C.H.; Sieber, R.N.; et al. Epidemiological characterisation of the first 785 SARS-CoV-2 Omicron variant cases in Denmark, December 2021. *Eurosurveillance* **2021**, *26*, 2101146. [CrossRef]
32. Fonager, J.; Bennedbæk, M.; Bager, P.; Wohlfahrt, J.; Ellegaard, K.M.; Ingham, A.C.; Edslev, S.M.; Stegger, M.; Sieber, R.N.; Lassauniere, R.; et al. Molecular epidemiology of the SARS-CoV-2 variant Omicron BA. 2 sub-lineage in Denmark, 29 November 2021 to 2 January 2022. *Eurosurveillance* **2022**, *27*, 2200181. [CrossRef] [PubMed]
33. Whaley, M.J.; Joseph, S.J.; Retchless, A.C.; Kretz, C.B.; Blain, A.; Hu, F.; Chang, H.Y.; Mbaeyi, S.A.; MacNeil, J.R.; Read, T.D.; et al. Whole genome sequencing for investigations of meningococcal outbreaks in the United States: A retrospective analysis. *Sci. Rep.* **2018**, *8*, 15803. [CrossRef] [PubMed]

34. Gilchrist, C.A.; Turner, S.D.; Riley, M.F.; Petri, W.A., Jr.; Hewlett, E.L. Whole-genome sequencing in outbreak analysis. *Clin. Microbiol. Rev.* **2015**, *28*, 541–563. [CrossRef] [PubMed]
35. Kwong, J.C.; McCallum, N.; Sintchenko, V.; Howden, B.P. Whole genome sequencing in clinical and public health microbiology. *Pathology* **2015**, *47*, 199–210. [CrossRef] [PubMed]
36. Hay, A.J.; McCauley, J.W. The WHO global influenza surveillance and response system (GISRS)—A future perspective. *Influenza Other Respir. Viruses* **2018**, *12*, 551–557. [CrossRef] [PubMed]
37. Carter, L.L.; Yu, M.A.; Sacks, J.A.; Barnadas, C.; Pereyaslov, D.; Cognat, S.; Briand, S.; Ryan, M.J.; Samaan, G. Global genomic surveillance strategy for pathogens with pandemic and epidemic potential 2022–2032. *Bull. World Health Organ.* **2022**, *100*, 239–239A. [CrossRef] [PubMed]
38. Towner, J.S.; Sealy, T.K.; Khristova, M.L.; Albariño, C.G.; Conlan, S.; Reeder, S.A.; Quan, P.L.; Lipkin, W.I.; Downing, R.; Tappero, J.W.; et al. Newly discovered ebola virus associated with hemorrhagic fever outbreak in Uganda. *PLoS Pathog.* **2008**, *4*, e1000212. [CrossRef]
39. Lam, T.T.Y.; Jia, N.; Zhang, Y.W.; Shum, M.H.H.; Jiang, J.F.; Zhu, H.C.; Tong, Y.G.; Shi, Y.X.; Ni, X.B.; Liao, Y.S.; et al. Identifying SARS-CoV-2-related coronaviruses in *Malayan pangolins*. *Nature* **2020**, *583*, 282–285. [CrossRef]
40. Walker, T.M.; Kohl, T.A.; Omar, S.V.; Hedge, J.; Elias, C.D.O.; Bradley, P.; Iqbal, Z.; Feuerriegel, S.; Niehaus, K.E.; Wilson, D.J.; et al. Whole-genome sequencing for prediction of *Mycobacterium tuberculosis* drug susceptibility and resistance: A retrospective cohort study. *Lancet Infect. Dis.* **2015**, *15*, 1193–1202. [CrossRef]
41. Lockhart, S.R.; Etienne, K.A.; Vallabhaneni, S.; Farooqi, J.; Chowdhary, A.; Govender, N.P.; Colombo, A.L.; Calvo, B.; Cuomo, C.A.; Desjardins, C.A.; et al. Simultaneous emergence of multidrug-resistant *Candida auris* on 3 continents confirmed by whole-genome sequencing and epidemiological analyses. *Clin. Infect. Dis.* **2017**, *64*, 134–140. [CrossRef]
42. Sardi, S.I.; Carvalho, R.H.; Pacheco, L.G.C.; Almeida, J.P.P.d.; Belitardo, E.M.M.d.A.; Pinheiro, C.S.; Campos, G.S.; Aguiar, E.R.G.R. High-quality resolution of the outbreak-related Zika virus genome and discovery of new viruses using ion torrent-based metatranscriptomics. *Viruses* **2020**, *12*, 782. [CrossRef] [PubMed]
43. Poon, L.L.; Song, T.; Rosenfeld, R.; Lin, X.; Rogers, M.B.; Zhou, B.; Sebra, R.; Halpin, R.A.; Guan, Y.; Twaddle, A.; et al. Quantifying influenza virus diversity and transmission in humans. *Nat. Genet.* **2016**, *48*, 195–200. [CrossRef] [PubMed]
44. Faria, N.R.; Quick, J.; Claro, I.M.; Theze, J.; de Jesus, J.G.; Giovanetti, M.; Kraemer, M.U.; Hill, S.C.; Black, A.; da Costa, A.C.; et al. Establishment and cryptic transmission of Zika virus in Brazil and the Americas. *Nature* **2017**, *546*, 406–410. [CrossRef] [PubMed]
45. Quick, J.; Loman, N.J.; Duraffour, S.; Simpson, J.T.; Severi, E.; Cowley, L.; Bore, J.A.; Koundouno, R.; Dudas, G.; Mikhail, A.; et al. Real-time, portable genome sequencing for Ebola surveillance. *Nature* **2016**, *530*, 228–232. [CrossRef]
46. Miller, R.R.; Montoya, V.; Gardy, J.L.; Patrick, D.M.; Tang, P. Metagenomics for pathogen detection in public health. *Genome Med.* **2013**, *5*, 81. [CrossRef]
47. Piombo, E.; Abdelfattah, A.; Droby, S.; Wisniewski, M.; Spadaro, D.; Schena, L. Metagenomics approaches for the detection and surveillance of emerging and recurrent plant pathogens. *Microorganisms* **2021**, *9*, 188. [CrossRef] [PubMed]
48. Kingry, L.; Sheldon, S.; Oatman, S.; Pritt, B.; Anacker, M.; Bjork, J.; Neitzel, D.; Strain, A.; Berry, J.; Sloan, L.; et al. Targeted metagenomics for clinical detection and discovery of bacterial tick-borne pathogens. *J. Clin. Microbiol.* **2020**, *58*, e00147-20. [CrossRef]
49. Balière, C.; Hourdel, V.; Kwasiborski, A.; Grassin, Q.; Feher, M.; Hoinard, D.; Vanhomwegen, J.; Taieb, F.; Consigny, P.H.; Manuguerra, J.C.; et al. Complete genome sequences of monkeypox virus from a French clinical sample and the corresponding isolated strain, obtained using nanopore sequencing. *Microbiol. Resour. Announc.* **2023**, *12*, e0000923. [CrossRef]
50. Gauthier, N.P.; Nelson, C.; Bonsall, M.B.; Locher, K.; Charles, M.; MacDonald, C.; Krajden, M.; Chorlton, S.D.; Manges, A.R. Nanopore metagenomic sequencing for detection and characterization of SARS-CoV-2 in clinical samples. *PLoS ONE* **2021**, *16*, e0259712. [CrossRef]
51. Kugelman, J.R.; Wiley, M.R.; Mate, S.; Ladner, J.T.; Beitzel, B.; Fakoli, L.; Taweh, F.; Prieto, K.; Diclaro, J.W.; Minogue, T.; et al. Monitoring of Ebola virus Makona evolution through establishment of advanced genomic capability in Liberia. *Emerg. Infect. Dis.* **2015**, *21*, 1135. [CrossRef] [PubMed]
52. Claro, I.M.; Romano, C.M.; Candido, D.D.S.; Lima, E.L.D.; Lindoso, J.A.L.; Ramundo, M.S.; Moreira, F.R.R.; Barra, L.A.C.; Borges, L.M.S.; Medeiros, L.A.; et al. Shotgun metagenomic sequencing of the first case of monkeypox virus in Brazil, 2022. *Rev. Inst. Med. Trop. São Paulo* **2022**, *64*. [CrossRef]
53. Greninger, A.L.; Chen, E.C.; Sittler, T.; Scheinerman, A.; Roubinian, N.; Yu, G.; Kim, E.; Pillai, D.R.; Guyard, C.; Mazzulli, T.; et al. A metagenomic analysis of pandemic influenza A (2009 H1N1) infection in patients from North America. *PLoS ONE* **2010**, *5*, e13381. [CrossRef] [PubMed]
54. Cotten, M.; Watson, S.J.; Kellam, P.; Al-Rabeeah, A.A.; Makhdoom, H.Q.; Assiri, A.; Al-Tawfiq, J.A.; Alhakeem, R.F.; Madani, H.; AlRabiah, F.A.; et al. Transmission and evolution of the Middle East respiratory syndrome coronavirus in Saudi Arabia: A descriptive genomic study. *Lancet* **2014**, *382*, 1993–2002. [CrossRef] [PubMed]
55. Seekatz, A.M.; Aas, J.; Gessert, C.E.; Rubin, T.A.; Saman, D.M.; Bakken, J.S.; Young, V.B. Recovery of the gut microbiome following fecal microbiota transplantation. *mBio* **2014**, *5*, e00893-14. [CrossRef] [PubMed]
56. Sheahan, T.; Hakstol, R.; Kailasam, S.; Glaister, G.D.; Hudson, A.J.; Wieden, H.J. Rapid metagenomics analysis of EMS vehicles for monitoring pathogen load using nanopore DNA sequencing. *PLoS ONE* **2019**, *14*, e0219961. [CrossRef] [PubMed]

57. Thannesberger, J.; Rascovan, N.; Eisenmann, A.; Klymiuk, I.; Zittra, C.; Fuehrer, H.P.; Scantlebury-Manning, T.; Gittens-St Hilaire, M.; Austin, S.; Landis, R.C.; et al. Viral metagenomics reveals the presence of novel Zika virus variants in Aedes mosquitoes from Barbados. *Parasites Vectors* **2021**, *14*, 343. [CrossRef]
58. Li, T.; Mbala-Kingebeni, P.; Naccache, S.N.; Thézé, J.; Bouquet, J.; Federman, S.; Somasekar, S.; Yu, G.; Sanchez-San Martin, C.; Achari, A.; 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*, e00827-19. [CrossRef]
59. Aljabr, W.; Alruwaili, M.; Penrice-Randal, R.; Alrezaihi, A.; Harrison, A.J.; Ryan, Y.; Bentley, E.; Jones, B.; Alhatlani, B.Y.; AlShahrani, D.; et al. Amplicon and metagenomic analysis of middle east respiratory syndrome (MERS) coronavirus and the microbiome in patients with severe MERS. *Msphere* **2021**, *6*, e0021921. [CrossRef]
60. Souza, J.V.C.; Santos, H.D.O.; Leite, A.B.; Giovanetti, M.; Bezerra, R.D.S.; Carvalho, E.D.; Bernardino, J.D.S.T.; Viala, V.L.; Haddad, R.; Ciccozzi, M.; et al. Viral metagenomics for the identification of emerging infections in clinical samples with inconclusive Dengue, Zika, and Chikungunya viral amplification. *Viruses* **2022**, *14*, 1933. [CrossRef]
61. Qiu, Y.; Wang, S.; Huang, B.; Zhong, H.; Pan, Z.; Zhuang, Q.; Peng, C.; Hou, G.; Wang, K. Viral infection detection using metagenomics technology in six poultry farms of eastern China. *PLoS ONE* **2019**, *14*, e0211553. [CrossRef] [PubMed]
62. Lewandowski, K.; Xu, Y.; Pullan, S.T.; Lumley, S.F.; Foster, D.; Sanderson, N.; Vaughan, A.; Morgan, M.; Bright, N.; Kavanagh, J.; et al. Metagenomic nanopore sequencing of influenza virus direct from clinical respiratory samples. *J. Clin. Microbiol.* **2019**, *58*, e00963-19. [CrossRef] [PubMed]
63. Loman, N.J.; Constantinidou, C.; Christner, M.; Rohde, H.; Chan, J.Z.M.; Quick, J.; Weir, J.C.; Quince, C.; Smith, G.P.; Betley, J.R.; et al. A culture-independent sequence-based metagenomics approach to the investigation of an outbreak of Shiga-toxigenic *Escherichia coli* O104: H4. *JAMA* **2013**, *309*, 1502–1510. [CrossRef] [PubMed]
64. Chain, P.; Kurtz, S.; Ohlebusch, E.; Slezak, T. An applications-focused review of comparative genomics tools: Capabilities, limitations and future challenges. *Brief. Bioinform.* **2003**, *4*, 105–123. [CrossRef]
65. Miller, W.; Makova, K.D.; Nekrutenko, A.; Hardison, R.C. Comparative genomics. *Annu. Rev. Genom. Hum. Genet.* **2004**, *5*, 15–56. [CrossRef] [PubMed]
66. Eichler, E.E. Genetic variation, comparative genomics, and the diagnosis of disease. *N. Engl. J. Med.* **2019**, *381*, 64–74. [CrossRef]
67. Zhang, Q.; Jing, S.; Cheng, Z.; Yu, Z.; Dehghan, S.; Shamsaddini, A.; Yan, Y.; Li, M.; Seto, D. Comparative genomic analysis of two emergent human adenovirus type 14 respiratory pathogen isolates in China reveals similar yet divergent genomes. *Emerg. Microbes Infect.* **2017**, *6*, e92. [CrossRef]
68. Rasko, D.A.; Webster, D.R.; Sahl, J.W.; Bashir, A.; Boisen, N.; Scheutz, F.; Paxinos, E.E.; Sebra, R.; Chin, C.S.; Iliopoulos, D.; et al. Origins of the *E. coli* strain causing an outbreak of hemolytic–uremic syndrome in Germany. *N. Engl. J. Med.* **2011**, *365*, 709–717. [CrossRef]
69. Ahammad, I.; Hossain, M.U.; Rahman, A.; Chowdhury, Z.M.; Bhattacharjee, A.; Das, K.C.; Keya, C.A.; Salimullah, M. Wave-wise comparative genomic study for revealing the complete scenario and dynamic nature of COVID-19 pandemic in Bangladesh. *PLoS ONE* **2021**, *16*, e0258019. [CrossRef]
70. Asrani, P.; Hasan, G.M.; Sohal, S.S.; Hassan, M.I. Molecular basis of pathogenesis of coronaviruses: A comparative genomics approach to planetary health to prevent zoonotic outbreaks in the 21st century. *Omics J. Integr. Biol.* **2020**, *24*, 634–644. [CrossRef]
71. Khan, M.I.; Khan, Z.A.; Baig, M.H.; Ahmad, I.; Farouk, A.E.; Song, Y.G.; Dong, J.J. Comparative genome analysis of novel coronavirus (SARS-CoV-2) from different geographical locations and the effect of mutations on major target proteins: An in silico insight. *PLoS ONE* **2020**, *15*, e0238344. [CrossRef]
72. Cheng, Z.; Yan, Y.; Jing, S.; Li, W.G.; Chen, W.W.; Zhang, J.; Li, M.; Zhao, S.; Cao, N.; Ou, J.; et al. Comparative genomic analysis of re-emergent human adenovirus type 55 pathogens associated with adult severe community-acquired pneumonia reveals conserved genomes and capsid proteins. *Front. Microbiol.* **2018**, *9*, 1180. [CrossRef] [PubMed]
73. Jun, S.R.; Leuze, M.R.; Nookaew, I.; Uberbacher, E.C.; Land, M.; Zhang, Q.; Wanchai, V.; Chai, J.; Nielsen, M.; Trolle, T.; et al. Ebolavirus comparative genomics. *FEMS Microbiol. Rev.* **2015**, *39*, 764–778. [CrossRef] [PubMed]
74. Brinkman, F.S.; Leipe, D.D. *Phylogenetic Analysis. Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins*, 2nd ed.; John Wiley & Sons, Inc: New York, NY, USA, 2001; Volume 2, p. 349.
75. Behl, A.; Nair, A.; Mohagaonkar, S.; Yadav, P.; Gambhir, K.; Tyagi, N.; Sharma, R.K.; Butola, B.S.; Sharma, N. Threat, challenges, and preparedness for future pandemics: A descriptive review of phylogenetic analysis based predictions. *Infect. Genet. Evol.* **2022**, *98*, 105217. [CrossRef] [PubMed]
76. Hodcroft, E.B.; Zuber, M.; Nadeau, S.; Vaughan, T.G.; Crawford, K.H.; Althaus, C.L.; Reichmuth, M.L.; Bowen, J.E.; Walls, A.C.; Corti, D.; et al. Spread of a SARS-CoV-2 variant through Europe in the summer of 2020. *Nature* **2021**, *595*, 707–712. [CrossRef] [PubMed]
77. Zhou, P.; Yang, X.L.; Wang, X.G.; Hu, B.; Zhang, L.; Zhang, W.; Si, H.R.; Zhu, Y.; Li, B.; Huang, C.L.; et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **2020**, *579*, 270–273. [CrossRef]
78. Forster, P.; Forster, L.; Renfrew, C.; Forster, M. Phylogenetic network analysis of SARS-CoV-2 genomes. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 9241–9243. [CrossRef] [PubMed]
79. Dudas, G.; Carvalho, L.M.; Rambaut, A.; Bedford, T. MERS-CoV spillover at the camel-human interface. *eLife* **2017**, *6*, e31257.
80. Yebra, G.; Ragonnet-Cronin, M.; Ssemwanga, D.; Parry, C.M.; Logue, C.H.; Cane, P.A.; Kaleebu, P.; Brown, A.J.L. Analysis of the history and spread of HIV-1 in Uganda using phylodynamics. *J. Gen. Virol.* **2015**, *96*, 1890. [CrossRef]

81. Benvenuto, D.; Cella, E.; Fogolari, M.; De Florio, L.; Borsetti, A.; Donati, D.; Garilli, F.; Spoto, S.; Ceccarelli, G.; Angeletti, S.; et al. The transmission dynamic of Madariaga Virus by bayesian phylogenetic analysis: Molecular surveillance of an emergent pathogen. *Microb. Pathog.* **2019**, *132*, 80–86. [CrossRef]
82. Chen, J.S.; Hsu, B.M.; Tsai, H.C.; Chen, Y.P.; Huang, T.Y.; Li, K.Y.; Ji, D.D.; Lee, H.S. Molecular surveillance of Vittaforma-like microsporidia by a small-volume procedure in drinking water source in Taiwan: Evidence for diverse and emergent pathogens. *Environ. Sci. Pollut. Res.* **2018**, *25*, 18823–18837. [CrossRef] [PubMed]
83. Bowers, J.R.; Driebe, E.M.; Albrecht, V.; McDougal, L.K.; Granade, M.; Roe, C.C.; Lemmer, D.; Rasheed, J.K.; Engelthaler, D.M.; Keim, P.; et al. Improved subtyping of *Staphylococcus aureus* clonal complex 8 strains based on whole-genome phylogenetic analysis. *Msphere* **2018**, *3*, e00464-17. [CrossRef] [PubMed]
84. Lam, T.T.Y.; Hon, C.C.; Tang, J.W. Use of phylogenetics in the molecular epidemiology and evolutionary studies of viral infections. *Crit. Rev. Clin. Lab. Sci.* **2010**, *47*, 5–49. [CrossRef] [PubMed]
85. Lukashev, A.N.; Vakulenko, Y.A.; Turbabina, N.A.; Deviatkin, A.A.; Drexler, J.F. Molecular epidemiology and phylogenetics of human enteroviruses: Is there a forest behind the trees? *Rev. Med. Virol.* **2018**, *28*, e2002. [CrossRef] [PubMed]
86. Yang, H.; Zhang, Y.; Teng, X.; Hou, H.; Deng, R.; Li, J. CRISPR-based nucleic acid diagnostics for pathogens. *TrAC Trends Anal. Chem.* **2023**, *160*, 116980. [CrossRef] [PubMed]
87. Mustafa, M.I.; Makhawi, A.M. SHERLOCK and DETECTR: CRISPR-Cas systems as potential rapid diagnostic tools for emerging infectious diseases. *J. Clin. Microbiol.* **2021**, *59*, e00745-20. [CrossRef]
88. Chen, J.S.; Ma, E.; Harrington, L.B.; Da Costa, M.; Tian, X.; Palefsky, J.M.; Doudna, J.A. Crispr-Cas12a Target Binding Unleashes Indiscriminate Single-Stranded Dnase Activity. *Science* **2018**, *360*, 436–439. [CrossRef] [PubMed]
89. Sun, Y.; Yu, L.; Liu, C.; Ye, S.; Chen, W.; Li, D.; Huang, W. One-tube SARS-CoV-2 detection platform based on RT-RPA and CRISPR/Cas12a. *J. Transl. Med.* **2021**, *19*, 1–10. [CrossRef]
90. Broughton, J.P.; Deng, X.; Yu, G.; Fasching, C.L.; Servellita, V.; Singh, J.; Miao, X.; Streithorst, J.A.; Granados, A.; Sotomayor-Gonzalez, A.; et al. CRISPR–Cas12-based detection of SARS-CoV-2. *Nat. Biotechnol.* **2020**, *38*, 870–874. [CrossRef]
91. Patchsung, M.; Jantarug, K.; Pattama, A.; Aphicho, K.; Suraritdechachai, S.; Meesawat, P.; Sappakhaw, K.; Leelahakorn, N.; Ruenkam, T.; Wongsatit, T.; et al. Clinical Validation of a Cas13-Based Assay for the Detection of SARS-CoV-2 RNA. *Nat. Biomed. Eng.* **2020**, *4*, 1140–1149. [CrossRef]
92. Gootenberg, J.S.; Abudayyeh, O.O.; Lee, J.W.; Essletzbichler, P.; Dy, A.J.; Joung, J.; Verdine, V.; Donghia, N.; Daringer, N.M.; Freije, C.A.; et al. Nucleic acid detection with CRISPR-Cas13a/C2c2. *Science* **2017**, *356*, 438–442. [CrossRef]
93. Agrawal, S.; Fanton, A.; Chandrasekaran, S.S.; Charrez, B.; Escajeda, A.M.; Son, S.; Mcintosh, R.; Bhuiya, A.; de León Derby, M.D.; Switz, N.A.; et al. Rapid, point-of-care molecular diagnostics with Cas13. *MedRxiv* **2021**. [CrossRef]
94. Wang, X.; Xiong, E.; Tian, T.; Cheng, M.; Lin, W.; Wang, H.; Zhang, G.; Sun, J.; Zhou, X. Clustered regularly interspaced short palindromic repeats/Cas9-mediated lateral flow nucleic acid assay. *ACS Nano* **2020**, *14*, 2497–2508. [CrossRef]
95. Pardee, K.; Green, A.A.; Takahashi, M.K.; Braff, D.; Lambert, G.; Lee, J.W.; Ferrante, T.; Ma, D.; Donghia, N.; Fan, M.; et al. Rapid, low-cost detection of Zika virus using programmable biomolecular components. *Cell* **2016**, *165*, 1255–1266. [CrossRef] [PubMed]
96. Nguyen, L.T.; Smith, B.M.; Jain, P.K. Enhancement of Trans-Cleavage Activity of Cas12a with Engineered Crrna Enables Amplified Nucleic Acid Detection. *Nat. Commun.* **2020**, *11*, 4906. [CrossRef] [PubMed]
97. Miao, J.; Zuo, L.; He, D.; Fang, Z.; Berthet, N.; Yu, C.; Wong, G. Rapid detection of Nipah virus using the one-pot RPA-CRISPR/Cas13a assay. *Virus Res.* **2023**, *332*, 199130. [CrossRef] [PubMed]
98. Ganbaatar, U.; Liu, C. CRISPR-based COVID-19 testing: Toward next-generation point-of-care diagnostics. *Front. Cell. Infect. Microbiol.* **2021**, *11*, 663949. [CrossRef]
99. Qiu, E.; Jin, S.; Xiao, Z.; Chen, Q.; Wang, Q.; Liu, H.; Xie, C.; Chen, C.; Li, Z.; Han, S. CRISPR-based Detection of Helicobacter pylori in Stool Samples. *Helicobacter* **2021**, *26*, e12828. [CrossRef] [PubMed]
100. Ramachandran, A.; Huyke, D.A.; Sharma, E.; Sahoo, M.K.; Huang, C.; Banaei, N.; Pinsky, B.A.; Santiago, J.G. Electric field-driven microfluidics for rapid CRISPR-based diagnostics and its application to detection of SARS-CoV-2. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 29518–29525. [CrossRef]
101. Zhan, Y.; Gao, X.; Li, S.; Si, Y.; Li, Y.; Han, X.; Sun, W.; Li, Z.; Ye, F. Development and evaluation of rapid and accurate CRISPR/Cas13-based RNA diagnostics for *Pneumocystis jirovecii* pneumonia. *Front. Cell. Infect. Microbiol.* **2022**, *12*, 765. [CrossRef]
102. Joung, J.; Ladha, A.; Saito, M.; Segel, M.; Bruneau, R.; Mee-li, W.H.; Kim, N.G.; Yu, X.; Li, J.; Walker, B.D.; et al. Point-of-care testing for COVID-19 using SHERLOCK diagnostics. *MedRxiv* **2020**. [CrossRef]
103. Myhrvold, C.; Freije, C.A.; Gootenberg, J.S.; Abudayyeh, O.O.; Metsky, H.C.; Durbin, A.F.; Kellner, M.J.; Tan, A.L.; Paul, L.M.; Parham, L.A.; et al. Field-deployable viral diagnostics using CRISPR-Cas13. *Science* **2018**, *360*, 444–448. [CrossRef] [PubMed]
104. Wang, Y.; Li, J.; Li, S.; Zhu, X.; Wang, X.; Huang, J.; Yang, X.; Tai, J. LAMP-CRISPR-Cas12-based diagnostic platform for detection of *Mycobacterium tuberculosis* complex using real-time fluorescence or lateral flow test. *Microchim. Acta* **2021**, *188*, 1–9. [CrossRef]
105. Ai, J.W.; Zhou, X.; Xu, T.; Yang, M.; Chen, Y.; He, G.Q.; Pan, N.; Cai, Y.; Li, Y.; Wang, X.; et al. CRISPR-based rapid and ultra-sensitive diagnostic test for *Mycobacterium tuberculosis*. *Emerg. Microbes Infect.* **2019**, *8*, 1361–1369. [CrossRef] [PubMed]
106. Kham-Kjing, N.; Ngo-Giang-Huong, N.; Tragoolpua, K.; Khamduang, W.; Hongjaisee, S. Highly Specific and Rapid Detection of hepatitis C virus using RT-LAMP-coupled CRISPR–Cas12 assay. *Diagnostics* **2022**, *12*, 1524. [CrossRef]

107. Chen, Q.; Gul, I.; Liu, C.; Lei, Z.; Li, X.; Raheem, M.A.; He, Q.; Haihui, Z.; Leeansyah, E.; Zhang, C.Y.; et al. CRISPR–Cas12-based field-deployable system for rapid detection of synthetic DNA sequence of the monkeypox virus genome. *J. Med. Virol.* **2023**, *95*, e28385. [CrossRef] [PubMed]
108. Barnes, K.G.; Lachenauer, A.E.; Nitido, A.; Siddiqui, S.; Gross, R.; Beitzel, B.; Siddle, K.J.; Freije, C.A.; Dighero-Kemp, B.; Mehta, S.B.; et al. Deployable CRISPR-Cas13a diagnostic tools to detect and report Ebola and Lassa virus cases in real-time. *Nat. Commun.* **2020**, *11*, 4131. [CrossRef]
109. Zhou, H.; Bu, S.; Xu, Y.; Xue, L.; Li, Z.; Hao, Z.; Wan, J.; Tang, F. CRISPR/Cas13a combined with hybridization chain reaction for visual detection of influenza A (H1N1) virus. *Anal. Bioanal. Chem.* **2022**, *414*, 8437–8445. [CrossRef]
110. Kralik, P.; Ricchi, M. A basic guide to real time PCR in microbial diagnostics: Definitions, parameters, and everything. *Front. Microbiol.* **2017**, *8*, 108. [CrossRef]
111. Boyle, D.S.; Lehman, D.A.; Lillis, L.; Peterson, D.; Singhal, M.; Armes, N.; Parker, M.; Piepenburg, O.; Overbaugh, J. Rapid detection of HIV-1 proviral DNA for early infant diagnosis using recombinase polymerase amplification. *MBio* **2013**, *4*, e00135-13. [CrossRef]
112. Centers for Disease Control and Prevention. *CDC Real Time RT-PCR (rRT-PCR) Protocol for Detection and Characterization of Influenza 2009 A (H1N1) Pdm Virus—RUO International*; CDC: Atlanta, GA, USA, 2010.
113. Tahamtan, A.; Ardebili, A. Real-time RT-PCR in COVID-19 detection: Issues affecting the results. *Expert Rev. Mol. Diagn.* **2020**, *20*, 453–454. [CrossRef]
114. Vogels, C.B.; Brito, A.F.; Wyllie, A.L.; Fauver, J.R.; Ott, I.M.; Kalinich, C.C.; Petrone, M.E.; Casanovas-Massana, A.; Muenker, M.C.; Moore, A.J.; et al. Analytical sensitivity and efficiency comparisons of SARS-CoV-2 qRT-PCR primer-probe sets. *Nat. Microbiol.* **2020**, *5*, 1299–1305. [CrossRef]
115. Faye, O.; Faye, O.; Soropogui, B.; Patel, P.; El Wahed, A.A.; Loucoubar, C.; Fall, G.; Kiory, D.; Magassouba, N.F.; Keita, S.; et al. Development and deployment of a rapid recombinase polymerase amplification Ebola virus detection assay in Guinea in 2015. *Eurosurveillance* **2015**, *20*, 30053. [CrossRef] [PubMed]
116. Waggoner, J.J.; Pinsky, B.A. Zika virus: Diagnostics for an emerging pandemic threat. *J. Clin. Microbiol.* **2016**, *54*, 860–867. [CrossRef]
117. Huang, Y.; Wei, H.; Wang, Y.; Shi, Z.; Raoul, H.; Yuan, Z. Rapid detection of filoviruses by real-time TaqMan polymerase chain reaction assays. *Virol. Sin.* **2012**, *27*, 273–277. [CrossRef] [PubMed]
118. Seki, M.; Takahashi, H.; Yamamoto, N.; Hamaguchi, S.; Ojima, M.; Hirose, T.; Yoshiya, K.; Ogura, H.; Shimazu, T.; Tomono, K. Polymerase chain reaction-based active surveillance of MRSA in emergency department patients. *Infect. Drug Resist.* **2015**, *8*, 113–118. [CrossRef] [PubMed]
119. Kim, J.; Hwang, E.S. Multiplexed diagnosis of four serotypes of dengue virus by real-time RT-PCR. *BioChip J.* **2020**, *14*, 421–428. [CrossRef]
120. Tokarz, R.; Tagliafierro, T.; Cucura, D.M.; Rochlin, I.; Sameroff, S.; Lipkin, W.I. Detection of *Anaplasma phagocytophilum*, *Babesia microti*, *Borrelia burgdorferi*, *Borrelia miyamotoi*, and Powassan virus in ticks by a multiplex real-time reverse transcription-PCR assay. *MSphere* **2017**, *2*, e00151-17. [CrossRef]
121. Li, B.; Liu, H.; Wang, W. Multiplex real-time PCR assay for detection of *Escherichia coli* O157: H7 and screening for non-O157 Shiga toxin-producing *E. coli*. *BMC Microbiol.* **2017**, *17*, 215. [CrossRef]
122. Wei, C.; Zhong, J.; Hu, T.; Zhao, X. Simultaneous detection of *Escherichia coli* O157: H7, *Staphylococcus aureus* and Salmonella by multiplex PCR in milk. *3 Biotech* **2018**, *8*, 76. [CrossRef]
123. Yadav, P.D.; Sahay, R.R.; Balakrishnan, A.; Mohandas, S.; Radhakrishnan, C.; Gokhale, M.D.; Balasubramanian, R.; Abraham, P.; Gupta, N.; Sugunan, A.P.; et al. Nipah virus outbreak in Kerala State, India amidst of COVID-19 pandemic. *Front. Public Health* **2022**, *10*, 818545. [CrossRef] [PubMed]
124. Corman, V.M.; Landt, O.; Kaiser, M.; Molenkamp, R.; Meijer, A.; Chu, D.K.; Bleicker, T.; Brünink, S.; Schneider, J.; Schmidt, M.L.; et al. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Eurosurveillance* **2020**, *25*, 2000045. [CrossRef] [PubMed]
125. L'Huillier, A.G.; Lombos, E.; Tang, E.; Perusini, S.; Eshaghi, A.; Nagra, S.; Frantz, C.; Olsha, R.; Kristjanson, E.; Dimitrova, K.; et al. Evaluation of Altona Diagnostics RealStar Zika virus reverse transcription-PCR test kit for Zika virus PCR testing. *J. Clin. Microbiol.* **2017**, *55*, 1576–1584. [CrossRef] [PubMed]
126. Uhteg, K.; Mostafa, H.H. Validation and implementation of an orthopoxvirus qualitative real-time PCR for the diagnosis of monkeypox in the clinical laboratory. *J. Clin. Virol.* **2023**, *158*, 105327. [CrossRef]
127. Georgacopoulos, O.; Nunnally, N.S.; Le, N.; Lysen, C.; Welsh, R.M.; Kordalewska, M.; Perlin, D.S.; Berkow, E.L.; Sexton, D.J. Performance evaluation of culture-independent SYBR green *Candida auris* quantitative PCR diagnostics on anterior nares surveillance swabs. *J. Clin. Microbiol.* **2020**, *58*, e00690-20. [CrossRef] [PubMed]
128. Antón, A.; Marcos, M.A.; Codoñer, F.M.; de Molina, P.; Martínez, A.; Cardeñosa, N.; Godoy, P.; Torner, N.; Martínez, M.J.; Ramón, S.; et al. Influenza C virus surveillance during the first influenza A (H1N1) 2009 pandemic wave in Catalonia, Spain. *Diagn. Microbiol. Infect. Dis.* **2011**, *69*, 419–427. [CrossRef]
129. Jothikumar, N.; Lowther, J.A.; Henshilwood, K.; Lees, D.N.; Hill, V.R.; Vinjé, J. Rapid and sensitive detection of noroviruses by using TaqMan-based one-step reverse transcription-PCR assays and application to naturally contaminated shellfish samples. *Appl. Environ. Microbiol.* **2005**, *71*, 1870–1875. [CrossRef]

130. Fernández-Pinero, J.; Gallardo, C.; Elizalde, M.; Robles, A.; Gómez, C.; Bishop, R.; Heath, L.; Couacy-Hymann, E.; Fasina, F.O.; Pelayo, V.; et al. Molecular diagnosis of African swine fever by a new real-time PCR using universal probe library. *Transbound. Emerg. Dis.* **2013**, *60*, 48–58. [CrossRef]
131. Kim, S.H.; Chang, S.Y.; Sung, M.; Park, J.H.; Bin Kim, H.; Lee, H.; Choi, J.P.; Choi, W.S.; Min, J.Y. Extensive viable Middle East respiratory syndrome (MERS) coronavirus contamination in air and surrounding environment in MERS isolation wards. *Rev. Infect. Dis.* **2016**, *63*, 363–369. [CrossRef]
132. Bartlow, A.W.; Stromberg, Z.R.; Gleasner, C.D.; Hu, B.; Davenport, K.W.; Jakhar, S.; Li, P.E.; Vosburg, M.; Garimella, M.; Chain, P.S.; et al. Comparing variability in diagnosis of upper respiratory tract infections in patients using syndromic, next generation sequencing, and PCR-based methods. *PLoS Glob. Public Health* **2022**, *2*, e0000811. [CrossRef]
133. Anderson, E.M.; Maldarelli, F. Quantification of HIV DNA using droplet digital PCR techniques. *Curr. Protoc. Microbiol.* **2018**, *51*, e62. [CrossRef]
134. Fowler, V.L.; Armson, B.; Gonzales, J.L.; Wise, E.L.; Howson, E.L.; Vincent-Mistiaen, Z.; Fouch, S.; Maltby, C.J.; Grippon, S.; Munro, S.; et al. A highly effective reverse-transcription loop-mediated isothermal amplification (RT-LAMP) assay for the rapid detection of SARS-CoV-2 infection. *J. Infect.* **2021**, *82*, 117–125. [CrossRef]
135. Donatin, E.; Drancourt, M. DNA microarrays for the diagnosis of infectious diseases. *Médecine Mal. Infect.* **2012**, *42*, 453–459. [CrossRef]
136. Liu, H.; Bebu, I.; Li, X. Microarray probes and probe sets. *Front. Biosci. (Elite Ed.)* **2010**, *2*, 325. [CrossRef] [PubMed]
137. Wang, D.; Coscoy, L.; Zylberberg, M.; Avila, P.C.; Boushey, H.A.; Ganem, D.; DeRisi, J.L. Microarray-based detection and genotyping of viral pathogens. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 15687–15692. [CrossRef] [PubMed]
138. Wang, D.; Urisman, A.; Liu, Y.T.; Springer, M.; Ksiazek, T.G.; Erdman, D.D.; Mardis, E.R.; Hickenbotham, M.; Magrini, V.; Eldred, J.; et al. Viral discovery and sequence recovery using DNA microarrays. *PLoS Biol.* **2003**, *1*, e2. [CrossRef] [PubMed]
139. Ksiazek, T.G.; Erdman, D.; Goldsmith, C.S.; Zaki, S.R.; Peret, T.; Emery, S.; Tong, S.; Urbani, C.; Comer, J.A.; Lim, W.; et al. A novel coronavirus associated with severe acute respiratory syndrome. *N. Engl. J. Med.* **2003**, *348*, 1953–1966. [CrossRef] [PubMed]
140. Lin, B.; Wang, Z.; Vora, G.J.; Thornton, J.A.; Schnur, J.M.; Thach, D.C.; Blaney, K.M.; Ligler, A.G.; Malanoski, A.P.; Santiago, J.; et al. Broad-spectrum respiratory tract pathogen identification using resequencing DNA microarrays. *Genome Res.* **2006**, *16*, 527–535. [CrossRef] [PubMed]
141. Malanoski, A.P.; Lin, B.; Wang, Z.; Schnur, J.M.; Stenger, D.A. Automated identification of multiple micro-organisms from resequencing DNA microarrays. *Nucleic Acids Res.* **2006**, *34*, 5300–5311. [CrossRef] [PubMed]
142. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [CrossRef]
143. Leski, T.A.; Lin, B.; Malanoski, A.P.; Wang, Z.; Long, N.C.; Meador, C.E.; Barrows, B.; Ibrahim, S.; Hardick, J.P.; Aitichou, M.; et al. Testing and validation of high density resequencing microarray for broad range biothreat agents detection. *PLoS ONE* **2009**, *4*, e6569. [CrossRef]
144. Palacios, G.; Quan, P.L.; Jabado, O.J.; Conlan, S.; Hirschberg, D.L.; Liu, Y.; Zhai, J.; Renwick, N.; Hui, J.; Hegyi, H.; et al. Panmicrobial oligonucleotide array for diagnosis of infectious diseases. *Emerg. Infect. Dis.* **2007**, *13*, 73. [CrossRef]
145. Quan, P.L.; Palacios, G.; Jabado, O.J.; Conlan, S.; Hirschberg, D.L.; Pozo, F.; Jack, P.J.; Cisterna, D.; Renwick, N.; Hui, J.; et al. Detection of respiratory viruses and subtype identification of influenza A viruses by GreeneChipResp oligonucleotide microarray. *J. Clin. Microbiol.* **2007**, *45*, 2359–2364. [CrossRef] [PubMed]
146. Iuchi, H.; Kawasaki, J.; Kubo, K.; Fukunaga, T.; Hokao, K.; Yokoyama, G.; Ichinose, A.; Suga, K.; Hamada, M. Bioinformatics approaches for unveiling virus—Host interactions. *Comput. Struct. Biotechnol. J.* **2023**, *21*, 1774–1784. [CrossRef] [PubMed]
147. Saeb, A.T.; Abouelhoda, M.; Selvaraju, M.; Althawadi, S.I.; Mutabagani, M.; Adil, M.; Al Hokail, A.; Tayeb, H.T. The use of next-generation sequencing in the identification of a fastidious pathogen: A lesson from a clinical setup. *Evol. Bioinform.* **2017**, *13*, 1176934316686072. [CrossRef] [PubMed]
148. Weinstock, G.M. Genomic approaches to studying the human microbiota. *Nature* **2012**, *489*, 250–256. [CrossRef] [PubMed]
149. Wilson, M.R.; Naccache, S.N.; Samayoa, E.; Biagtan, M.; Bashir, H.; Yu, G.; Salamat, S.M.; Somasekar, S.; Federman, S.; Miller, S.; et al. Actionable diagnosis of neuroleptospirosis by next-generation sequencing. *N. Engl. J. Med.* **2014**, *370*, 2408–2417. [CrossRef] [PubMed]
150. Skwor, T. The Use of DNASTAR Lasergene Educational Software with Molecular Techniques to Support Bacterial Identification. *Proc. Assoc. Biol. Lab. Educ.* **2012**, *33*, 327–334.
151. Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C.; et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **2012**, *28*, 1647–1649. [CrossRef]
152. Darling, A.E.; Treangen, T.J.; Messeguer, X.; Perna, N.T. Analyzing patterns of microbial evolution using the mauve genome alignment system. *Comp. Genom.* **2007**, *396*, 135–152.
153. Bragin, E.; Chatzimichali, E.A.; Wright, C.F.; Hurles, M.E.; Firth, H.V.; Bevan, A.P.; Swaminathan, G.J. DECIPHER: Database for the interpretation of phenotype-linked plausibly pathogenic sequence and copy-number variation. *Nucleic Acids Res.* **2014**, *42*, D993–D1000. [CrossRef] [PubMed]

154. Haas, B.J.; Gevers, D.; Earl, A.M.; Feldgarden, M.; Ward, D.V.; Giannoukos, G.; Ciulla, D.; Tabbaa, D.; Highlander, S.K.; Sodergren, E.; et al. Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Res.* **2011**, *21*, 494–504. [CrossRef] [PubMed]
155. Schloss, P.D.; Westcott, S.L.; Ryabin, T.; Hall, J.R.; Hartmann, M.; Hollister, E.B.; Lesniewski, R.A.; Oakley, B.B.; Parks, D.H.; Robinson, C.J.; et al. Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **2009**, *75*, 7537–7541. [CrossRef] [PubMed]
156. *Pathogen Detection Beta [Internet]*; National Library of Medicine (US), National Center for Biotechnology Information: Bethesda, MD, USA, 2004. Available online: https://www.ncbi.nlm.nih.gov/pathogens/ (accessed on 15 August 2023).
157. Snyder, E.E.; Kampanya, N.; Lu, J.; Nordberg, E.K.; Karur, H.R.; Shukla, M.; Soneja, J.; Tian, Y.; Xue, T.; Yoo, H.; et al. PATRIC: The VBI pathosystems resource integration center. *Nucleic Acids Res.* **2007**, *35* (Suppl. S1), D401–D406. [CrossRef] [PubMed]
158. Cosentino, S.; Voldby Larsen, M.; Møller Aarestrup, F.; Lund, O. PathogenFinder-distinguishing friend from foe using bacterial whole genome sequence data. *PLoS ONE* **2013**, *8*, e77302. [CrossRef]
159. Deneke, C.; Rentzsch, R.; Renard, B.Y. PaPrBaG: A machine learning approach for the detection of novel pathogens from NGS data. *Sci. Rep.* **2017**, *7*, 39194. [CrossRef] [PubMed]
160. Joensen, K.G.; Scheutz, F.; Lund, O.; Hasman, H.; Kaas, R.S.; Nielsen, E.M.; Aarestrup, F.M. Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic *Escherichia coli*. *J. Clin. Microbiol.* **2014**, *52*, 1501–1510. [CrossRef] [PubMed]
161. Graham, R.M.A.; Doyle, C.J.; Jennison, A.V. Real-time investigation of a *Legionella pneumophila* outbreak using whole genome sequencing. *Epidemiol. Infect.* **2014**, *142*, 2347–2351. [CrossRef]
162. Inns, T.; Lane, C.; Peters, T.; Dallman, T.; Chatt, C.; McFarland, N.; Crook, P.; Bishop, T.; Edge, J.; Hawker, J.; et al. A multi-country Salmonella Enteritidis phage type 14b outbreak associated with eggs from a German producer: 'near real-time' application of whole genome sequencing and food chain investigations, United Kingdom, May to September 2014. *Eurosurveillance* **2015**, *20*, 21098. [CrossRef]
163. Quick, J.; Ashton, P.; Calus, S.; Chatt, C.; Gossain, S.; Hawker, J.; Nair, S.; Neal, K.; Nye, K.; Peters, T.; et al. Rapid draft sequencing and real-time nanopore sequencing in a hospital outbreak of Salmonella. *Genome Biol.* **2015**, *16*, 114. [CrossRef]
164. Didelot, X.; Fraser, C.; Gardy, J.; Colijn, C. Genomic infectious disease epidemiology in partially sampled and ongoing outbreaks. *Mol. Biol. Evol.* **2017**, *34*, 997–1007. [CrossRef] [PubMed]
165. Woolhouse, M.; Gaunt, E. Ecological origins of novel human pathogens. *Crit. Rev. Microbiol.* **2007**, *33*, 231–242. [CrossRef] [PubMed]
166. Murray, K.A.; Preston, N.; Allen, T.; Zambrana-Torrelio, C.; Hosseini, P.R.; Daszak, P. Global biogeography of human infectious diseases. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 12746–12751. [CrossRef]
167. Kelly, T.R.; Karesh, W.B.; Johnson, C.K.; Gilardi, K.V.; Anthony, S.J.; Goldstein, T.; Olson, S.H.; Machalaba, C.; Predict Consortium; Mazet, J.A. One Health proof of concept: Bringing a transdisciplinary approach to surveillance for zoonotic viruses at the human-wild animal interface. *Prev. Vet. Med.* **2017**, *137*, 112–118. [CrossRef] [PubMed]
168. Smiley Evans, T.; Barry, P.A.; Gilardi, K.V.; Goldstein, T.; Deere, J.D.; Fike, J.; Yee, J.; Ssebide, B.J.; Karmacharya, D.; Cranfield, M.R.; et al. Optimization of a novel non-invasive oral sampling technique for zoonotic pathogen surveillance in nonhuman primates. *PLoS Neglected Trop. Dis.* **2015**, *9*, e0003813. [CrossRef]
169. Anthony, S.J.; Epstein, J.H.; Murray, K.A.; Navarrete-Macias, I.; Zambrana-Torrelio, C.M.; Solovyov, A.; Ojeda-Flores, R.; Arrigo, N.C.; Islam, A.; Ali Khan, S.; et al. A strategy to estimate unknown viral diversity in mammals. *MBio* **2013**, *4*, e00598-13. [CrossRef]
170. Anthony, S.J.; Islam, A.; Johnson, C.; Navarrete-Macias, I.; Liang, E.; Jain, K.; Hitchens, P.L.; Che, X.; Soloyvov, A.; Hicks, A.L.; et al. Non-random patterns in viral diversity. *Nat. Commun.* **2015**, *6*, 8147. [CrossRef]
171. GHRF Commission (Commission on a Global Health Risk Framework for the Future). *The Neglected Dimension of Global Security a Framework to Counter Infectious Disease Crises*; Commission on Global Health Risk Framework for the Future; National Academies Press: Washington, DC, USA, 2016.
172. Mandl, K.D.; Overhage, J.M.; Wagner, M.M.; Lober, W.B.; Sebastiani, P.; Mostashari, F.; Pavlin, J.A.; Gesteland, P.H.; Treadwell, T.; Koski, E.; et al. Implementing syndromic surveillance: A practical guide informed by the early experience. *J. Am. Med. Inform. Assoc.* **2004**, *11*, 141–150. [CrossRef]
173. Henning, K.J. What is syndromic surveillance? *Morb. Mortal. Wkly. Rep.* **2004**, *53*, 7–11.
174. Brownstein, J.S.; Freifeld, C.C.; Madoff, L.C. Digital disease detection—Harnessing the Web for public health surveillance. *N. Engl. J. Med.* **2009**, *360*, 2153. [CrossRef]
175. Smolinski, M.S.; Crawley, A.W.; Baltrusaitis, K.; Chunara, R.; Olsen, J.M.; Wójcik, O.; Santillana, M.; Nguyen, A.; Brownstein, J.S. Flu near you: Crowdsourced symptom reporting spanning 2 influenza seasons. *Am. J. Public Health* **2015**, *105*, 2124–2130. [CrossRef] [PubMed]
176. Hulth, A.; Rydevik, G.; Linde, A. Web queries as a source for syndromic surveillance. *PLoS ONE* **2009**, *4*, e4378. [CrossRef] [PubMed]
177. Carneiro, H.A.; Mylonakis, E. Google trends: A web-based tool for real-time surveillance of disease outbreaks. *Clin. Infect. Dis.* **2009**, *49*, 1557–1564. [CrossRef] [PubMed]

178. O'Shea, J. Digital disease detection: A systematic review of event-based internet biosurveillance systems. *Int. J. Med. Inform.* **2017**, *101*, 15–22. [CrossRef] [PubMed]
179. Brownstein, J.S.; Freifeld, C.C. HealthMap: The development of automated real-time internet surveillance for epidemic intelligence. *Eurosurveillance* **2007**, *12*, 3322. [CrossRef] [PubMed]
180. Neher, R.A.; Bedford, T. Real-time analysis and visualization of pathogen sequence data. *J. Clin. Microbiol.* **2018**, *56*, e00480-18. [CrossRef] [PubMed]
181. Bagabir, S.A.; Ibrahim, N.K.; Bagabir, H.A.; Ateeq, R. Covid-19 and Artificial Intelligence: Genome sequencing, drug development and vaccine discovery. *J. Infect. Public Health* **2022**, *15*, 289–296. [CrossRef]
182. Sundermann, A.; Chen, J.; Kumar, P.; Ayres, A.; Cho, S.; Ezeonwuka, C.; Griffith, M.P.; Miller, J.Z.; Mustapha, M.M.; Pasculle, A.W.; et al. Whole-Genome sequencing surveillance and machine learning of the Electronic Health Record for enhanced healthcare outbreak detection. *Clin. Infect. Dis.* **2021**, *75*, 476–482. [CrossRef]
183. Yu, K.; Beam, A.L.; Kohane, I.S. Artificial intelligence in healthcare. *Nat. Biomed. Eng.* **2018**, *2*, 719–731. [CrossRef]
184. Ross, E. *Perspectives on Data Sharing in Disease Surveillance*; Centre on Global Health Security; The Royal Institute of International Affairs Chatham House: London, UK, 2014.
185. Leber, A.L.; Peterson, E.; Bard, J.D. The hidden crisis in the times of COVID-19: Critical shortages of medical laboratory professionals in clinical microbiology. *J. Clin. Microbiol.* **2022**, *60*, e0024122. [CrossRef]
186. Cornish, N.E.; Bachmann, L.H.; Diekema, D.J.; McDonald, L.C.; McNult, P.; Stevens-Garcia, J.; Raphael, B.H.; Miller, M.B. Pandemic Demand for SARS-CoV-2 Testing Led to Critical Supply and Workforce Shortages in U.S. Clinical and Public Health Laboratories. *J. Clin. Microbiol.* **2023**, *61*, e0318920. [CrossRef] [PubMed]
187. Aarestrup, F.M.; Koopmans, M.G. Sharing data for global infectious disease surveillance and outbreak detection. *Trends Microbiol.* **2016**, *24*, 241–245. [CrossRef] [PubMed]
188. Sane, J.; Edelstein, M. *Overcoming Barriers to Data Sharing in Public Health. A Global Perspective*; Chatham House: London, UK, 2015.
189. Gans, J.; Wolinsky, M.; Dunbar, J. Computational improvements reveal great bacterial diversity and high metal toxicity in soil. *Science* **2005**, *309*, 1387–1390. [CrossRef] [PubMed]
190. Yuan, G.C.; Cai, L.; Elowitz, M.; Enver, T.; Fan, G.; Guo, G.; Irizarry, R.; Kharchenko, P.; Kim, J.; Orkin, S.; et al. Challenges and emerging directions in single-cell analysis. *Genome Biol.* **2017**, *18*, 1–8. [CrossRef]
191. Minoche, A.E.; Dohm, J.C.; Himmelbauer, H. Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and Genome Analyzer systems. *Genome Biol.* **2011**, *12*, R112. [CrossRef] [PubMed]
192. Rotimi, C.N.; Abayomi, A.; Abimiku, A.; Adabayeri, V.; Adebamowo, C.; Adebiyi, E.; Ademola, A.D.; Adeyemo, A.; Adu, D.; Affolabi, D.; et al. Enabling the genomic revolution in Africa. *Science* **2014**, *344*, 1346–1348. [CrossRef]
193. Gurdasani, D.; Carstensen, T.; Tekola-Ayele, F.; Pagani, L.; Tachmazidou, I.; Hatzikotoulas, K.; Karthikeyan, S.; Iles, L.; Pollard, M.; Choudhury, A.; et al. The African Genome Variation Project shapes medical genetics in Africa. *Nature* **2014**, *517*, 327–332. [CrossRef]
194. Moreno-Estrada, A.; Gignoux, C.R.; Fernández-López, J.C.; Zakharia, F.; Sikora, M.; Contreras, A.; Acuña-Alonzo, V.; Sandoval, K.; Eng, C.; Romero-Hidalgo, S.; et al. The genetics of Mexico recapitulates Native American substructure and affects biomedical traits. *Science* **2014**, *344*, 1280–1285. [CrossRef]
195. Jain, M.; Olsen, H.E.; Paten, B.; Akeson, M. Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat. Biotechnol.* **2018**, *36*, 338–345. [CrossRef]
196. Langmead, B.; Nellore, A. Cloud computing for genomic data analysis and collaboration. *Nat. Rev. Genet.* **2018**, *19*, 208–219. [CrossRef]
197. Martinez-Martin, N.; Magnus, D. Privacy and ethical challenges in next-generation sequencing. *Expert Rev. Precis. Med. Drug Dev.* **2019**, *4*, 95–104. [CrossRef]
198. Coltart, C.E.M.; Hoppé, A.; Parker, M.; Dawson, L.; Amon, J.J.; Simwinga, M.; Geller, G.; Henderson, G.E.; Laeyendecker, O.; Tucker, J.D.; et al. Ethical considerations in global HIV phylogenetic research. *Lancet HIV* **2018**, *5*, e656–e666. [CrossRef]
199. Mutenherwa, F.; Wassenaar, D.; De Oliveira, T. Experts' perspectives on key ethical issues associated with HIV phylogenetics as applied in HIV Transmission dynamics research. *J. Empir. Res. Hum. Res. Ethics* **2018**, *14*, 61–77. [CrossRef] [PubMed]
200. Degeling, C.; Johnson, J.; Gilbert, G.L. Perspectives of Australian policy-makers on the potential benefits and risks of technologically enhanced communicable disease surveillance—A modified Delphi survey. *Health Res. Policy Syst.* **2019**, *17*, 35. [CrossRef] [PubMed]
201. Rump, B.; Cornelis, C.; Woonink, F.; Verweij, M. The need for ethical reflection on the use of molecular microbial characterisation in outbreak management. *Eurosurveillance* **2013**, *18*, 20384. [CrossRef] [PubMed]
202. Ribeiro, C.D.S.; Van Roode, M.Y.; Haringhuizen, G.B.; Koopmans, M.; Claassen, E.; Van De Burgwal, L. How ownership rights over microorganisms affect infectious disease control and innovation: A root-cause analysis of barriers to data sharing as experienced by key stakeholders. *PLoS ONE* **2018**, *13*, e0195885. [CrossRef] [PubMed]
203. Knoppers, B.M.; Harris, J.R.; Budin-Ljøsne, I.; Dove, E.S. A framework for responsible sharing of genomic and health-related data. *Lancet Oncol.* **2014**, *15*, e224–e231. [CrossRef]
204. Kiermer, V.; Bourne, P.E.; Fullerton, S.M.; Chambers, C. The European General Data Protection Regulation: Challenges and considerations for research. *eLife* **2018**, *7*, e34473. [CrossRef]

205. Cremers, A.J.; Coolen, J.P.; Bleeker-Rovers, C.P.; van der Geest-Blankert, A.D.; Haverkate, D.; Hendriks, H.; Henriet, S.S.; Huynen, M.A.; Kolwijck, E.; Liem, D.; et al. Surveillance-embedded genomic outbreak resolution of methicillin-susceptible Staphylococcus aureus in a neonatal intensive care unit. *Sci. Rep.* **2020**, *10*, 2619. [CrossRef] [PubMed]
206. Buragohain, L.; Ghosh, M.; Kumar, R.; Dahiya, S.; Malik, Y.S.; Prasad, M. Application of Proteomics and Metabolomics in Disease Diagnosis. In *Advances in Animal Disease Diagnosis*; CRC Press, Taylor & Francis Group: Boca Raton, FL, USA, 2021; pp. 79–102.
207. Sahajpal, N.S.; Mondal, A.K.; Njau, A.; Petty, Z.; Chen, J.; Ananth, S.; Ahluwalia, P.; Williams, C.; Ross, T.M.; Chaubey, A.; et al. High-throughput next-generation sequencing respiratory viral panel: A diagnostic and epidemiologic tool for SARS-CoV-2 and other viruses. *Viruses* **2021**, *13*, 2063. [CrossRef]