Current Trends in Diagnostics of Viral Infections of Unknown Etiology

Daniel Kiselev [1,2] 🆔**, Alina Matsvay** [1,3] 🆔**, Ivan Abramov** [1]**, Vladimir Dedkov** [4,5]**,
German Shipulin** [1] **and Kamil Khafizov** [1,3,*] 🆔

[1] FSBI "Center of Strategic Planning" of the Ministry of Health, 119435 Moscow, Russia;
   neurolynx13@gmail.com (D.K.); arity767@gmail.com (A.M.); abriv@bk.ru (I.A.);
   shipgerman@gmail.com (G.S.)
[2] I.M. Sechenov First Moscow State Medical University, 119146 Moscow, Russia
[3] Moscow Institute of Physics and Technology, National Research University, 117303 Moscow, Russia
[4] Pasteur Institute, Federal Service on Consumers' Rights Protection and Human Well-Being Surveillance,
   197101 Saint-Petersburg, Russia; vgdedkov@gmail.com
[5] Martsinovsky Institute of Medical Parasitology, Tropical and Vector Borne Diseases, Sechenov First Moscow
   State Medical University, 119146 Moscow, Russia
* Correspondence: kkhafizov@gmail.com; Tel.: +7-917-597-2085

✅ check for **updates**

Abstract: Viruses are evolving at an alarming rate, spreading and inconspicuously adapting to cutting-edge therapies. Therefore, the search for rapid, informative and reliable diagnostic methods is becoming urgent as ever. Conventional clinical tests (PCR, serology, etc.) are being continually optimized, yet provide very limited data. Could high throughput sequencing (HTS) become the future gold standard in molecular diagnostics of viral infections? Compared to conventional clinical tests, HTS is universal and more precise at profiling pathogens. Nevertheless, it has not yet been widely accepted as a diagnostic tool, owing primarily to its high cost and the complexity of sample preparation and data analysis. Those obstacles must be tackled to integrate HTS into daily clinical practice. For this, three objectives are to be achieved: (1) designing and assessing universal protocols for library preparation, (2) assembling purpose-specific pipelines, and (3) building computational infrastructure to suit the needs and financial abilities of modern healthcare centers. Data harvested with HTS could not only augment diagnostics and help to choose the correct therapy, but also facilitate research in epidemiology, genetics and virology. This information, in turn, could significantly aid clinicians in battling viral infections.

1. Introduction

As some statistical models suggest, there are over 320,000 mammalian viruses in existence [1], a little over 200 of which are known to infect humans [2], with the number increasing steadily over decades [3]. The recent years have been marked by a rising frequency of emerging viral infections, e.g., SARS [4], Zika [5], or a more recent *Coronavirus* 2019-nCoV outbreak [6]. Bearing that in mind, the range of available diagnostic tools seems disproportionate to the growing number of diseases, because they detect only known pathogens, which constitute ~0.07% of viral entities. Metaphorically, we are attempting to look at a vast sea of threats through the eye of a needle.

As of this day, a multitude of clinical tests are available for detection of viruses, e.g., FISH, ELISA (enzyme-linked immunosorbent assay), and PCR- and their numerous modifications. The latter has

become a gold standard, for example, in cases of gastrointestinal [7] and respiratory infections [8]. Indeed, the presence of viral nucleic acids in biological media, particularly in blood plasma, while not definitive, is highly suspect of an ongoing infection, because it strongly implies viral replication. Furthermore, PCR not only enables us to detect pathogens, but also quantify them with great precision, based on the number of copies of a pathogen's DNA/RNA templates in a sample.

As effective as they are, PCR tests are sometimes prone to major pitfalls, chiefly owing to high mutability of a viral genome, which significantly perplexes primer design and requires researchers to update primer sequences continually. With the entirety of human-infecting viruses, most of which are yet to be revealed, it becomes increasingly difficult to maintain up-to-date primer panels while also ensuring the correct reaction conditions and high specificity for differentiating closely related viral species. This situation is aggravated by the low rates of correct pathogen identification [9].

Immunoassays have been extensively used in clinical laboratories for over 60 years now. Fast and relatively reliable screening tests for infectious diseases have been successfully introduced into healthcare practice, e.g., for HIV, HBV, HEV, HAV, HCV, and Rubella virus. Although ELISA does not quantify pathogens with great precision, it effectively detects pathogens based on their antigen structure. Useful as they are, immunoassays are also subject to biases and fundamental pitfalls, although they are often disregarded for the sake of low cost and convenience.

Problems with conventional diagnostic methods have been stated in several studies. For example, Arnold et al. [10] found that only 62% of viral respiratory infections among children could be confidently attributed to known pathogens, mostly to human metapneumovirus (63%) and adenoviruses (45%), while almost a third of all cases remained undiagnosed. In another study by Vu et al. [11], it was demonstrated that in 42.7% of cases of viral gastroenteritis the pathogen could not be identified with conventional methods. Kennedy et al. [12] outline that up to 40% of viral encephalitis infections remain undiagnosed with modern clinical tests.

High throughput sequencing (HTS) represents a range of technologies based on sequencing by synthesis, which allow retrieving of multiple nucleotide sequences at once with high reliability for further pathogen identification. Thus, HTS may tackle the aforementioned diagnostic issues, adding an extra layer of data for identification and further insight into a pathogen.

In this paper, we assess modern clinical tests for viral infections and explore the ways that HTS provides detailed data about a pathogen's genome, allowing for a deeper analysis of its features, such as drug resistance and phylogeny. We also hope to demonstrate how modern clinical laboratories would benefit from adopting HTS as a routine diagnostic technique.

2. Traditional Methods of Diagnosing Infections

Before molecular and serological diagnostics became widely accessible, detection of pathogens primarily relied on conventional microbiological methods, e.g., cultures on growth media, in cells, and in laboratory animals with further microscopy, identification of antigenic and pathogenic characteristics and the metabolic profile.

ELISA has become one of the most widely used tests, owing to high sensitivity and specificity (both usually reported at >90% [13–15]), low price per sample, and the simplicity of sample processing and result interpretation. Modern commercial kits are designed to cover a large variety of antigens, bacterial and viral, even at femto- and attomole concentrations. All these factors contribute to the popularity of ELISA, turning it into a universal tool for disease screening and prognosis. At the same time, strong limitations restrict its use. Because the test basically relies on antigen detection, cross-reactivity between heterologous antigens can lead to false positive results, like with influenza virus-specific CD8+ T cell and hepatitis C virus antigens [16]. Other fundamental methodological flaws have been described by Hoofnagle and Wener (2009) [17] and include the hook effect, possible presence of anti-reagent and autoantibodies and lack of concordance. It is also of note that ELISA does not provide any extra data beyond approximate viral burden and limited detection of a pathogen, rendering the identification of new viruses impossible.

The development of molecular cloning and commercial solutions for Sanger sequencing aided the identification of new pathogens. The core principle of this approach includes three steps: (1) amplification of a selected region via PCR, (2) molecular cloning of the amplicons, and (3) Sanger sequencing of the molecular clones. In some cases, hybridization is used to detect viral nucleic acids. This requires at least some knowledge of the target sequences, as it utilizes a microarray with oligonucleotides nested on its surface that are complementary to the conservative sequences in the genomes of the chosen viral species. The technique never grew to become popular in diagnostics, because (1) content of viral nucleic acids in the sample had to be high, and (2) high levels of the host's DNA and the DNA of the microbiome interfered with amplification [18].

A number of alternative approaches have been proposed to detect target sequences in the pathogen's genome. Among them are SISPA (sequence-independent single primer amplification) and its modifications [19–23], and VIDISCA (virus discovery based on cDNA-AFLP), which is based on the AFLP (amplified fragment length polymorphism) method that was specifically adapted for studying viromes [23,24].

In spite of a multitude of available technologies, real-time PCR remains, as of now, one of the most powerful and common, often regarded as a reference diagnostic test. RT-PCR has been extensively used [25] for studying viral genera [1,25–27], for which a significant number of primers have been designed [28–30]. Since the potential of this method became apparent, many enterprises have stepped in, developing commercial PCR kits for reliable fast-track diagnosis of infections like enterovirus, influenza, herpes simplex, etc. While effective and highly sensitive, these panels by design cannot identify new strains and types with altered target sequences. Because of possible genetic similarities between known and previously undiscovered viruses, primers might anneal incorrectly, resulting in false positives and taxonomic misidentification. Multiplexing of primers and probes can be challenging, as there is always a risk of cross-specificity and primer-dimer formation. While separating reactions solves this issue, it requires greater volumes of clinical samples, which are usually limited. Furthermore, in the case of probes, the number of targets is strictly limited to a handful of available fluorophores.

The choice of a diagnostic test is normally based on the circumstances. For instance, while popular and simple methods (PCR, ELISA) are quick, they are less informative than Sanger or next generation sequencing, and it is obvious that the latter could supply extra information about pathogens. (Figure 1)
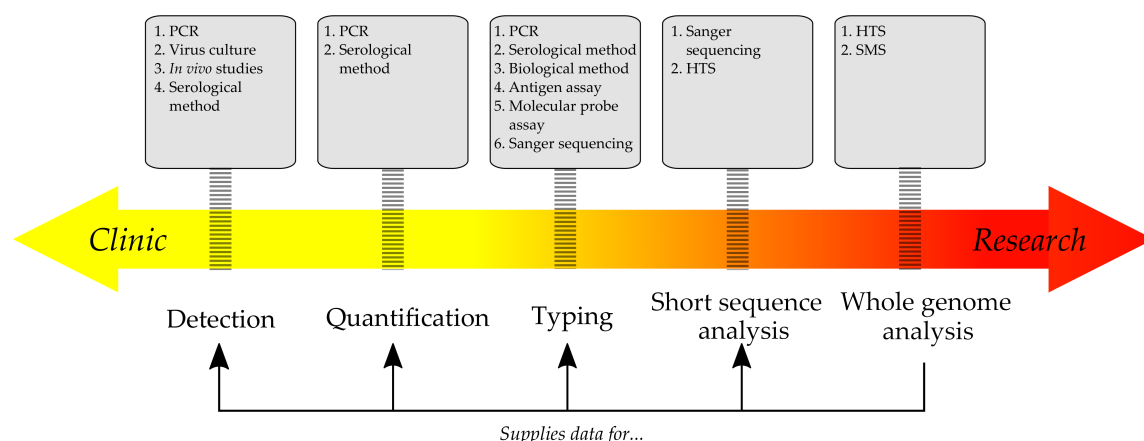

Figure 1. The prevalence of clinical or scientific application depends on the method and the type of data it yields. Classic approaches, like serology and PCR, are quick, but naturally limited to only the known pathogens. More advanced methods, such as HTS, could supply vital data for diagnostics (e.g., optimal target regions for PCR) and further clinical research.

3. Studying Viral Pathogens with High Throughput Sequencing (HTS)

The invention of HTS—alias next generation sequencing (NGS)—has revolutionized research in biology and medical science, including virology, allowing new viruses to be genetically analyzed

without prior cultivation. Sequencing of viral genomes has already become a pivotal part of virological research [31], aiding clinicians in identifying complex infections [32]. It is gradually growing to become one of the main methods in molecular diagnostics [33], development of vaccines [34,35] and searches for new therapeutic strategies [36,37]. Over the years, it has pushed the frontiers of molecular epidemiology [38] and evolutionary genomics [39,40].

Another type of NGS-based analysis is metagenomics, which was comprehensively defined by Chen and Pachter as "the application of modern genomics technique without the need for isolation and lab cultivation of individual species" [41]. In essence, it offers the means to investigate the taxonomic diversity of a microbial community in an environmental or biological sample as the entirety of present nucleic acids is studied. Further applications of this method extend to transcriptomics and proteomics, allowing for a deeper insight into several aspects of pathogens, such as drug resistance, adaptation mechanisms and vast communication networks between species.

3.1. Metagenomic Approach

Metagenomics allows for identification and in-depth studies of unknown viral pathogens by using shotgun sequencing [42–44], which has been extensively applied in clinical and environmental research [45,46]. Nucleic acids from a sample undergo virtually unbiased sequencing, i.e., with minimum prejudice towards specific organisms [47,48]; in theory, the method can be used to analyze a potentially unlimited range of targets [49,50]. Nevertheless, there is evidence that metagenomic sequencing is restricted by multiple pitfalls and biases, based on the pathogen's structure, extraction method, GC content, and other factors [51–53]. Therefore, there are five major limitations: (1) sequences of interest should share at least low identity with the analogous sequences in a reference genome, ensuring correct mapping; (2) the analytical complexity of obtained data often requires employment of a qualified bioinformatician and the use of a specialized computational infrastructure (including both hardware and software; (3) hidden experimental and methodological prejudices towards certain taxa; (4) defining the method's sensitivity and properly measuring it against a relevant reference; (5) dealing with the abundance of host-cell, bacterial and fungal nucleic acids.

Unlike traditional tests [54], metagenomics removes the necessity of designing and synthesizing PCR primers and probes. This reduces time consumption, which is critical during outbreaks of viral infections, such as Zika virus [55] or Ebola virus [56], when fast unbiased pathogen identification is crucial for effective disease containment. Unbiased HTS can also aid in investigating cases of unknown clustered viral infections, when other diagnostic tools do not provide sufficient information, like in the case of the new *Arenavirus* [57].

A metagenomic approach is deemed most helpful when heterogeneous infections share an identical clinical presentation [58] or when genetic markers of antibiotic resistance are in question [50]. Data supplied using NGS also augment diagnostics of respiratory infections: whereas modern multiplex molecular assays reveal the etiology of respiratory infections in approximately 40–80% of cases [58–62], metagenomics successfully tackles this limitation [63].

Furthermore, some respiratory viruses (e.g., *Rhinoviruses*) frequently appear to be the only or the most abundant pathogens in samples from patients with respiratory infections [64]. This pattern implies that either the diagnostic tools are erroneous [65–68] or the groups of related viruses cause infections with a variable clinical manifestation [69]. Metagenomics reveals actual etiological factors, allowing for genotyping, identifying antibiotic resistance markers [70] and supplying data molecular epidemiology [71]. Recently, the applicability of HTS protocols to analyzing clinical samples has been evaluated, demonstrating high potential benefits for clinical studies [58,59,72].

Graf et al. (2016) [47] compared the sensitivity of two pathogen identification methods: metagenomic RNA sequencing (RNA-Seq) with further Taxonomer processing and commercial FDA-approved respiratory virus panel (RVP) for GenMark eSensor. They used 42 virus-positive controls from pediatric patients and 67 unidentified samples [47]. Metagenomic analysis revealed 86% of known respiratory pathogens, with supplementary PCR corroborating the finding in a mere

third of discordant samples. However, for unknown samples, consistency between the two methods reached as high as 93%. Still, metagenomic analysis uncovered 12 extra viruses that were either not targeted by the RVP or failed to bind to complementary nucleotide strands on the chip because of significantly mutated genome sequences. A metagenomic approach not only aids in identification of a pathogen, but also grants an insight into its nucleotide sequence. Although viral sequences constituted only a small fraction of total reads, sufficient data for 84% of samples were gathered to allow for high-resolution genotyping.

Therefore, despite technological encumbrances, a metagenomic approach has a great potential for clinical application. It could provide the means to identify both known and novel pathogens, establish phylogenetic connections and detect drug-resistant and highly virulent strains of common viruses [47].

3.2. Problems of Metagenomic Approach

Even though the cost per raw megabase of DNA sequence has plummeted over the years [73], HTS remains a costly and time-consuming method, requiring a complex procedure of sample preparation, expensive equipment and at least some bioinformatic training with laboratory personnel, therefore rendering it unsuitable for large-scale screening studies [74]. For example, library preparation in the aforementioned study by Graf et al. took no less than 14 h. A HiSeq 2500 (Illumina, San Diego, California, USA) was used to conduct sequencing that lasted 11 consecutive days. Hypothetically, a metagenomic approach could be adapted to a clinical lab routine, provided that library preparation and run processing are optimized to deliver results as fast as conventional tests and present them in an easily interpretable form. Outsourcing of data processing is one option to solve this problem. Ranges of bioinformatics tools facilitate data analysis, e.g., Kraken [75], Kaiju [76], VirusFinder [77]. Different workflows have recently been benchmarked and comprehensively overviewed elsewhere [78–80].

In-house analysis is another alternative, given that the algorithms for sequence processing include criteria that would return a reliable interpretation of assessed reads. Ideally, such algorithms would not only filter out artefacts and perform mapping to the reference genome, but also suggest the most represented pathogen in each sample, narrowing down the spectrum of considered pathogens. This would create a fast definitive NGS-based diagnostic tool [81,82].

Library preparation requires the use of total DNA/RNA from a sample, implying the presence of nucleic acids from other organisms. Direct shotgun sequencing (i.e., without contaminant nucleic acid removal) yields most reads from a host, whereas target reads constitute a mere fraction of total data [50], leading to a significant drop in the method's sensitivity. Researchers are forced to search for target sequences amongst a handful of "junk" reads, employing time and computer power to solve this "needle in a haystack" sort of challenge.

A few strategies can tackle this methodological nuisance. In principal, they are based on either one or both of the following approaches [83]: (1) in silico separation of host sequences based on their similarity with the hosts reference genome; (2) conducting of taxonomic classification for the whole set of sequences using comprehensive databases which includes sequences of both target and host genome. The former option is very demanding in terms of computational power, since millions of reads undergo a complex analysis [84]. The latter is subject to faults of its own: partial or low-quality reference sequences, false mapping and high genetic diversity resulting in numerous polymorphisms [79].

A pathogen's content in a sample is unknown a priori and can only be approximated, for example, based on severity of an onset [85]. Consequently, it becomes difficult to evaluate a minimally required number of sequencing reads per sample (SRPS) sufficient for a reliable in-depth analysis. Underestimation of SRPS results in a low yield of target sequences, for example, 0.008% for Epstein-Barr virus (EBV) [86], 0.0003% for Lassa virus [87] and 0.3% for Zika virus [88]. Considering that metagenomics focuses on relative quantities of nucleic acids rather than absolute counts, establishing appropriate thresholds for positive and negative results becomes challenging. In cases of cell cultures and tissues, calculating copies per cell (CPC) has been suggested [89]; however, this parameter would not work for other types of samples, e.g., swabs and washes. High content of a host's nucleic acids

might artificially lower CPC, falsely attributing negative results to samples. As we have mentioned, the sequencing depth is another aspect to this issue, because shallow sequencing provides little data with poor scalability, to the point where an extra round of sequencing is required to verify the results based on low SRPS data [54]. Therefore, taking the aforementioned factors into account and designing a functional SRPS assessment tool poses an important objective for further research.

Furthermore, DNA/RNA samples might be subject to contamination, which creates artifacts [90]. Thus, it is crucial to keep track of any possible contamination within the laboratory. It would also help to scan scientific magazines periodically for the discovery of new microorganisms in normal human microbiota to adjust the sequence data accordingly.

Metagenomics was initially suggested for expanding the knowledge on existing families, rather than discovering new ones [91], which naturally limits its application to the broad-range search. In practice, reads very often fail to map to reference genomes due to a lack of even minimally homologous sequences [72]. This situation is anticipated in cases of organisms that have not been screened for viral infections before or whenever strains of known viruses in the sample possess unique sequences that have not been described yet. Thus, a better understanding of the virosphere requires not only expansive sets of samples, but also improvements to data processing algorithms. Paradoxically, the more we explore the biology of viruses, the more apparent it becomes that only a negligible part of their taxonomy is being studied and, to make matters worse, mainly that of closely related viruses, rather than the principally new ones [92,93]. In this way, while rapidly gaining profoundness, our knowledge of viruses is progressively narrowing.

Standardization of the method is yet another issue. Because metagenomics is aimed at describing a multitude of microorganisms in a sample, and also due to biases and host-cell DNA and RNA abundance, robust references and normalization methods have to be developed. A possible solution has been proposed by Wilder et al. [94], who suggested creating reference metagenomic libraries for sequencing with unknown samples. Indeed, this approach seems logical and reliable in cases when an abundance of a particular virus is expected. For example, comparing reference samples with a clinical sample, where an abundance of a particular viral species is observed, would indicate a possible on-going infection. Nevertheless, confidence intervals have to be set to ensure that the differences are significant, and studying more samples is required for an accurate evaluation, so further research is required. However, it is challenging to conceptualize standards when the search for new viruses is concerned, because there is no apparent candidate for a standard. In this case, limits on the sequencing quality can be imposed that would help differentiate between sequencing artifacts and the actual discovery of a new pathogen. Validation could include cultivation with supplementary PCR and further whole genome sequencing for new pathogens.

Finally, complex techniques of sample preparation significantly drop reproducibility owing to numerous steps and variations in protocols, resulting in higher error rates. This is a major stumbling block for clinical application of HTS, because there is not much room for error in the medical field, with people's lives being at stake [95].

In summary, the most notable deterrents for clinical metagenomics are: (1) the complexity and high costs of sample preparation, (2) the requirement for special bioinformatic education and skills, (3) the need for a powerful data processing infrastructure and (4) the possible inconsistency between results owing to uneven pathogen distribution in a body and/or quality of sampling. Nevertheless, the mentioned limitations can potentially be alleviated with advances in sequencing technologies [96]. It is important to remember that nearly 40 years ago Sanger sequencing appeared cost-ineffective and overly complex for clinical application [95]; yet, it is one of the widely celebrated diagnostic tools in healthcare [97,98].

In most NGS-based pathogen studies, clutter reads are a problem. Whenever clinical samples are used, the presence of a host's nucleic acids becomes an inevitability, requiring additional computational power for filtering them out and, consequently, adding to the gross sequencing cost. To make matters worse, target DNA/RNA usually constitutes only a small fraction of total DNA/RNA. However, target

amplification combined with enrichment become a game-changer, allowing for both concentration of target templates and removal of unwanted sequences.

3.3. Methods for Improving Sequencing Output

Numerous methods are widely used: affinity enrichment [99], filtration [100], ultracentrifugation [101] and depletion of the host's nucleic acids [90], including protocols utilizing saponin [102] and propidium monoazide [103]. Depletion is a powerful tool, but potential loss of target sequences restricts its use for samples with a low content of viral nucleic acids. Their fraction is usually lost during ethanol cleaning and stems from either weak probe affinity, competitive binding, low ethanol concentration or contamination with strong bases. Additionally, listed techniques are expensive and laborious. Unspecific amplification methods, e.g., multiple displacement amplification (MDA), utilize random primers and Phi29 polymerase to increase reaction output, however not necessarily with high enrichment quality [104], and may lead to additional contamination, uneven fragment distribution and amplification errors, lowering the overall library quality [105]. Sometimes depletion protocols do not effectively remove host-cell DNA [106]. Let us look further into specific techniques for improving NGS library concentration. (Figure 2)
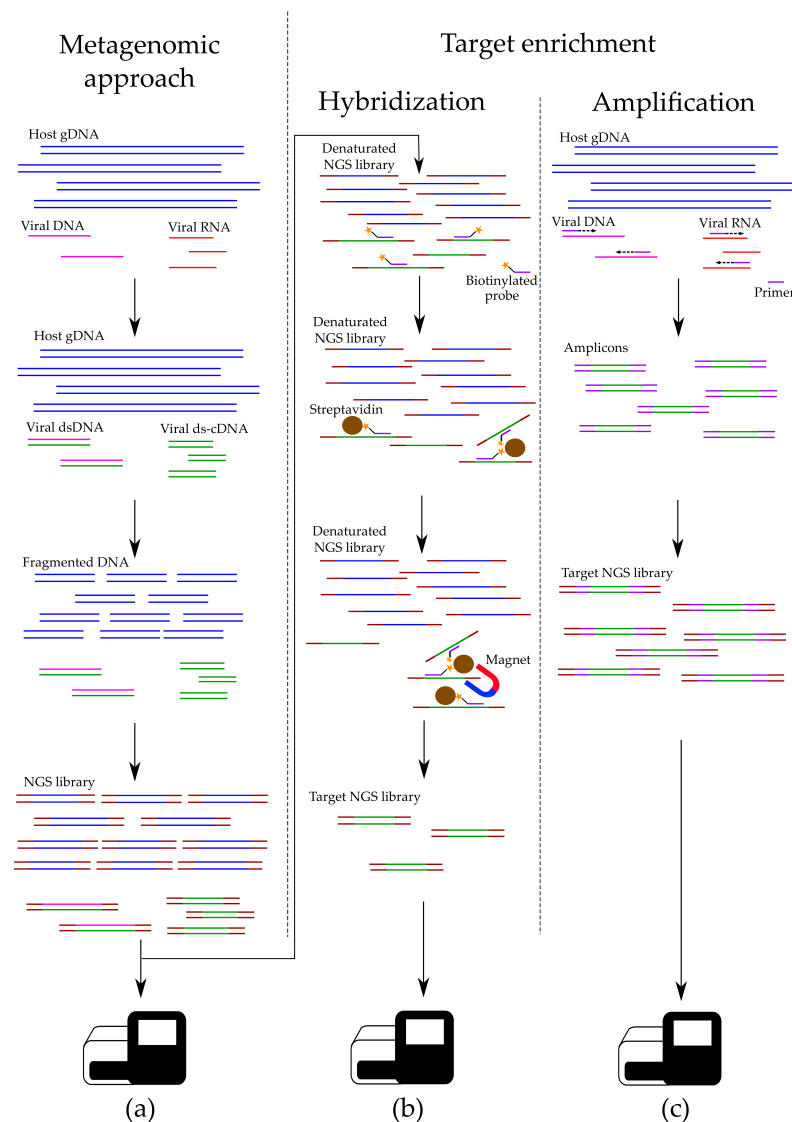

Figure 2. Unlike metagenomics (**a**), target HTS takes on an extra step of hybridization enrichment (**b**) and/or target amplification (**c**), during which the concentration of selected sequences is increased.

3.3.1. Nucleic Acids Depletion

Various types of clinical specimens routinely contain a large portion of host-cell nucleic acids, which can interfere with target PCR [18] and decrease sequencing efficiency and, therefore, have to be removed in a process called depletion. Utilizing biological, physical or chemical principles, it significantly decreases the content of host-cell DNA and/or RNA, while only negligibly lowering concentration of a pathogen's nucleic acids. Most depletion kits are based on hybridization of molecular probes with target sequences, which is a lengthy but effective process. Commercial depletion kits are widely available (Ribo-Zero, RiboCop, HostZERO, NEBNext, MolYsis).

In a protocol by Gu et al. [107], Cas9 utilizes a set of guide RNAs to target unwanted nucleic acids during depletion of abundant sequences by hybridization (DASH). Cleavage-mediated depletion resulted in an up to 99% reduction of mitochondrial DNA. This approach can be used for targeting human genomic DNA in samples with viral nucleic acids.

Despite its beneficial role in sample processing, depletion poses a few risks for the output quality. Since it acts as a bottleneck, which allows only certain sequences through, biases are inherently imposed on the selection process. Cross-hybridization [108] is one example: recombinations in a viral genome or flaws in guide RNA design cause unspecific annealing to homologous non-target sequences, resulting in an erroneous cleavage.

Although separation of target sequences from "junk" nucleic acids is the primary goal of depletion, high levels of host-cell nucleic acids can, in theory, impede probe-sequence binding by competitive mechanisms, in which case the effectiveness of depletion would sag. However, some modified protocols have been developed that tackle this issue [103,109].

All points considered, nucleic acid depletion is an effective tool for decreasing concentrations of unwanted sequences in a sample, despite it possessing a few limitations that could lower the sequencing output by removing target nucleic acids.

3.3.2. Hybridization-Based Enrichment

Experiments that compare modern clinical assays with HTS-based tests conclude that data derived from sequencing might be helpful in several ways. It can serve for pathogen detection and identification and also supply clinically valuable information on the presence of resistance genes or resistance-yielding mutations in a pathogen's genome [110]. It could help healthcare professionals choose the correct therapy without having to wait for microbiological tests and prescribing broad-spectrum drugs that are constantly shown to have adverse side effects [111]. Of note is that NGS finds pathogens beyond standard testing panels, unlike traditional tests. However, genetic libraries must be created, often from samples with low concentrations of a pathogen's nucleic acids and high content of nucleic acids from a host and other microorganisms. An elegant solution has been proposed by two independent teams of researchers, resulting in the development of a novel commercially available enrichment tool [112] based on hybridization.

The most notable feature of this approach is that molecular probes, instead of targeting selected viral species or families, bind to certain sequences in all known viral genera infecting vertebrates, thus supposedly creating the most comprehensive tool for viral screening. This enrichment technique greatly increases sequencing depth, while also raising the fraction of viral reads by 100–10,000 times compared to standard library preparation protocols [113,114]. A diagnostic tool of such sensitivity could serve as an alternative to traditional methods when they fail to detect the pathogen. In contrast, total sample sequencing could at least partly detect a new pathogen's sequences, revealing its identity.

Target probe-based enrichment (also known as hybridization-based enrichment) is used for studying viral genomes without prior cultivation or clonal amplification [105–117]. For this, short DNA or RNA probes are used, which are complementary to the chosen sequences in a viral genome. The trick is to use several probes for one viral genome, thus ensuring its full coverage and guaranteed extraction: even if the genome contains major mutations, there is a distinct possibility that some target sequences remain unchanged, allowing for binding with a probe. Probes themselves are biotinylated

at one end, binding strongly to beads covered in streptavidin. After that, a magnet separator is used to wash the beads from any redundant DNA and RNA with further elution of target sequences [118].

Hybridization-based enrichment holds a few advantages over PCR-based techniques. Alterations to a pathogen's sequence impede primer annealing, whereas probes tend to bind to target sites, given proper reaction conditions and time, even if the target sequence contains polymorphisms [112]. Besides, hybridization works well for covering large genomes, demonstrating less coverage biases [119].

This approach has been successfully used to describe clinically significant viruses with varying genome sizes, such as HCV [54], HSV-1 [120], VZV [115], EBV [121], HCMV [122], HHV6 [119] and HHV7 [123]. Hybridization can be set up in a single tube and is easily automated [121]. Moreover, the use of probes yields results more reliable than those after cultivation. This is because acquired sequences are almost totally identical to original templates, whereas cultivation creates quasispecies, distorting actual genetic structure of a viral population [115,122].

Sensitivity of hybridization is increased when probes are designed using a large set of reference sequences, resulting in a better coverage of all conceivable genetic variants. This is achieved because target enrichment is still possible when target sequences differ from a probe sequence. However, full sequences are needed when creating hybridization probes, while designing PCR primers requires sequences only of the flanking regions. Another advantage of probe-based methods is that if some oligonucleotides fail to anneal during hybridization, overlapping probes might compensate [122,124].

Briese et al. [112] developed VirCapSeq-VERT—a platform for concentrating nucleic acids of vertebrate viruses. It contains approximately two million biotinylated probes for target enrichment. VirCapSeq-VERT detects virtually all human viruses, including new ones, featuring up to 40% of unknown sequences. Another team of researchers created an analogous approach, called ViroCap [117], which has been tested in a virological lab on actual clinical samples. In 26 samples, metagenomic shotgun sequencing (MSS) with ViroCap detected all anticipated viruses plus 30 that had not been identified before. An experiment carried out later [125] demonstrated that out of 30 viruses identified with NGS and missed by clinical tests, 18 were identified by MSS and ViroCap, and an additional 12 with ViroCap only. Another powerful computational tool for designing probes has been recently developed by Metsky et al. [126]. CATCH (compact aggregation of targets for comprehensive hybridization) analyzes target viral genomes and selects a minimum number of probes for optimal coverage of multiple viruses. This results in an effective hybridization-based enrichment, while preserving the original viral diversity of a sample. Reduction in the number of probes curtails experiment costs, while increasing the efficiency of metagenomic sequencing.

Nevertheless, the price per sample remains high for this approach, and it has not gained much popularity yet due to insufficient data on its relative performance. In addition, hybridization protocols are time-consuming. According to official protocols, hybridization may take up to 24 h in some cases, after which there are additional steps required for library preparation, consuming a few extra hours. Secondly, viral genomes that differ from probe sequences for more than 40% will most likely fail to undergo enrichment [112]. Although this does not guarantee the loss of these viruses, the effectiveness of the procedure nosedives. For these reasons, MSS and PCR with degenerate primers might prove more effective, albeit yielding more "noise" data in the output, which is removable during processing. In the end, the choice of enrichment method boils down to the experiment design and quality of the input material.

3.3.3. Target Amplification

Target PCR is an effective and simple alternative to metagenomics, working well for both individual genes and small viral genomes [127]. PCR-based amplification is often used for whole viral genome sequencing of samples with low viral load [54], e.g., during the investigation of the measles outbreak during the Olympics in 2010 [128] and epidemics of Ebola [56] and Zika diseases [129]. Sequencing of long (2.5–3.0 kb) amplified fragments has clarified the variability of a *Norovirus* genome and its spreading among patients in a few hospitals in Vietnam [130,131]. A similar approach has been

utilized to measure the specificity and sensitivity of the Illumina platform for detection of minor polymorphisms in mixed HIV populations [132]. Deep sequencing of PCR-amplified viral genomes yielded complete genome sequences for the influenza virus [133], Dengue virus [134] and HCV [135].

This approach proves apt for investigating small viral genomes that can be covered with only a few PCR amplicons. However, vast heterogeneity of RNA viruses (e.g., HCV27, *Noroviruses* and *Rhabdoviruses*) might require the use of multiple primer sets to ensure amplification of all known genotypes [130,136]. Some researchers propose that this method, coupled with NGS, is used for large viral genomes, like HCMV [19]. In this case, PCR allows long sequences to be "split" into shorter overlapping fragments that can be assembled into a full sequence during data analysis, which increases sequencing depth.

According to Reyes and Kim (1991) [19], another type of amplification is SISPA, a method that serves well when "nucleotide sequence of the desired molecule is both unknown and present in limited amounts making its recovery by standard cloning procedures technically difficult." Recently, it has been used for quick and reliable identification and characterization of viruses [137].

As stated above, PCR works perfectly for amplification of small viral genomes, like those of HIV and influenza virus. For that purpose, primers are designed so that they cover a whole genome, either in a few fragments or in a single molecule. In theory, cloning shorter sequences is simpler, more reliable and increases sequencing depth, whilst also requiring compilation of primer panels for each genus. Nevertheless, target amplification has its own shortcomings. For instance, it has low data scalability, while requiring a relatively large amount of input material to ensure proper site coverage without biases. For this reason, its application is mainly restricted to samples that meet robust requirements (amount of input DNA, material quality, absence of PCR inhibitors, etc.). For instance, amplification of the whole Ebola virus genome utilizes 11 or 19 primer pairs, covering over 97% of the full sequence [138]. Two experiments studying *Noroviruses* included amplification with 14 and 22 primer pairs, respectively [137,138]. Finally, sequencing of the Paramurshir virus genome was conducted with a set of 60 PCR reactions and additional Sanger sequencing to cover the unamplified fragments [139].

Apparently, sometimes amplification, as described above, overloads laboratory workflow with numerous reactions and mandatory normalization of the products' concentrations. There is always room for error, e.g., primers failing to anneal to targets due to unmatching primer sequences (particularly in rapidly mutating viruses [139]). It is also important to consider the costs of primer synthesis and reagents, along with the amount of labor required for setting up multiple reactions per each sample [140]. Thus, even though this method allows for amplification and further sequencing of large viral genomes, technical complexity and low cost-effectiveness render it inapt for massive clinical research, reserving it primarily for scientific purposes.

Multiple PCRs require more input material. In clinical practice, an amount of sample drawn from a patient is usually limited. Multiplex PCR allows for parallel amplification of target sequences in a single tube, thus utilizing a smaller sample volume. In this case, compatibility of primers has to be assessed beforehand using bioinformatic tools to avoid artifacts, such as primer-dimers and false priming.

Target PCR becomes a challenge when organisms with high genome mutability are studied, e.g., HCV [112], influenza virus or *Noroviruses*. Frequent changes in their genetic sequences interfere with primer annealing [54,130,141]. A more subtle approach to designing primers could improve the outcome, but PCR alone is by definition incapable of identifying new viruses.

We suppose that new methods based on metabarcoding might aid identification of viruses in the future. Originally, barcoding—another type of target amplification-mediated approach—was developed to assess bacterial associations (e.g., gut microbiota), based on reliable universal phylogenetic markers, such as 16S, where short taxon-specific sequences consisting of a variable fragment flanked by conservative fragments are used for tagging organisms. Metabarcoding combines a metagenomic approach (i.e., studying multiple organisms in a sample) with barcoding, making it a powerful tool for

studying complex microbial associations [142], fungi and eukaryotes [143]. Usually, the scalability of a metabarcoding approach is limited by genus level; however, this limitation is compensated for by a significant cost reduction in comparison with metagenomics. Unfortunately, no such loci have been described for viruses, owing primarily to their genetic diversity; however, some attempts have been made at solving this issue with broad-range primer panels, created by computational methods [144,145] with promising results.

3.4. Whole Viral Genome Sequencing

Once a pathogen is identified, whole genome sequencing can be engaged for further research into its genome. This could prove beneficial for the knowledge of its genetic patterns, drug resistance and possible targets for antiviral therapy.

Comparatively small genomes, like those of human immunodeficiency virus (HIV), influenza virus, hepatitis B virus (HBV) and hepatitis C virus (HCV), are frequently sequenced in research purposes, and attempts are being made to integrate this procedure into clinical practice [146]. Until recently, detection of drug-resistant strains required sequencing of only certain parts of viral genome. Whole genome sequencing (WGS) was considered far too expensive for this purpose; however, we are obtaining more data on evolving mechanisms of drug resistance, including genetic mutations that define it [147], and with the cost of WGS steadily decreasing, the use of this method becomes justified.

Routine use of NGS for diagnostic purposes is expected to hold great advantages over existing techniques [148]. For instance, the genotyping of *Coronavirus* virus, which has been drawing attention of epidemiologists worldwide, is already influencing decision-making in healthcare [149]. Additionally, genomes of drug-resistant HIV strains are used to study the evolution of the virus [150] and complex genotype/phenotype associations between the virus and the host [151,152].

At present, WGS remains costly, even for small viral genomes, compared to the price of target sequencing. Nevertheless, additional data produced using WGS might supplement our knowledge of genetic substrates for drug resistance and uncover complex intergenic associations at its core.

3.5. Methods of Sequencing Data Analysis

Whenever metagenomics is used for the detection of pathogens, it is crucial to use reliable bioinformatic tools and specialized databases that would help decide whether discovered microorganisms indeed have caused the infection or act merely as artifacts. Such processing normally demands lots of computing power and knowledge and skills in bioinformatics. Typically, data analysis mainly involves comparing obtained reads with reference genomes. Quite a few algorithms have been developed for this purpose [153], with BLAST being the most widely used [154]. However, BLAST works slowly for the analysis of NGS data [155], and processing times can take several days, or even weeks, especially when calculations are performed for amino acids. Other often-used programs are Bowtie [156] and BWA (Burrows-Wheeler Aligner) [157], which are usually employed as filtration tools, and DIAMOND as an alternative to BLAST [158].

As for the published bioinformatic pipelines, in general they adapt either of the two strategies: (1) firsthand removing the host's reads by mapping them to the host's or another non-target reference genome and subsequently analyzing the remaining sequences; (2) assembling short reads into larger contigs, and only then comparing assembled sequences to reference genomes, including those of viruses [80]. (Figure 3)
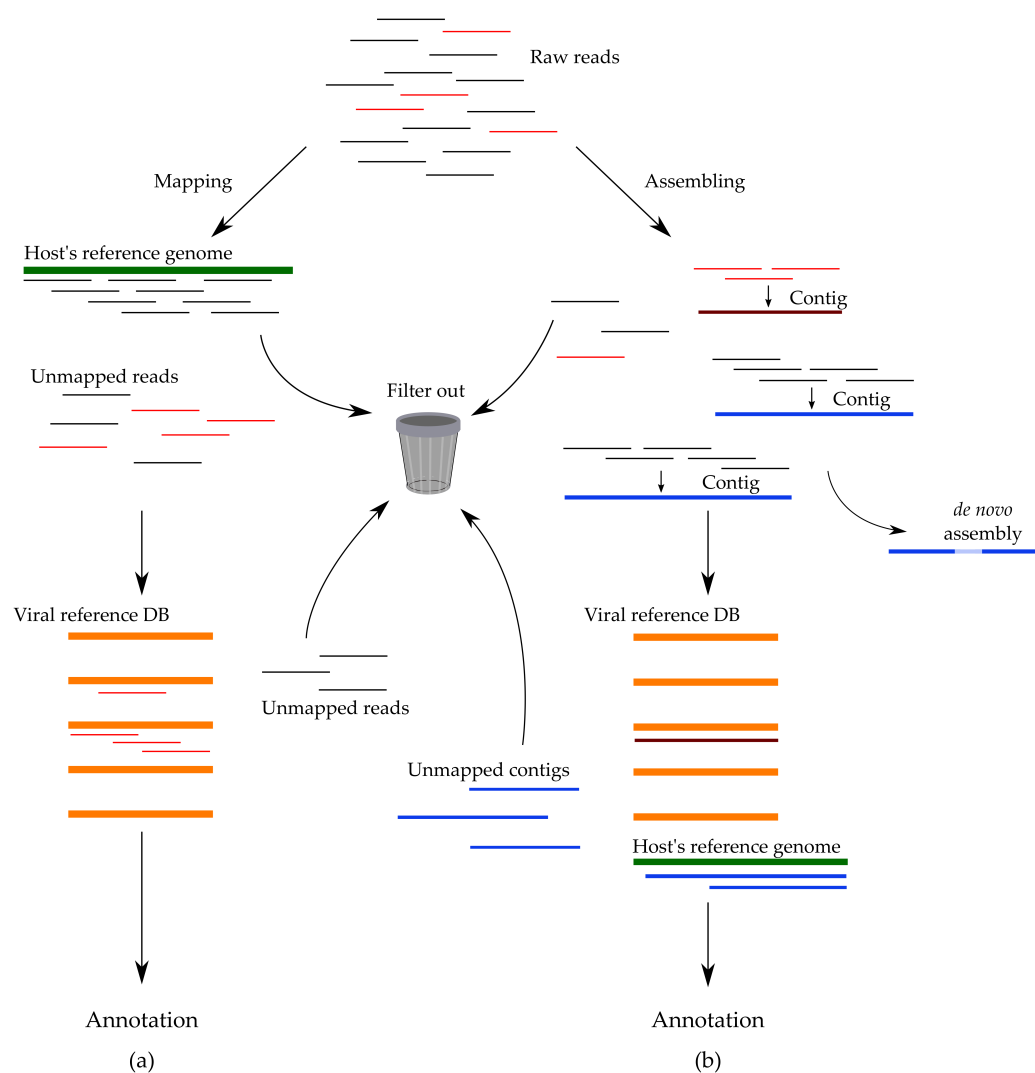
Figure 3. Pipelines for processing reads utilize two main strategies for filtering out "junk" data: (**a**) mapping raw reads onto the reference genome of the host and removing them, while preserving unmapped sequences for further analysis; (**b**) assembling short reads into contigs and comparing them against the host's reference genome. Annotation follows both strategies.

First filtering of the host reads (Figure 3a) works well only when the host's genome has been thoroughly studied and described in detail, although it is not a problem for detection of human viral infections. As HTS technologies advance, more complete genome sequences become publicly available. Another problem is that viruses may contain nucleotide fragments that are similar to certain regions in the human (or other host) genomes, leading to false negative results. The second approach requires high coverage to work, thus significantly increasing both the processing time and the experiment cost. Moreover, whenever input data are insufficient, spaces might form between reads, impairing their assembly into contigs, thus halting virus classification. In principal, two pipelines utilize the same concept of comparing sequences that "survived" filtering against reference viral genomes [80]. Depending on the sample type and expected contaminants, the filtering step may also include rRNA, mtRNA, mRNA, bacterial or fungal sequences or non-human host genomes.

Apart from the differences in the methodology between first filtering out the host reads or leaving them included, there are other serious challenges, such as the proper taxonomic assignment of viral reads or contigs. Homology or alignment-based similarity search methods are often used, but a composition search, in which oligonucleotide frequencies or k-mer counts are matched to reference

sequences, is also an option. At the same time, a composition search requires the program to be "trained" on reference data and it is not used much in viral genomics.

Dismissing the host's sequences is an important step that has been introduced into numerous processing algorithms, such as VirusFinder [77], VirusSeq [159] or Vy-PER [160]. This step helps remove false positives caused by similarities between some regions of the human and viral genomes.

One popular pipeline for analysis of viral reads is by Petty et al. [161]—ezVIR, which was designed to process HTS data from any of the sequencing platforms, and which initially compares raw reads with reference human genome, subsequently removing them and analyzing the remaining sequences against the viral database.

VirusFinder, for instance, first utilizes Bowtie2 [156,162] to detect and remove sequences derived from human genome. Next, it engages BLAT (BLAST-like alignment tool) [163] to align the remaining sequences to the database of viral genomes. During the final step, the short reads, supposedly viral, are assembled into longer sequences—contigs—using Trinity [164]. VirusFinder, developed primarily to identify viral integration sites within the human genome, produces the best results when given sequencing reads with the maximum depth possible. In the original article [77], coverage varies between 31× and 121×. Acquired contigs are then used for phylogenetic analysis.

VirusHunter [165], on the other hand, utilizes BLASTn to filter out human-related sequences after quality evaluation. Sequences that pass filtration are taxonomically classified according to BLASTn and BLASTx algorithms. Thus, VirusHunter requires a high-quality host's genome for sorting the sequences and significant computational power to run.

VirusSeq [159] is intended for detection of viral sequences in tumor tissues. First, it removes human sequences by comparing them against the reference. The MOSAIK program [166] is used both prior to filtration and afterwards to ensure quality sequence detection. VirusSeq sets limits on the minimally acceptable number of reads and coverage based on the size of viral genome. For example, it demands at least 1000 reads per virus given that the sequencing depth equals at least 30×. Although the threshold can be adjusted, this tool has been developed for reads with high coverage and is consequently not recommended for processing data with a low percentage of viral reads.

Vy-Per [160] is another bioinformatic instrument that utilizes reference human genome for filtering out reads of a host's DNA. Sequences that are not dismissed during this step are compared with the data in the NCBI database using BLAT tool [163]. Although experiments set up to test Vy-Per use samples with a rather high coverage (80× for samples and 40× for controls), it is not mandatory, but lowers the risk of false positives.

PathSeq [167] is a powerful computational tool for analyzing the non-host part of sequencing data, which is able to detect the presence of both known and new pathogens. The PathSeq approach begins with a subtraction phase in which reads are aligned with human reference genome to subsequently exclude them, in an attempt to concentrate pathogen-derived data. This is followed by the analytical phase, in which the remaining reads are aligned to the microbial reference sequences and assembled de novo. The formation of large contigs, including several unmapped reads that do not have a significant similarity to alignment with any sequence in the referenced databases, may suggest a previously undetected organism.

SURPI ("sequence-based ultrarapid pathogen identification") [168] is one more example of the pipelines for complex metagenomic NGS data generated from clinical samples that first filter out the non-target reads using 29 databases and then identify viral, bacterial, parasitic and fungal reads, which also involves de novo contig assembly.

There are many other tools available, and we refer to recent reviews by Nooij et al. (2018) [80] and Fonseca et al. [169].

The search for viral pathogens is often impeded by their genetic variability, caused by a multitude of factors: gene duplications and exchanges, frequent single-nucleotide mutations, gene-large insertions and rapid adaptation of viruses to new hosts. These become a particular nuisance whenever large sets of samples from various organisms are handled. Firstly, reliable reference genomes have only

been assembled for a limited number of organisms, although, as previously mentioned, this issue is being addressed. Secondly, the amount of usable viral nucleic acids in a sample depends on numerous factors, including the stage of a pathogen's life cycle and the overall quality of material and its type, all of which might cause prejudice towards certain viral species and drop the amount of usable viral DNA and RNA. This complicates the assembly of contigs, because it requires the maximum number of reads possible.

All search methods rely on genome reference databases, such as the NCBI GenBank, RefSeq or BLAST nucleotide (nt) and non-redundant protein (nr) databases. The poor quality of the reference sequences is a major obstacle to processing data. Because of this, specialized databases are being compiled manually to ensure a strict quality control, e.g., ViPR [170], RVDB [171] and viruSITE [172]. However, they are reliable for the same reason that they are limited, because only a small fraction of all published sequences ever make it to these databases. As a consequence, reads obtained from supposedly new strains are frequently left out. Contrarily, a vast NCBI-based GenBank database is brimming with viral sequences, both complete and partial; however, the price for the quantity is the quality of the assembled data. Even so, GenBank is far from being truly comprehensive, but this is merely a question of time. Protein databases are also used, for example, Pfam [173] and UniProt [174]. Protein-level searches can usually detect more distant homology due to the usage of sensitive amino acid similarity matrices, which can improve the detection of divergent viruses, but untranslated regions of the genome remain unused.

4. Long Read Sequencing

Despite being one of the most powerful research tools, the NGS sequencing, which is sometimes called second-generation sequencing, is not flawless. Namely, not a single platform can read fragments more than 700 bp [175], thus necessitating DNA fragmentation prior to sequencing. Additionally, second-generation platforms amplify the signal prior to detection. PCR is used for molecular cloning of the fragmented DNA, which inevitably results in overrepresentation of some amplicons and uneven sequence coverage. Small length of the reads sometimes becomes a trouble for haplotyping, significantly complicating localization of the questioned loci. It also causes problems when analyzing repeats and recombinant regions of the DNA due to mapping issues.

In contrast, long read sequencing (LRS) technologies do not require significant DNA fragmentation or additional template amplification. For these reasons, they are frequently referred to as the "third-generation" sequencing technologies, although this distinction sometimes may not be accurate. Platforms such as PacBio (Pacific Biosciences), PromethION, GridION and MinION (ONT, Oxford Nanopore Technologies) are capable of reading exceptionally long fragments, including whole viral genomes. Secondly, these platforms potentially allow ultra-high coverage to be easily reached and even quantify the sample, because the viral load would be directly proportional to the number of detectable viral genomes in the sample. Thus, samples with high viral titer can provide high-quality sequences without prior amplification. Another feature of the ONT platforms is the sample turn-around time, which is often critical in the clinical environment; point-of-care testing makes it possible to identify the infectious agent virtually at the patient's bed. In this setting, the correct therapy can be started within hours or days of disease manifestation, prospectively improving the patient care outcome.

One of the most popular ONT sequencers—the MinION—has been praised for possessing multiple boons: relatively low cost, when compared to platforms by Illumina and Thermo Fisher Scientific, and mobility, owing to compact dimensions of the sequencer, simplicity of protocols, and rapid sample processing. Another great advantage that ONT platforms hold over second-generation sequencing is the length of analyzed fragments. Because fragments over 2 Mb can be read in a single run with maintained quality, in metagenomic experiments, genes of interest, like those of drug resistance [176], can be easily attributed to relevant microorganisms, as opposed to reading shorter fragments, when identifying the resistant strains can, in theory, pose a significant challenge. Furthermore, available protocols include not only those for metagenomics, but also for studying select organisms, e.g., viruses.

Both parallel and consecutive sequencing of samples can be performed, owing to a multitude of pores in a cell.

MinION has been extensively used in detecting known [177–181] and new [182] pathogen strains, for instance, of Ebola virus [183], *Begomovirus* [184] and *Papillomavirus* [185]. Curiously, MinION allows samples to be detected and genotyped within mere hours from sample collection, even "in the field", sometimes literally [186,187]. It has also been successfully used for rapid investigation of Dengue virus outbreak in Angola [188] and Ebola epidemic in Guinea [189]. Another experiment [190] allowed for deciphering full genomes of three *Avipoxvirus* strains in real-time, skipping extraction and enrichment altogether. Another analogous example is Everglades Virus (EVEV), which has been detected and subgenotyped on-site [191].

Another great advantage of Nanopore technology is that unique protocols for direct RNA sequencing have been developed, meaning that RNA viruses could be analyzed right away, without reverse transcription [192,193]. Even more, depending on the experiment objective, either differential RNA sequencing (dRNA-Seq) or mRNA could be investigated, the latter being achieved by ligating an adapter only to the poly-A tail, thus allowing for transcriptome analysis [192,194]. Direct RNA-Seq helps avoid distortions of fragments' representation [195] in the sample, revolutionizing the approach to studying RNA-viruses and the intricacies of their life cycle [196].

Despite the apparent benefits of this approach, there are a few drawbacks to consider, e.g., significant error rates that require a large coverage for their partial compensation. This poses a problem that is particularly important for the research of drug resistance, the substrate for which is often a few single-nucleotide polymorphisms, particularly in RNA-viruses [132]. Xu et al. suggest that incorrect sample differentiation and chimeric reads that emerge during sequencing with MinION might significantly alter the experiment outcome [197].

Combining target enrichment with long-range sequencing could potentially solve the problem of high error rate of the sequencing method [198,199] by increasing coverage and improving processing methods [200,201]. For instance, in one metagenomic study of sea water [202], the LASL method (long-read linker-amplified shotgun library) was used to produce enough DNA templates for successful MinION sequencing of a sample with initially low target DNA content (VirION system).

To summarize, small physical dimensions of MinION, along with other ONT sequencers, availability of DNA and RNA sequencing protocols, potentially high precision and high data processing rates create promising prospects for its clinical application [179]. So far, it has already been tested in veterinary science for detection of canine distemper virus (CDV) [203,204] and in biology for identification of numerous pathogens in plants [205] and fish [206]. Nevertheless, prior to extrapolating these positive results into the medical field, we need to develop and validate the workflows that would easily integrate into the everyday work of a clinical laboratory and not require any special bioinformatic training in the lab staff [207].

5. Obstacles to Overcome in the Nearest Future

The risk of contamination and the issues with sensitivity are important to sequencing viral nucleic acids, because they cause false positive and false negative results, respectively. High-sensitivity sequencing (be it metagenomic, PCR-based or probe-enriched) allows for detection of minimal quantities of contaminating viral nucleic acids [208,209]. For example, murine leukemia virus (MLV) [198,210] and *Parvovirus*-like sequences [99,211] are only two kinds of the vast range of common contaminants found even in commercial laboratory supplies [212].

Like with other highly sensitive technologies, only robust validated protocols and their strict implementation can effectively reduce the probability of contamination. This principle has been observed in action for quite a long time in paleogenetics [213], although methods in this field might prove somewhat exorbitant in everyday NGS research with samples not as valuable and unique. Nevertheless, some practices common in this field surely will find application in clinical NGS, when human wellbeing and life are at stake, compared to an abstract scientific interest of pure research.

It is also notable that detection of viral nucleic acids does not necessarily imply an ongoing infection, and therefore, positive results should be validated using supplementary methods targeted at pathogens suggested by the NGS tests. For instance, in the case of idiopathic encephalitis, positive results of NGS can be corroborated either using immunohistochemical assay [214,215], electron microscopy or microbiological methods, i.e., cultivation [216]. At any rate, setting validation thresholds and developing a method for NGS standardization is a significant challenge that has to be addressed prior to introducing NGS into clinical routine.

Undoubtedly, standardization of NGS-based methods, including data processing, will lay a foundation for their widespread integration into clinical routine. Existing algorithms for analyzing sequence output, given user-friendly graphical interfaces, should become easy to handle even without expert knowledge of Linux commands. Another important step is the creation of reliable, comprehensive drug-resistance databases that will enable detection of clinically significant strains required for the special treatment. Such databases are already available for HIV [217], HBV [218,219] and HCV [220].

Proper quality control is another problem. It is difficult to set a valid standard when the method is aimed at discovering new types of organisms. To confirm the NGS data, the pathogen in question has to be extracted, cultured as a pure strain and genetically characterized. Perhaps, the samples subject to NGS assays should be analyzed at several facilities to confirm the reproducibility of the results. The use of different sequencing platforms would ensure even better reproducibility and help prevent biases and typical mistakes that various sequencing technologies are prone to, while also excluding the human factor.

Despite multiple reasons clearly pointing to the necessity of NGS-based tests in virology, most importantly WGS, it is crucial to convince clinical laboratories that it is absolutely worth the effort. For this, clinicians should receive indisputable evidence of the advantages of this approach for the patient, institution budget and long-term prospects for the healthcare system. As for the laboratory staff, the workflow should be clearly structured, time-efficient, scalable and, preferably, automatable, rendering WGS a competitive approach with a highly convenient, reliable and useful outcome.

6. Conclusions

Studying viral genetics with NGS methods is rapidly gaining clinical significance, be it in diagnostics, epidemiological research, the hunt for drug resistance strains or infection control. Various approaches exist that feature WGS, amplicon sequencing, enrichment sequencing and metagenomics, the choice depending on the type of pathogen and experiment objective. Metagenomics, for instance, works best with unknown viruses, whereas PCR-based techniques are apt for samples with low diversity and short genomes. Target enrichment is preferred whenever viruses with large genomes are concerned or if the sample contains a multitude of heterogeneous viruses that feature well-identified nucleic acid sequences.

At present, two major research vectors exist: (1) developing techniques for depletion of bacterial and host sequences that would spare viral nucleic acids and (2) evolving long-range sequencing to the point where it becomes financially and qualitatively comparable to the second-generation methods. The key to success in the struggle against viruses and their shifty nature lies within combining current methods and thus potentiating the creation of the ultimate diagnostic tool. Ideally, it has to be able to detect and describe viral pathogens and predict the evolution of viruses to let clinicians combat them with the utmost effectiveness.

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

References

1. 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*. [CrossRef] [PubMed]

2. Woolhouse, M.; Scott, F.; Hudson, Z.; Howey, R.; Chase-Topping, M. Human viruses: Discovery and emergence. *Philos. Trans. R. Soc. B Biol. Sci.* **2012**, *367*, 2864–2871. [CrossRef] [PubMed]

3. 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] [PubMed]

4. Vijayanand, P.; Wilkins, E.; Woodhead, M. Severe acute respiratory syndrome (SARS): A review. *Clin. Med.* **2004**, *4*, 152–160. [CrossRef]

5. Rather, I.A.; Lone, J.B.; Bajpai, V.K.; Paek, W.K.; Lim, J. Zika Virus: An Emerging Worldwide Threat. *Front. Microbiol.* **2017**, *8*, 1417. [CrossRef]

6. Hui, D.S.; I Azhar, E.; Madani, T.A.; Ntoumi, F.; Kock, R.; Dar, O.; Ippolito, G.; Mchugh, T.D.; Memish, Z.A.; Drosten, C.; et al. The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health—The latest 2019 novel coronavirus outbreak in Wuhan, China. *Int. J. Infect. Dis.* **2020**, *91*, 264–266. [CrossRef]

7. Beal, S.G.; Tremblay, E.E.; Toffel, S.; Velez, L.; Rand, K.H. A Gastrointestinal PCR Panel Improves Clinical Management and Lowers Health Care Costs. *J. Clin. Microbiol.* **2017**, *56*, e01457-17. [CrossRef]

8. Beck, E.T.; Henrickson, K.J. Molecular diagnosis of respiratory viruses. *Future Microbiol.* **2010**, *5*, 901–916. [CrossRef]

9. Schlaberg, R.; Chiu, C.Y.; Miller, S.; Procop, G.W.; Weinstock, G.; the Professional Practice Committee and Committee on Laboratory Practices of the American Society for Microbiology; the Microbiology Resource Committee of the College of American Pathologists. Validation of Metagenomic Next-Generation Sequencing Tests for Universal Pathogen Detection. *Arch. Pathol. Lab. Med.* **2017**, *141*, 776–786. [CrossRef]

10. Arnold, J.C.; Singh, K.K.; Spector, S.A.; Sawyer, M.H. Undiagnosed Respiratory Viruses in Children. *Pediatrics* **2008**, *121*, e631–e637. [CrossRef]

11. Vu, D.-L.; Sabrià, A.; Aregall, N.; Michl, K.; Rodriguez Garrido, V.; Goterris, L.; Bosch, A.; Pintó, R.M.; Guix, S. Novel Human Astroviruses: Prevalence and Association with Common Enteric Viruses in Undiagnosed Gastroenteritis Cases in Spain. *Viruses* **2019**, *11*, 585. [CrossRef] [PubMed]

12. Kennedy, P.; Quan, P.-L.; Lipkin, W. Viral Encephalitis of Unknown Cause: Current Perspective and Recent Advances. *Viruses* **2017**, *9*, 138. [CrossRef] [PubMed]

13. Abdel-Hamid, M.; El-Daly, M.; El-Kafrawy, S.; Mikhail, N.; Strickland, G.T.; Fix, A.D. Comparison of Second- and Third-Generation Enzyme Immunoassays for Detecting Antibodies to Hepatitis C Virus. *J. Clin. Microbiol.* **2002**, *40*, 1656–1659. [CrossRef] [PubMed]

14. Döller, G.; Schuy, W.; Tjhen, K.Y.; Stekeler, B.; Gerth, H.J. Direct detection of influenza virus antigen in nasopharyngeal specimens by direct enzyme immunoassay in comparison with quantitating virus shedding. *J. Clin. Microbiol.* **1992**, *30*, 866–869. [CrossRef] [PubMed]

15. Hammond, G.W.; Smith, S.J.; Noble, G.R. Sensitivity and Specificity of Enzyme Immunoassay for Serodiagnosis of Influenza A Virus Infections. *J. Infect. Dis.* **1980**, *141*, 644–651. [CrossRef] [PubMed]

16. Kasprowicz, V.; Ward, S.M.; Turner, A.; Grammatikos, A.; Nolan, B.E.; Lewis-Ximenez, L.; Sharp, C.; Woodruff, J.; Fleming, V.M.; Sims, S.; et al. Defining the directionality and quality of influenza virus–specific CD8+ T cell cross-reactivity in individuals infected with hepatitis C virus. *J. Clin. Investig.* **2008**, JCI33082. [CrossRef]

17. Hoofnagle, A.N.; Wener, M.H. The fundamental flaws of immunoassays and potential solutions using tandem mass spectrometry. *J. Immunol. Methods* **2009**, *347*, 3–11. [CrossRef]

18. Cogswell, F.B.; Bantar, C.E.; Hughes, T.G.; Gu, Y.; Philipp, M.T. Host DNA can interfere with detection of Borrelia burgdorferi in skin biopsy specimens by PCR. *J. Clin. Microbiol.* **1996**, *34*, 980–982. [CrossRef]

19. Reyes, G.R.; Kim, J.P. Sequence-independent, single-primer amplification (SISPA) of complex DNA populations. *Mol. Cell. Probes* **1991**, *5*, 473–481. [CrossRef]

20. Castrignano, S.B.; Nagasse-Sugahara, T.K.; Kisielius, J.J.; Ueda-Ito, M.; Brandão, P.E.; Curti, S.P. Two novel circo-like viruses detected in human feces: Complete genome sequencing and electron microscopy analysis. *Virus Res.* **2013**, *178*, 364–373. [CrossRef]

21. DePew, J.; Zhou, B.; McCorrison, J.M.; Wentworth, D.E.; Purushe, J.; Koroleva, G.; Fouts, D.E. Sequencing viral genomes from a single isolated plaque. *Virol. J.* **2013**, *10*, 181. [CrossRef] [PubMed]

22. Djikeng, A.; Halpin, R.; Kuzmickas, R.; DePasse, J.; Feldblyum, J.; Sengamalay, N.; Afonso, C.; Zhang, X.; Anderson, N.G.; Ghedin, E.; et al. Viral genome sequencing by random priming methods. *BMC Genom.* **2008**, *9*, 5. [CrossRef] [PubMed]

23. Ambrose, H.E.; Clewley, J.P. Virus discovery by sequence-independent genome amplification. *Rev. Med. Virol.* **2006**, *16*, 365–383. [CrossRef] [PubMed]

24. de Vries, M.; Deijs, M.; Canuti, M.; van Schaik, B.D.C.; Faria, N.R.; van de Garde, M.D.B.; Jachimowski, L.C.M.; Jebbink, M.F.; Jakobs, M.; Luyf, A.C.M.; et al. A Sensitive Assay for Virus Discovery in Respiratory Clinical Samples. *PLoS ONE* **2011**, *6*, e16118. [CrossRef] [PubMed]

25. VanDevanter, D.R.; Warrener, P.; Bennett, L.; Schultz, E.R.; Coulter, S.; Garber, R.L.; Rose, T.M. Detection and analysis of diverse herpesviral species by consensus primer PCR. *J. Clin. Microbiol.* **1996**, *34*, 1666–1671. [CrossRef]

26. Weiss, S.; Witkowski, P.T.; Auste, B.; Nowak, K.; Weber, N.; Fahr, J.; Mombouli, J.-V.; Wolfe, N.D.; Drexler, J.F.; Drosten, C.; et al. Hantavirus in Bat, Sierra Leone. *Emerg. Infect. Dis.* **2012**, *18*, 159–161. [CrossRef]

27. Drosten, C.; Günther, S.; Preiser, W.; van der Werf, S.; Brodt, H.-R.; Becker, S.; Rabenau, H.; Panning, M.; Kolesnikova, L.; Fouchier, R.A.M.; et al. Identification of a Novel Coronavirus in Patients with Severe Acute Respiratory Syndrome. *N. Engl. J. Med.* **2003**, *348*, 1967–1976. [CrossRef]

28. Zheng, L.; Tang, J.; Clover, G.R.G.; Spackman, M.E.; Freeman, A.J.; Rodoni, B.C. Novel genus-specific broad range primers for the detection of furoviruses, hordeiviruses and rymoviruses and their application in field surveys in South-east Australia. *J. Virol. Methods* **2015**, *214*, 1–9. [CrossRef]

29. Choi, S.K.; Choi, J.K.; Park, W.M.; Ryu, K.H. RT–PCR detection and identification of three species of cucumoviruses with a genus-specific single pair of primers. *J. Virol. Methods* **1999**, *83*, 67–73. [CrossRef]

30. Proebster, B.; Kinney, R.M.; Kaaden, O.-R.; Pfeffer, M. Genus-Specific Detection of Alphaviruses by a Semi-Nested Reverse Transcription-Polymerase Chain Reaction. *Am. J. Trop. Med. Hyg.* **1997**, *57*, 709–718. [CrossRef]

31. Radford, A.D.; Chapman, D.; Dixon, L.; Chantrey, J.; Darby, A.C.; Hall, N. Application of next-generation sequencing technologies in virology. *J. Gen. Virol.* **2012**, *93*, 1853–1868. [CrossRef]

32. Jerome, H.; Taylor, C.; Sreenu, V.B.; Klymenko, T.; Filipe, A.D.S.; Jackson, C.; Davis, C.; Ashraf, S.; Wilson-Davies, E.; Jesudason, N.; et al. Metagenomic next-generation sequencing aids the diagnosis of viral infections in febrile returning travellers. *J. Infect.* **2019**, *79*, 383–388. [CrossRef] [PubMed]

33. Prachayangprecha, S.; Schapendonk, C.M.E.; Koopmans, M.P.; Osterhaus, A.D.M.E.; Schurch, A.C.; Pas, S.D.; van der Eijk, A.A.; Poovorawan, Y.; Haagmans, B.L.; Smits, S.L. Exploring the Potential of Next-Generation Sequencing in Detection of Respiratory Viruses. *J. Clin. Microbiol.* **2014**, *52*, 3722–3730. [CrossRef]

34. Dhiman, N.; Smith, D.I.; Poland, G.A. Next-generation sequencing: A transformative tool for vaccinology. *Expert Rev. Vaccines* **2009**, *8*, 963–967. [CrossRef] [PubMed]

35. Luciani, F.; Bull, R.A.; Lloyd, A.R. Next generation deep sequencing and vaccine design: Today and tomorrow. *Trends Biotechnol.* **2012**, *30*, 443–452. [CrossRef] [PubMed]

36. Kruglyak, K.M.; Lin, E.; Ong, F.S. Next-Generation Sequencing and Applications to the Diagnosis and Treatment of Lung Cancer. In *Lung Cancer and Personalized Medicine: Novel Therapies and Clinical Management*; Ahmad, A., Gadgeel, S.M., Eds.; Springer International Publishing: Cham, Switzerland, 2016; Volume 890, pp. 123–136. ISBN 978-3-319-24931-5.

37. Derache, A.; Iwuji, C.C.; Baisley, K.; Danaviah, S.; Marcelin, A.-G.; Calvez, V.; de Oliveira, T.; Dabis, F.; Porter, K.; Pillay, D. Impact of Next-generation Sequencing Defined Human Immunodeficiency Virus Pretreatment Drug Resistance on Virological Outcomes in the ANRS 12249 Treatment-as-Prevention Trial. *Clin. Infect. Dis.* **2019**, *69*, 207–214. [CrossRef] [PubMed]

38. Riley, L.W.; Blanton, R.E. Advances in Molecular Epidemiology of Infectious Diseases: Definitions, Approaches, and Scope of the Field *. *Microbiol. Spectr.* **2018**, *6*. [CrossRef]

39. Wilson, M.R.; Sample, H.A.; Zorn, K.C.; Arevalo, S.; Yu, G.; Neuhaus, J.; Federman, S.; Stryke, D.; Briggs, B.; Langelier, C.; et al. Clinical Metagenomic Sequencing for Diagnosis of Meningitis and Encephalitis. *N. Engl. J. Med.* **2019**, *380*, 2327–2340. [CrossRef]

40. Sukhum, K.V.; Diorio-Toth, L.; Dantas, G. Genomic and Metagenomic Approaches for Predictive Surveillance of Emerging Pathogens and Antibiotic Resistance. *Clin. Pharmacol. Ther.* **2019**, *106*, 512–524. [CrossRef]

41. Chen, K.; Pachter, L. Bioinformatics for Whole-Genome Shotgun Sequencing of Microbial Communities. *PLoS Comput. Biol.* **2005**, *1*, e24. [CrossRef]

42. Worthey, E.A.; Mayer, A.N.; Syverson, G.D.; Helbling, D.; Bonacci, B.B.; Decker, B.; Serpe, J.M.; Dasu, T.; Tschannen, M.R.; Veith, R.L.; et al. Making a definitive diagnosis: Successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease. *Genet. Med.* **2011**, *13*, 255–262. [CrossRef] [PubMed]

43. Li, C.-X.; Shi, M.; Tian, J.-H.; Lin, X.-D.; Kang, Y.-J.; Chen, L.-J.; Qin, X.-C.; Xu, J.; Holmes, E.C.; Zhang, Y.-Z. Unprecedented genomic diversity of RNA viruses in arthropods reveals the ancestry of negative-sense RNA viruses. *eLife* **2015**, *4*, e05378. [CrossRef] [PubMed]

44. Shi, M.; Lin, X.-D.; Tian, J.-H.; Chen, L.-J.; Chen, X.; Li, C.-X.; Qin, X.-C.; Li, J.; Cao, J.-P.; Eden, J.-S.; et al. Redefining the invertebrate RNA virosphere. *Nature* **2016**, *540*, 539–543. [CrossRef] [PubMed]

45. Venter, J.C. Environmental Genome Shotgun Sequencing of the Sargasso Sea. *Science* **2004**, *304*, 66–74. [CrossRef] [PubMed]

46. Mulcahy-O'Grady, H.; Workentine, M.L. The Challenge and Potential of Metagenomics in the Clinic. *Front. Immunol.* **2016**, *7*. [CrossRef] [PubMed]

47. Graf, E.H.; Simmon, K.E.; Tardif, K.D.; Hymas, W.; Flygare, S.; Eilbeck, K.; Yandell, M.; Schlaberg, R. Unbiased Detection of Respiratory Viruses by Use of RNA Sequencing-Based Metagenomics: A Systematic Comparison to a Commercial PCR Panel. *J. Clin. Microbiol.* **2016**, *54*, 1000–1007. [CrossRef]

48. Fischer, N.; Indenbirken, D.; Meyer, T.; Lütgehetmann, M.; Lellek, H.; Spohn, M.; Aepfelbacher, M.; Alawi, M.; Grundhoff, A. Evaluation of Unbiased Next-Generation Sequencing of RNA (RNA-seq) as a Diagnostic Method in Influenza Virus-Positive Respiratory Samples. *J. Clin. Microbiol.* **2015**, *53*, 2238–2250. [CrossRef]

49. Alquezar-Planas, D.E.; Mourier, T.; Bruhn, C.A.W.; Hansen, A.J.; Vitcetz, S.N.; Mørk, S.; Gorodkin, J.; Nielsen, H.A.; Guo, Y.; Sethuraman, A.; et al. Discovery of a divergent HPIV4 from respiratory secretions using second and third generation metagenomic sequencing. *Sci. Rep.* **2013**, *3*, 2468. [CrossRef]

50. Gu, W.; Miller, S.; Chiu, C.Y. Clinical Metagenomic Next-Generation Sequencing for Pathogen Detection. *Annu. Rev. Pathol. Mech. Dis.* **2019**, *14*, 319–338. [CrossRef]

51. McLaren, M.R.; Willis, A.D.; Callahan, B.J. Consistent and correctable bias in metagenomic sequencing experiments. *eLife* **2019**, *8*, e46923. [CrossRef]

52. Boers, S.A.; Jansen, R.; Hays, J.P. Understanding and overcoming the pitfalls and biases of next-generation sequencing (NGS) methods for use in the routine clinical microbiological diagnostic laboratory. *Eur. J. Clin. Microbiol. Infect. Dis.* **2019**, *38*, 1059–1070. [CrossRef] [PubMed]

53. Chen, Y.-C.; Liu, T.; Yu, C.-H.; Chiang, T.-Y.; Hwang, C.-C. Effects of GC Bias in Next-Generation-Sequencing Data on De Novo Genome Assembly. *PLoS ONE* **2013**, *8*, e62856. [CrossRef] [PubMed]

54. Thomson, E.; Ip, C.L.C.; Badhan, A.; Christiansen, M.T.; Adamson, W.; Ansari, M.A.; Bibby, D.; Breuer, J.; Brown, A.; Bowden, R.; et al. Comparison of Next-Generation Sequencing Technologies for Comprehensive Assessment of Full-Length Hepatitis C Viral Genomes. *J. Clin. Microbiol.* **2016**, *54*, 2470–2484. [CrossRef] [PubMed]

55. Mlakar, J.; Korva, M.; Tul, N.; Popović, M.; Poljšak-Prijatelj, M.; Mraz, J.; Kolenc, M.; Resman Rus, K.; Vesnaver Vipotnik, T.; Fabjan Vodušek, V.; et al. Zika Virus Associated with Microcephaly. *N. Engl. J. Med.* **2016**, *374*, 951–958. [CrossRef]

56. 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]

57. Palacios, G.; Druce, J.; Du, L.; Tran, T.; Birch, C.; Briese, T.; Conlan, S.; Quan, P.-L.; Hui, J.; Marshall, J.; et al. A New Arenavirus in a Cluster of Fatal Transplant-Associated Diseases. *N. Engl. J. Med.* **2008**, *358*, 991–998. [CrossRef]

58. Kustin, T.; Ling, G.; Sharabi, S.; Ram, D.; Friedman, N.; Zuckerman, N.; Bucris, E.D.; Glatman-Freedman, A.; Stern, A.; Mandelboim, M. A method to identify respiratory virus infections in clinical samples using next-generation sequencing. *Sci. Rep.* **2019**, *9*, 2606. [CrossRef]

59. Choi, S.-H.; Hong, S.-B.; Ko, G.-B.; Lee, Y.; Park, H.J.; Park, S.-Y.; Moon, S.M.; Cho, O.-H.; Park, K.-H.; Chong, Y.P.; et al. Viral Infection in Patients with Severe Pneumonia Requiring Intensive Care Unit Admission. *Am. J. Respir. Crit. Care Med.* **2012**, *186*, 325–332. [CrossRef]

60. Karhu, J.; Ala-Kokko, T.I.; Vuorinen, T.; Ohtonen, P.; Syrjala, H. Lower Respiratory Tract Virus Findings in Mechanically Ventilated Patients With Severe Community-Acquired Pneumonia. *Clin. Infect. Dis.* **2014**, *59*, 62–70. [CrossRef]

61. Honkinen, M.; Lahti, E.; Österback, R.; Ruuskanen, O.; Waris, M. Viruses and bacteria in sputum samples of children with community-acquired pneumonia. *Clin. Microbiol. Infect.* **2012**, *18*, 300–307. [CrossRef]

62. Jain, S.; Williams, D.J.; Arnold, S.R.; Ampofo, K.; Bramley, A.M.; Reed, C.; Stockmann, C.; Anderson, E.J.; Grijalva, C.G.; Self, W.H.; et al. Community-Acquired Pneumonia Requiring Hospitalization among U.S. Children. *N. Engl. J. Med.* **2015**, *372*, 835–845. [CrossRef] [PubMed]

63. Thorburn, F.; Bennett, S.; Modha, S.; Murdoch, D.; Gunson, R.; Murcia, P.R. The use of next generation sequencing in the diagnosis and typing of respiratory infections. *J. Clin. Virol.* **2015**, *69*, 96–100. [CrossRef] [PubMed]

64. Heinonen, S.; Jartti, T.; Garcia, C.; Oliva, S.; Smitherman, C.; Anguiano, E.; de Steenhuijsen Piters, W.A.A.; Vuorinen, T.; Ruuskanen, O.; Dimo, B.; et al. Rhinovirus Detection in Symptomatic and Asymptomatic Children: Value of Host Transcriptome Analysis. *Am. J. Respir. Crit. Care Med.* **2016**, *193*, 772–782. [CrossRef] [PubMed]

65. Pavia, A.T. Viral Infections of the Lower Respiratory Tract: Old Viruses, New Viruses, and the Role of Diagnosis. *Clin. Infect. Dis.* **2011**, *52*, S284–S289. [CrossRef] [PubMed]

66. Ruuskanen, O.; Jarvinen, A. Editorial Commentary: What is the Real Role of Respiratory Viruses in Severe Community-Acquired Pneumonia? *Clin. Infect. Dis.* **2014**, *59*, 71–73. [CrossRef] [PubMed]

67. Ruuskanen, O.; Lahti, E.; Jennings, L.C.; Murdoch, D.R. Viral pneumonia. *The Lancet* **2011**, *377*, 1264–1275. [CrossRef]

68. Self, W.H.; Williams, D.J.; Zhu, Y.; Ampofo, K.; Pavia, A.T.; Chappell, J.D.; Hymas, W.C.; Stockmann, C.; Bramley, A.M.; Schneider, E.; et al. Respiratory Viral Detection in Children and Adults: Comparing Asymptomatic Controls and Patients With Community-Acquired Pneumonia. *J. Infect. Dis.* **2016**, *213*, 584–591. [CrossRef]

69. Pérez-Sautu, U.; Wiley, M.R.; Iglesias-Caballero, M.; Pozo, F.; Prieto, K.; Chitty, J.A.; García-García, M.L.; Calvo, C.; Casas, I.; Palacios, G. Target-independent high-throughput sequencing methods provide evidence that already known human viral pathogens play a main role in respiratory infections with unexplained etiology. *Emerg. Microbes Infect.* **2019**, *8*, 1054–1065. [CrossRef]

70. Otto, M. Next-generation sequencing to monitor the spread of antimicrobial resistance. *Genome Med.* **2017**, *9*, 68. [CrossRef]

71. Trotter, A.J.; Aydin, A.; Strinden, M.J.; O'Grady, J. Recent and emerging technologies for the rapid diagnosis of infection and antimicrobial resistance. *Curr. Opin. Microbiol.* **2019**, *51*, 39–45. [CrossRef]

72. Datta, S. Next-generation sequencing in clinical virology: Discovery of new viruses. *World J. Virol.* **2015**, *4*, 265. [CrossRef]

73. DNA Sequencing Costs: Data. Available online: https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data (accessed on 26 September 2019).

74. Hillmann, B.; Al-Ghalith, G.A.; Shields-Cutler, R.R.; Zhu, Q.; Gohl, D.M.; Beckman, K.B.; Knight, R.; Knights, D. Evaluating the Information Content of Shallow Shotgun Metagenomics. *mSystems* **2018**, *3*, e00069-18. [CrossRef] [PubMed]

75. Wood, D.E.; Salzberg, S.L. Kraken: Ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* **2014**, *15*, R46. [CrossRef] [PubMed]

76. Menzel, P.; Ng, K.L.; Krogh, A. Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nat. Commun.* **2016**, *7*, 11257. [CrossRef]

77. Wang, Q.; Jia, P.; Zhao, Z. VirusFinder: Software for Efficient and Accurate Detection of Viruses and Their Integration Sites in Host Genomes through Next Generation Sequencing Data. *PLoS ONE* **2013**, *8*, e64465. [CrossRef] [PubMed]

78. Ye, S.H.; Siddle, K.J.; Park, D.J.; Sabeti, P.C. Benchmarking Metagenomics Tools for Taxonomic Classification. *Cell* **2019**, *178*, 779–794. [CrossRef] [PubMed]

79. Breitwieser, F.P.; Lu, J.; Salzberg, S.L. A review of methods and databases for metagenomic classification and assembly. *Brief. Bioinform.* **2019**, *20*, 1125–1136. [CrossRef]

80. Nooij, S.; Schmitz, D.; Vennema, H.; Kroneman, A.; Koopmans, M.P.G. Overview of Virus Metagenomic Classification Methods and Their Biological Applications. *Front. Microbiol.* **2018**, *9*, 749. [CrossRef]

81. Gargis, A.S.; Kalman, L.; Berry, M.W.; Bick, D.P.; Dimmock, D.P.; Hambuch, T.; Lu, F.; Lyon, E.; Voelkerding, K.V.; Zehnbauer, B.A.; et al. Assuring the quality of next-generation sequencing in clinical laboratory practice. *Nat. Biotechnol.* **2012**, *30*, 1033–1036. [CrossRef]

82. Gargis, A.S.; Kalman, L.; Bick, D.P.; da Silva, C.; Dimmock, D.P.; Funke, B.H.; Gowrisankar, S.; Hegde, M.R.; Kulkarni, S.; Mason, C.E.; et al. Good laboratory practice for clinical next-generation sequencing informatics pipelines. *Nat. Biotechnol.* **2015**, *33*, 689–693. [CrossRef]

83. Oulas, A.; Pavloudi, C.; Polymenakou, P.; Pavlopoulos, G.A.; Papanikolaou, N.; Kotoulas, G.; Arvanitidis, C.; Iliopoulos, I. Metagenomics: Tools and Insights for Analyzing Next-Generation Sequencing Data Derived from Biodiversity Studies. *Bioinforma. Biol. Insights* **2015**, *9*. [CrossRef]

84. Prakash, T.; Taylor, T.D. Functional assignment of metagenomic data: Challenges and applications. *Brief. Bioinform.* **2012**, *13*, 711–727. [CrossRef] [PubMed]

85. Hijano, D.R.; Brazelton de Cardenas, J.; Maron, G.; Garner, C.D.; Ferrolino, J.A.; Dallas, R.H.; Gu, Z.; Hayden, R.T. Clinical correlation of influenza and respiratory syncytial virus load measured by digital PCR. *PLoS ONE* **2019**, *14*, e0220908. [CrossRef] [PubMed]

86. Allen, U.D.; Hu, P.; Pereira, S.L.; Robinson, J.L.; Paton, T.A.; Beyene, J.; Khodai-Booran, N.; Dipchand, A.; Hébert, D.; Ng, V.; et al. The genetic diversity of Epstein-Barr virus in the setting of transplantation relative to non-transplant settings: A feasibility study. *Pediatr. Transplant.* **2016**, *20*, 124–129. [CrossRef] [PubMed]

87. Matranga, C.B.; Andersen, K.G.; Winnicki, S.; Busby, M.; Gladden, A.D.; Tewhey, R.; Stremlau, M.; Berlin, A.; Gire, S.K.; England, E.; et al. Enhanced methods for unbiased deep sequencing of Lassa and Ebola RNA viruses from clinical and biological samples. *Genome Biol.* **2014**, *15*, 519. [CrossRef] [PubMed]

88. Calvet, G.; Aguiar, R.S.; Melo, A.S.O.; Sampaio, S.A.; de Filippis, I.; Fabri, A.; Araujo, E.S.M.; de Sequeira, P.C.; de Mendonça, M.C.L.; de Oliveira, L.; et al. Detection and sequencing of Zika virus from amniotic fluid of fetuses with microcephaly in Brazil: A case study. *Lancet Infect. Dis.* **2016**, *16*, 653–660. [CrossRef]

89. Khan, A.S.; Ng, S.H.S.; Vandeputte, O.; Aljanahi, A.; Deyati, A.; Cassart, J.-P.; Charlebois, R.L.; Taliaferro, L.P. A Multicenter Study To Evaluate the Performance of High-Throughput Sequencing for Virus Detection. *mSphere* **2017**, *2*, e00307-17. [CrossRef]

90. Oechslin, C.P.; Lenz, N.; Liechti, N.; Ryter, S.; Agyeman, P.; Bruggmann, R.; Leib, S.L.; Beuret, C.M. Limited Correlation of Shotgun Metagenomics Following Host Depletion and Routine Diagnostics for Viruses and Bacteria in Low Concentrated Surrogate and Clinical Samples. *Front. Cell. Infect. Microbiol.* **2018**, *8*, 375. [CrossRef]

91. Kunin, V.; Copeland, A.; Lapidus, A.; Mavromatis, K.; Hugenholtz, P. A Bioinformatician's Guide to Metagenomics. *Microbiol. Mol. Biol. Rev.* **2008**, *72*, 557–578. [CrossRef]

92. Shi, M.; Lin, X.-D.; Chen, X.; Tian, J.-H.; Chen, L.-J.; Li, K.; Wang, W.; Eden, J.-S.; Shen, J.-J.; Liu, L.; et al. The evolutionary history of vertebrate RNA viruses. *Nature* **2018**, *556*, 197–202. [CrossRef]

93. Zhang, Y.-Z.; Shi, M.; Holmes, E.C. Using Metagenomics to Characterize an Expanding Virosphere. *Cell* **2018**, *172*, 1168–1172. [CrossRef] [PubMed]

94. Wilder, C.N.; Lopera, J.; Mittar, D.; Horton, C. Optimizing microbiome research applications with cutting-edge standards. *Nature*. Available online: https://www.nature.com/articles/d42473-019-00005-x (accessed on 26 September 2019). [CrossRef]

95. Kulkarni, P.; Frommolt, P. Challenges in the Setup of Large-scale Next-Generation Sequencing Analysis Workflows. *Comput. Struct. Biotechnol. J.* **2017**, *15*, 471–477. [CrossRef] [PubMed]

96. Greninger, A.L. The challenge of diagnostic metagenomics. *Expert Rev. Mol. Diagn.* **2018**, *18*, 605–615. [CrossRef] [PubMed]

97. Dewey, F.E.; Pan, S.; Wheeler, M.T.; Quake, S.R.; Ashley, E.A. DNA Sequencing: Clinical Applications of New DNA Sequencing Technologies. *Circulation* **2012**, *125*, 931–944. [CrossRef] [PubMed]

98. Totomoch-Serra, A.; Marquez, M.F.; Cervantes-Barragán, D.E. Sanger sequencing as a first-line approach for molecular diagnosis of Andersen-Tawil syndrome. *F1000Research* **2017**, *6*, 1016. [CrossRef] [PubMed]

99. Oude Munnink, B.B.; Jazaeri Farsani, S.M.; Deijs, M.; Jonkers, J.; Verhoeven, J.T.P.; Ieven, M.; Goossens, H.; de Jong, M.D.; Berkhout, B.; Loens, K.; et al. Autologous Antibody Capture to Enrich Immunogenic Viruses for Viral Discovery. *PLoS ONE* **2013**, *8*, e78454. [CrossRef] [PubMed]

100. Rabinowicz, P.D. Constructing Gene-Enriched Plant Genomic Libraries Using Methylation Filtration Technology. In *Plant Functional Genomics*; Humana Press: Totowa, NJ, USA, 2003; Volume 236, pp. 21–36. ISBN 978-1-59259-413-9.

101. Kohl, C.; Brinkmann, A.; Dabrowski, P.W.; Radonić, A.; Nitsche, A.; Kurth, A. Protocol for Metagenomic Virus Detection in Clinical Specimens1. *Emerg. Infect. Dis.* **2015**, *21*. [CrossRef]

102. Charalampous, T.; Kay, G.L.; Richardson, H.; Aydin, A.; Baldan, R.; Jeanes, C.; Rae, D.; Grundy, S.; Turner, D.J.; Wain, J.; et al. Nanopore metagenomics enables rapid clinical diagnosis of bacterial lower respiratory infection. *Nat. Biotechnol.* **2019**, *37*, 783–792. [CrossRef]

103. Marotz, C.A.; Sanders, J.G.; Zuniga, C.; Zaramela, L.S.; Knight, R.; Zengler, K. Improving saliva shotgun metagenomics by chemical host DNA depletion. *Microbiome* **2018**, *6*, 42. [CrossRef]

104. Sabina, J.; Leamon, J.H. Bias in Whole Genome Amplification: Causes and Considerations. In *Whole Genome Amplification*; Kroneis, T., Ed.; Springer New York: New York, NY, USA, 2015; Volume 1347, pp. 15–41. ISBN 978-1-4939-2989-4.

105. Jensen, R.H.; Mollerup, S.; Mourier, T.; Hansen, T.A.; Fridholm, H.; Nielsen, L.P.; Willerslev, E.; Hansen, A.J.; Vinner, L. Target-Dependent Enrichment of Virions Determines the Reduction of High-Throughput Sequencing in Virus Discovery. *PLoS ONE* **2015**, *10*, e0122636. [CrossRef]

106. Denesvre, C.; Dumarest, M.; Rémy, S.; Gourichon, D.; Eloit, M. Chicken skin virome analyzed by high-throughput sequencing shows a composition highly different from human skin. *Virus Genes* **2015**, *51*, 209–216. [CrossRef] [PubMed]

107. Gu, W.; Crawford, E.D.; O'Donovan, B.D.; Wilson, M.R.; Chow, E.D.; Retallack, H.; DeRisi, J.L. Depletion of Abundant Sequences by Hybridization (DASH): Using Cas9 to remove unwanted high-abundance species in sequencing libraries and molecular counting applications. *Genome Biol.* **2016**, *17*, 41. [CrossRef] [PubMed]

108. Kraus, A.J.; Brink, B.G.; Siegel, T.N. Efficient and specific oligo-based depletion of rRNA. *Sci. Rep.* **2019**, *9*, 12281. [CrossRef] [PubMed]

109. Nelson, M.T.; Pope, C.E.; Marsh, R.L.; Wolter, D.J.; Weiss, E.J.; Hager, K.R.; Vo, A.T.; Brittnacher, M.J.; Radey, M.C.; Hayden, H.S.; et al. Human and Extracellular DNA Depletion for Metagenomic Analysis of Complex Clinical Infection Samples Yields Optimized Viable Microbiome Profiles. *Cell Rep.* **2019**, *26*, 2227–2240.e5. [CrossRef] [PubMed]

110. Tsalik, E.L.; Bonomo, R.A.; Fowler, V.G. New Molecular Diagnostic Approaches to Bacterial Infections and Antibacterial Resistance. *Annu. Rev. Med.* **2018**, *69*, 379–394. [CrossRef] [PubMed]

111. Lindbaek, M. Broad-spectrum antibiotics gave no clinical benefit and more adverse effects than narrow-spectrum antibiotics in treating acute respiratory tract infections in US children. *BMJ Evid.-Based Med.* **2019**, *24*, 119–120. [CrossRef]

112. Briese, T.; Kapoor, A.; Mishra, N.; Jain, K.; Kumar, A.; Jabado, O.J.; Lipkin, W.I. Virome Capture Sequencing Enables Sensitive Viral Diagnosis and Comprehensive Virome Analysis. *mBio* **2015**, *6*, e01491-15. [CrossRef]

113. Vasilyev, E.V.; Trofimov, D.Y.; Tonevitsky, A.G.; Ilinsky, V.V.; Korostin, D.O.; Rebrikov, D.V. Torque Teno Virus (TTV) distribution in healthy Russian population. *Virol. J.* **2009**, *6*, 134. [CrossRef]

114. Hino, S. TTV, a new human virus with single stranded circular DNA genome. *Rev. Med. Virol.* **2002**, *12*, 151–158. [CrossRef]

115. Depledge, D.P.; Kundu, S.; Jensen, N.J.; Gray, E.R.; Jones, M.; Steinberg, S.; Gershon, A.; Kinchington, P.R.; Schmid, D.S.; Balloux, F.; et al. Deep Sequencing of Viral Genomes Provides Insight into the Evolution and Pathogenesis of Varicella Zoster Virus and Its Vaccine in Humans. *Mol. Biol. Evol.* **2014**, *31*, 397–409. [CrossRef]

116. Tsangaras, K.; Wales, N.; Sicheritz-Pontén, T.; Rasmussen, S.; Michaux, J.; Ishida, Y.; Morand, S.; Kampmann, M.-L.; Gilbert, M.T.P.; Greenwood, A.D. Hybridization Capture Using Short PCR Products Enriches Small Genomes by Capturing Flanking Sequences (CapFlank). *PLoS ONE* **2014**, *9*, e109101. [CrossRef] [PubMed]

117. Wylie, T.N.; Wylie, K.M.; Herter, B.N.; Storch, G.A. Enhanced virome sequencing using targeted sequence capture. *Genome Res.* **2015**, *25*, 1910–1920. [CrossRef] [PubMed]

118. Lovett, M.; Kere, J.; Hinton, L.M. Direct selection: A method for the isolation of cDNAs encoded by large genomic regions. *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 9628–9632. [CrossRef] [PubMed]

119. Tweedy, J.; Spyrou, M.A.; Donaldson, C.D.; Depledge, D.; Breuer, J.; Gompels, U.A. Complete Genome Sequence of the Human Herpesvirus 6A Strain AJ from Africa Resembles Strain GS from North America. *Genome Announc.* **2015**, *3*, e01498-14. [CrossRef]

120. Ebert, K.; Depledge, D.P.; Breuer, J.; Harman, L.; Elliott, G. Mode of Virus Rescue Determines the Acquisition of VHS Mutations in VP22-Negative Herpes Simplex Virus 1. *J. Virol.* **2013**, *87*, 10389–10393. [CrossRef]

121. Palser, A.L.; Grayson, N.E.; White, R.E.; Corton, C.; Correia, S.; Ba abdullah, M.M.; Watson, S.J.; Cotten, M.; Arrand, J.R.; Murray, P.G.; et al. Genome Diversity of Epstein-Barr Virus from Multiple Tumor Types and Normal Infection. *J. Virol.* **2015**, *89*, 5222–5237. [CrossRef]

122. Depledge, D.P.; Palser, A.L.; Watson, S.J.; Lai, I.Y.-C.; Gray, E.R.; Grant, P.; Kanda, R.K.; Leproust, E.; Kellam, P.; Breuer, J. Specific Capture and Whole-Genome Sequencing of Viruses from Clinical Samples. *PLoS ONE* **2011**, *6*, e27805. [CrossRef]

123. Donaldson, C.D.; Clark, D.A.; Kidd, I.M.; Breuer, J.; Depledge, D.D. Genome Sequence of Human Herpesvirus 7 Strain UCL-1. *Genome Announc.* **2013**, *1*, e00830-13. [CrossRef]

124. Mamanova, L.; Coffey, A.J.; Scott, C.E.; Kozarewa, I.; Turner, E.H.; Kumar, A.; Howard, E.; Shendure, J.; Turner, D.J. Target-enrichment strategies for next-generation sequencing. *Nat. Methods* **2010**, *7*, 111–118. [CrossRef]

125. Wylie, K.M.; Wylie, T.N.; Buller, R.; Herter, B.; Cannella, M.T.; Storch, G.A. Detection of Viruses in Clinical Samples by Use of Metagenomic Sequencing and Targeted Sequence Capture. *J. Clin. Microbiol.* **2018**, *56*, e01123-18. [CrossRef]

126. Viral Hemorrhagic Fever Consortium; Metsky, H.C.; Siddle, K.J.; Gladden-Young, A.; Qu, J.; Yang, D.K.; Brehio, P.; Goldfarb, A.; Piantadosi, A.; Wohl, S.; et al. Capturing sequence diversity in metagenomes with comprehensive and scalable probe design. *Nat. Biotechnol.* **2019**, *37*, 160–168. [CrossRef] [PubMed]

127. Cottam, E.M.; Wadsworth, J.; Knowles, N.J.; King, D.P. Full Sequencing of Viral Genomes: Practical Strategies Used for the Amplification and Characterization of Foot-and-Mouth Disease Virus. In *Molecular Epidemiology of Microorganisms*; Caugant, D.A., Ed.; Humana Press: Totowa, NJ, USA, 2009; Volume 551, pp. 217–230. ISBN 978-1-60327-998-7.

128. Gardy, J.L.; Naus, M.; Amlani, A.; Chung, W.; Kim, H.; Tan, M.; Severini, A.; Krajden, M.; Puddicombe, D.; Sahni, V.; et al. Whole-Genome Sequencing of Measles Virus Genotypes H1 and D8 During Outbreaks of Infection Following the 2010 Olympic Winter Games Reveals Viral Transmission Routes. *J. Infect. Dis.* **2015**, *212*, 1574–1578. [CrossRef] [PubMed]

129. Singh, R.K.; Dhama, K.; Karthik, K.; Tiwari, R.; Khandia, R.; Munjal, A.; Iqbal, H.M.N.; Malik, Y.S.; Bueno-Marí, R. Advances in Diagnosis, Surveillance, and Monitoring of Zika Virus: An Update. *Front. Microbiol.* **2018**, *8*, 2677. [CrossRef] [PubMed]

130. Cotten, M.; Petrova, V.; Phan, M.V.T.; Rabaa, M.A.; Watson, S.J.; Ong, S.H.; Kellam, P.; Baker, S. Deep Sequencing of Norovirus Genomes Defines Evolutionary Patterns in an Urban Tropical Setting. *J. Virol.* **2014**, *88*, 11056–11069. [CrossRef] [PubMed]

131. Kundu, S.; Lockwood, J.; Depledge, D.P.; Chaudhry, Y.; Aston, A.; Rao, K.; Hartley, J.C.; Goodfellow, I.; Breuer, J. Next-Generation Whole Genome Sequencing Identifies the Direction of Norovirus Transmission in Linked Patients. *Clin. Infect. Dis.* **2013**, *57*, 407–414. [CrossRef] [PubMed]

132. Kireev, D.E.; Lopatukhin, A.E.; Murzakova, A.V.; Pimkina, E.V.; Speranskaya, A.S.; Neverov, A.D.; Fedonin, G.G.; Fantin, Y.S.; Shipulin, G.A. Evaluating the accuracy and sensitivity of detecting minority HIV-1 populations by Illumina next-generation sequencing. *J. Virol. Methods* **2018**, *261*, 40–45. [CrossRef]

133. Watson, S.J.; Langat, P.; Reid, S.M.; Lam, T.T.-Y.; Cotten, M.; Kelly, M.; Van Reeth, K.; Qiu, Y.; Simon, G.; Bonin, E.; et al. Molecular Epidemiology and Evolution of Influenza Viruses Circulating within European Swine between 2009 and 2013. *J. Virol.* **2015**, *89*, 9920–9931. [CrossRef]

134. Parameswaran, P.; Charlebois, P.; Tellez, Y.; Nunez, A.; Ryan, E.M.; Malboeuf, C.M.; Levin, J.Z.; Lennon, N.J.; Balmaseda, A.; Harris, E.; et al. Genome-Wide Patterns of Intrahuman Dengue Virus Diversity Reveal Associations with Viral Phylogenetic Clade and Interhost Diversity. *J. Virol.* **2012**, *86*, 8546–8558. [CrossRef]

135. Newman, R.M.; Kuntzen, T.; Weiner, B.; Berical, A.; Charlebois, P.; Kuiken, C.; Murphy, D.G.; Simmonds, P.; Bennett, P.; Lennon, N.J.; et al. Whole Genome Pyrosequencing of Rare Hepatitis C Virus Genotypes Enhances Subtype Classification and Identification of Naturally Occurring Drug Resistance Variants. *J. Infect. Dis.* **2013**, *208*, 17–31. [CrossRef]

136. Miia, J.-V.; Tiina, N.; Tarja, S.; Olli, V.; Liisa, S.; Anita, H. Evolutionary trends of European bat lyssavirus type 2 including genetic characterization of Finnish strains of human and bat origin 24 years apart. *Arch. Virol.* **2015**, *160*, 1489–1498. [CrossRef]

137. Chrzastek, K.; Lee, D.; Smith, D.; Sharma, P.; Suarez, D.L.; Pantin-Jackwood, M.; Kapczynski, D.R. Use of Sequence-Independent, Single-Primer-Amplification (SISPA) for rapid detection, identification, and characterization of avian RNA viruses. *Virology* **2017**, *509*, 159–166. [CrossRef] [PubMed]

138. 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] [PubMed]

139. Safonova, M.V.; Shchelkanov, M.Y.; Khafizov, K.; Matsvay, A.D.; Ayginin, A.A.; Dolgova, A.S.; Shchelkanov, E.M.; Pimkina, E.V.; Speranskaya, A.S.; Galkina, I.V.; et al. Sequencing and genetic characterization of two strains Paramushir virus obtained from the Tyuleniy Island in the Okhotsk Sea (2015). *Ticks Tick-Borne Dis.* **2019**, *10*, 269–279. [CrossRef] [PubMed]

140. Bialasiewicz, S.; McVernon, J.; Nolan, T.; Lambert, S.B.; Zhao, G.; Wang, D.; Nissen, M.D.; Sloots, T.P. Detection of a divergent Parainfluenza 4 virus in an adult patient with influenza like illness using next-generation sequencing. *BMC Infect. Dis.* **2014**, *14*, 275. [CrossRef] [PubMed]

141. Brown, J.R.; Roy, S.; Ruis, C.; Yara Romero, E.; Shah, D.; Williams, R.; Breuer, J. Norovirus Whole-Genome Sequencing by SureSelect Target Enrichment: A Robust and Sensitive Method. *J. Clin. Microbiol.* **2016**, *54*, 2530–2537. [CrossRef]

142. Cristescu, M.E. From barcoding single individuals to metabarcoding biological communities: Towards an integrative approach to the study of global biodiversity. *Trends Ecol. Evol.* **2014**, *29*, 566–571. [CrossRef]

143. DeSalle, R.; Goldstein, P. Review and Interpretation of Trends in DNA Barcoding. *Front. Ecol. Evol.* **2019**, *7*, 302. [CrossRef]

144. Deng, X.; Achari, A.; Federman, S.; Yu, G.; Somasekar, S.; Bártolo, I.; Yagi, S.; Mbala-Kingebeni, P.; Kapetshi, J.; Ahuka-Mundeke, S.; et al. Metagenomic sequencing with spiked primer enrichment for viral diagnostics and genomic surveillance. *Nat. Microbiol.* **2020**. [CrossRef]

145. Ayginin, A.A.; Pimkina, E.V.; Matsvay, A.D.; Speranskaya, A.S.; Safonova, M.V.; Blinova, E.A.; Artyushin, I.V.; Dedkov, V.G.; Shipulin, G.A.; Khafizov, K. The Study of Viral RNA Diversity in Bird Samples Using De Novo Designed Multiplex Genus-Specific Primer Panels. *Adv. Virol.* **2018**, *2018*, 1–10. [CrossRef]

146. Dessilly, G.; Goeminne, L.; Vandenbroucke, A.; Dufrasne, F.E.; Martin, A.; Kabamba-Mukabi, B. First evaluation of the Next-Generation Sequencing platform for the detection of HIV-1 drug resistance mutations in Belgium. *PLoS ONE* **2018**, *13*, e0209561. [CrossRef]

147. McKeegan, K.S.; Borges-Walmsley, M.I.; Walmsley, A.R. Microbial and viral drug resistance mechanisms. *Trends Microbiol.* **2002**, *10*, s8–s14. [CrossRef]

148. Aanensen, D.M.; Feil, E.J.; Holden, M.T.G.; Dordel, J.; Yeats, C.A.; Fedosejev, A.; Goater, R.; Castillo-Ramírez, S.; Corander, J.; Colijn, C.; et al. Whole-Genome Sequencing for Routine Pathogen Surveillance in Public Health: A Population Snapshot of Invasive *Staphylococcus aureus* in Europe. *mBio* **2016**, *7*, e00444-16. [CrossRef] [PubMed]

149. Lu, R.; Zhao, X.; Li, J.; Niu, P.; Yang, B.; Wu, H.; Wang, W.; Song, H.; Huang, B.; Zhu, N.; et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: Implications for virus origins and receptor binding. *Lancet* **2020**. [CrossRef]

150. Mbisa, J.L.; Fearnhill, E.; Dunn, D.T.; Pillay, D.; Asboe, D.; Cane, P.A. Evidence of Self-Sustaining Drug Resistant HIV-1 Lineages Among Untreated Patients in the United Kingdom. *Clin. Infect. Dis.* **2015**, *61*, 829–836. [CrossRef] [PubMed]

151. Bartha, I.; Carlson, J.M.; Brumme, C.J.; McLaren, P.J.; Brumme, Z.L.; John, M.; Haas, D.W.; Martinez-Picado, J.; Dalmau, J.; López-Galíndez, C.; et al. A genome-to-genome analysis of associations between human genetic variation, HIV-1 sequence diversity, and viral control. *eLife* **2013**, *2*, e01123. [CrossRef] [PubMed]

152. Power, R.A.; Davaniah, S.; Derache, A.; Wilkinson, E.; Tanser, F.; Gupta, R.K.; Pillay, D.; de Oliveira, T. Genome-Wide Association Study of HIV Whole Genome Sequences Validated using Drug Resistance. *PLoS ONE* **2016**, *11*, e0163746. [CrossRef] [PubMed]

153. Artimo, P.; Jonnalagedda, M.; Arnold, K.; Baratin, D.; Csardi, G.; de Castro, E.; Duvaud, S.; Flegel, V.; Fortier, A.; Gasteiger, E.; et al. ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res.* **2012**, *40*, W597–W603. [CrossRef]

154. 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]

155. Niu, B.; Zhu, Z.; Fu, L.; Wu, S.; Li, W. FR-HIT, a very fast program to recruit metagenomic reads to homologous reference genomes. *Bioinformatics* **2011**, *27*, 1704–1705. [CrossRef]

156. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359. [CrossRef]

157. Li, H.; Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinforma. Oxf. Engl.* **2010**, *26*, 589–595. [CrossRef] [PubMed]

158. Buchfink, B.; Xie, C.; Huson, D.H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **2015**, *12*, 59–60. [CrossRef] [PubMed]

159. Chen, Y.; Yao, H.; Thompson, E.J.; Tannir, N.M.; Weinstein, J.N.; Su, X. VirusSeq: Software to identify viruses and their integration sites using next-generation sequencing of human cancer tissue. *Bioinformatics* **2013**, *29*, 266–267. [CrossRef] [PubMed]

160. UFO Sequencing Consortium within the I-BFM Study Group; Forster, M.; Szymczak, S.; Ellinghaus, D.; Hemmrich, G.; Rühlemann, M.; Kraemer, L.; Mucha, S.; Wienbrandt, L.; Stanulla, M.; et al. Vy-PER: Eliminating false positive detection of virus integration events in next generation sequencing data. *Sci. Rep.* **2015**, *5*, 11534. [CrossRef] [PubMed]

161. Petty, T.J.; Cordey, S.; Padioleau, I.; Docquier, M.; Turin, L.; Preynat-Seauve, O.; Zdobnov, E.M.; Kaiser, L. Comprehensive Human Virus Screening Using High-Throughput Sequencing with a User-Friendly Representation of Bioinformatics Analysis: A Pilot Study. *J. Clin. Microbiol.* **2014**, *52*, 3351–3361. [CrossRef]

162. Langmead, B.; Trapnell, C.; Pop, M.; Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **2009**, *10*, R25. [CrossRef]

163. Kent, W.J. BLAT—The BLAST-Like Alignment Tool. *Genome Res.* **2002**, *12*, 656–664. [CrossRef]

164. Grabherr, M.G.; Haas, B.J.; Yassour, M.; Levin, J.Z.; Thompson, D.A.; Amit, I.; Adiconis, X.; Fan, L.; Raychowdhury, R.; Zeng, Q.; et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **2011**, *29*, 644–652. [CrossRef]

165. Zhao, G.; Krishnamurthy, S.; Cai, Z.; Popov, V.L.; Travassos da Rosa, A.P.; Guzman, H.; Cao, S.; Virgin, H.W.; Tesh, R.B.; Wang, D. Identification of Novel Viruses Using VirusHunter – an Automated Data Analysis Pipeline. *PLoS ONE* **2013**, *8*, e78470. [CrossRef]

166. Lee, W.-P.; Stromberg, M.P.; Ward, A.; Stewart, C.; Garrison, E.P.; Marth, G.T. MOSAIK: A Hash-Based Algorithm for Accurate Next-Generation Sequencing Short-Read Mapping. *PLoS ONE* **2014**, *9*, e90581. [CrossRef]

167. Kostic, A.D.; Ojesina, A.I.; Pedamallu, C.S.; Jung, J.; Verhaak, R.G.W.; Getz, G.; Meyerson, M. PathSeq: Software to identify or discover microbes by deep sequencing of human tissue. *Nat. Biotechnol.* **2011**, *29*, 393–396. [CrossRef] [PubMed]

168. Naccache, S.N.; Federman, S.; Veeraraghavan, N.; Zaharia, M.; Lee, D.; Samayoa, E.; Bouquet, J.; Greninger, A.L.; Luk, K.-C.; Enge, B.; et al. A cloud-compatible bioinformatics pipeline for ultrarapid pathogen identification from next-generation sequencing of clinical samples. *Genome Res.* **2014**, *24*, 1180–1192. [CrossRef] [PubMed]

169. Fonseca, N.A.; Rung, J.; Brazma, A.; Marioni, J.C. Tools for mapping high-throughput sequencing data. *Bioinformatics* **2012**, *28*, 3169–3177. [CrossRef] [PubMed]

170. Pickett, B.E.; Sadat, E.L.; Zhang, Y.; Noronha, J.M.; Squires, R.B.; Hunt, V.; Liu, M.; Kumar, S.; Zaremba, S.; Gu, Z.; et al. ViPR: An open bioinformatics database and analysis resource for virology research. *Nucleic Acids Res.* **2012**, *40*, D593–D598. [CrossRef]

171. Goodacre, N.; Aljanahi, A.; Nandakumar, S.; Mikailov, M.; Khan, A.S. A Reference Viral Database (RVDB) To Enhance Bioinformatics Analysis of High-Throughput Sequencing for Novel Virus Detection. *mSphere* **2018**, *3*, e00069-18. [CrossRef]

172. Stano, M.; Beke, G.; Klucar, L. viruSITE—integrated database for viral genomics. *Database* **2016**, *2016*, baw162. [CrossRef]

173. El-Gebali, S.; Mistry, J.; Bateman, A.; Eddy, S.R.; Luciani, A.; Potter, S.C.; Qureshi, M.; Richardson, L.J.; Salazar, G.A.; Smart, A.; et al. The Pfam protein families database in 2019. *Nucleic Acids Res.* **2019**, *47*, D427–D432. [CrossRef]

174. The UniProt Consortium UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Res.* **2019**, *47*, D506–D515. [CrossRef]

175. Ari, Ş.; Arikan, M. Next-Generation Sequencing: Advantages, Disadvantages, and Future. In *Plant Omics: Trends and Applications*; Hakeem, K.R., Tombuloğlu, H., Tombuloğlu, G., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 109–135. ISBN 978-3-319-31701-4.

176. Břinda, K.; Callendrello, A.; Ma, K.C.; MacFadden, D.R.; Charalampous, T.; Lee, R.S.; Cowley, L.; Wadsworth, C.B.; Grad, Y.H.; Kucherov, G.; et al. Rapid heuristic inference of antibiotic resistance and susceptibility by genomic neighbor typing. *Bioinformatics* **2018**. [CrossRef]

177. Ueda, Y. New sequencing technology, third-generation Nanopore sequencing, for hepatitis C virus research. *Hepatol. Res.* **2019**. [CrossRef]

178. Astbury, S.; Soares, M.M.D.C.N.; Peprah, E.; King, B.J.; Jardim, A.C.G.; Shimizu, J.F.; Jalal, P.J.; Saeed, C.H.; Sabeer, F.T.; Irving, W.L.; et al. Extraction-free direct PCR from dried serum spots permits HBV genotyping and RAS identification by Sanger and minION sequencing. *Microbiology* **2019**. [CrossRef]

179. Kafetzopoulou, L.E.; Efthymiadis, K.; Lewandowski, K.; Crook, A.; Carter, D.; Osborne, J.; Aarons, E.; Hewson, R.; Hiscox, J.A.; Carroll, M.W.; et al. Assessment of metagenomic Nanopore and Illumina sequencing for recovering whole genome sequences of chikungunya and dengue viruses directly from clinical samples. *Eurosurveillance* **2018**, *23*. [CrossRef] [PubMed]

180. Faria, N.R.; Quick, J.; Claro, I.M.; Thézé, J.; de Jesus, J.G.; Giovanetti, M.; Kraemer, M.U.G.; 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]

181. Batovska, J.; Lynch, S.E.; Rodoni, B.C.; Sawbridge, T.I.; Cogan, N.O. Metagenomic arbovirus detection using MinION nanopore sequencing. *J. Virol. Methods* **2017**, *249*, 79–84. [CrossRef]

182. Tyler, A.D.; Mataseje, L.; Urfano, C.J.; Schmidt, L.; Antonation, K.S.; Mulvey, M.R.; Corbett, C.R. Evaluation of Oxford Nanopore's MinION Sequencing Device for Microbial Whole Genome Sequencing Applications. *Sci. Rep.* **2018**, *8*, 10931. [CrossRef]

183. Wawina-Bokalanga, T.; Vanmechelen, B.; Martí-Carreras, J.; Vergote, V.; Vermeire, K.; Muyembe-Tamfum, J.-J.; Ahuka-Mundeke, S.; Maes, P. Complete Genome Sequence of a New Ebola Virus Strain Isolated during the 2017 Likati Outbreak in the Democratic Republic of the Congo. *Microbiol. Resour. Announc.* **2019**, *8*, e00360-19. [CrossRef]

184. Naito, F.Y.B.; Melo, F.L.; Fonseca, M.E.N.; Santos, C.A.F.; Chanes, C.R.; Ribeiro, B.M.; Gilbertson, R.L.; Boiteux, L.S.; de Cássia Pereira-Carvalho, R. Nanopore sequencing of a novel bipartite New World begomovirus infecting cowpea. *Arch. Virol.* **2019**, *164*, 1907–1910. [CrossRef]

185. Vanmechelen, B.; Bertelsen, M.F.; Rector, A.; Van den Oord, J.J.; Laenen, L.; Vergote, V.; Maes, P. Identification of a novel species of papillomavirus in giraffe lesions using nanopore sequencing. *Vet. Microbiol.* **2017**, *201*, 26–31. [CrossRef]

186. Hoenen, T.; Groseth, A.; Rosenke, K.; Fischer, R.J.; Hoenen, A.; Judson, S.D.; Martellaro, C.; Falzarano, D.; Marzi, A.; Squires, R.B.; et al. Nanopore Sequencing as a Rapidly Deployable Ebola Outbreak Tool. *Emerg. Infect. Dis.* **2016**, *22*. [CrossRef]

187. Kilianski, A.; Roth, P.A.; Liem, A.T.; Hill, J.M.; Willis, K.L.; Rossmaier, R.D.; Marinich, A.V.; Maughan, M.N.; Karavis, M.A.; Kuhn, J.H.; et al. Use of Unamplified RNA/cDNA–Hybrid Nanopore Sequencing for Rapid Detection and Characterization of RNA Viruses. *Emerg. Infect. Dis.* **2016**, *22*, 1448–1451. [CrossRef]

188. Hill, S.C.; de Vasconcelos, J.N.; Granja, B.G.; Thézé, J.; Jandondo, D.; Neto, Z.; Mirandela, M.; dos Santos Sebastião, C.; Cândido, A.L.M.; Clemente, C.; et al. Early Genomic Detection of Cosmopolitan Genotype of Dengue Virus Serotype 2, Angola, 2018. *Emerg. Infect. Dis.* **2019**, *25*, 784–787. [CrossRef] [PubMed]

189. Check Hayden, E. Pint-sized DNA sequencer impresses first users. *Nature* **2015**, *521*, 15–16. [CrossRef] [PubMed]

190. Croville, G.; Le Loc'h, G.; Zanchetta, C.; Manno, M.; Camus-Bouclainville, C.; Klopp, C.; Delverdier, M.; Lucas, M.-N.; Donnadieu, C.; Delpont, M.; et al. Rapid whole-genome based typing and surveillance of avipoxviruses using nanopore sequencing. *J. Virol. Methods* **2018**, *261*, 34–39. [CrossRef] [PubMed]

191. Russell, J.A.; Campos, B.; Stone, J.; Blosser, E.M.; Burkett-Cadena, N.; Jacobs, J.L. Unbiased Strain-Typing of Arbovirus Directly from Mosquitoes Using Nanopore Sequencing: A Field-forward Biosurveillance Protocol. *Sci. Rep.* **2018**, *8*, 5417. [CrossRef] [PubMed]

192. Garalde, D.R.; Snell, E.A.; Jachimowicz, D.; Sipos, B.; Lloyd, J.H.; Bruce, M.; Pantic, N.; Admassu, T.; James, P.; Warland, A.; et al. Highly parallel direct RNA sequencing on an array of nanopores. *Nat. Methods* **2018**, *15*, 201–206. [CrossRef] [PubMed]

193. Keller, M.W.; Rambo-Martin, B.L.; Wilson, M.M.; Ridenour, C.A.; Shepard, S.S.; Stark, T.J.; Neuhaus, E.B.; Dugan, V.G.; Wentworth, D.E.; Barnes, J.R. Direct RNA Sequencing of the Coding Complete Influenza A Virus Genome. *Sci. Rep.* **2018**, *8*, 14408. [CrossRef]

194. Jenjaroenpun, P.; Wongsurawat, T.; Pereira, R.; Patumcharoenpol, P.; Ussery, D.W.; Nielsen, J.; Nookaew, I. Complete genomic and transcriptional landscape analysis using third-generation sequencing: A case study of Saccharomyces cerevisiae CEN.PK113-7D. *Nucleic Acids Res.* **2018**, *46*, e38. [CrossRef]

195. Depledge, D.P.; Srinivas, K.P.; Sadaoka, T.; Bready, D.; Mori, Y.; Placantonakis, D.G.; Mohr, I.; Wilson, A.C. Direct RNA sequencing on nanopore arrays redefines the transcriptional complexity of a viral pathogen. *Nat. Commun.* **2019**, *10*, 754. [CrossRef]

196. Boldogkői, Z.; Moldován, N.; Balázs, Z.; Snyder, M.; Tombácz, D. Long-Read Sequencing—A Powerful Tool in Viral Transcriptome Research. *Trends Microbiol.* **2019**, *27*, 578–592. [CrossRef]

197. Xu, Y.; Lewandowski, K.; Lumley, S.; Pullan, S.; Vipond, R.; Carroll, M.; Foster, D.; Matthews, P.C.; Peto, T.; Crook, D. Detection of Viral Pathogens With Multiplex Nanopore MinION Sequencing: Be Careful With Cross-Talk. *Front. Microbiol.* **2018**, *9*, 2225. [CrossRef]

198. Erlwein, O.; Robinson, M.J.; Dustan, S.; Weber, J.; Kaye, S.; McClure, M.O. DNA Extraction Columns Contaminated with Murine Sequences. *PLoS ONE* **2011**, *6*, e23484. [CrossRef] [PubMed]

199. Karamitros, T.; Magiorkinis, G. A novel method for the multiplexed target enrichment of MinION next generation sequencing libraries using PCR-generated baits. *Nucleic Acids Res.* **2015**, *43*, e152. [CrossRef] [PubMed]

200. Vinaya Kumar, K.; Shekhar, M.S.; Otta, S.K.; Karthic, K.; Ashok Kumar, J.; Gopikrishna, G.; Vijayan, K.K. First Report of a Complete Genome Sequence of *White spot syndrome virus* from India. *Genome Announc.* **2018**, *6*, e00055-18. [CrossRef] [PubMed]

201. Günther, T.; Haas, L.; Alawi, M.; Wohlsein, P.; Marks, J.; Grundhoff, A.; Becher, P.; Fischer, N. Recovery of the first full-length genome sequence of a parapoxvirus directly from a clinical sample. *Sci. Rep.* **2017**, *7*, 3734. [CrossRef]

202. Warwick-Dugdale, J.; Solonenko, N.; Moore, K.; Chittick, L.; Gregory, A.C.; Allen, M.J.; Sullivan, M.B.; Temperton, B. Long-read viral metagenomics captures abundant and microdiverse viral populations and their niche-defining genomic islands. *PeerJ* **2019**, *7*, e6800. [CrossRef]

203. Peserico, A.; Marcacci, M.; Malatesta, D.; Di Domenico, M.; Pratelli, A.; Mangone, I.; D'Alterio, N.; Pizzurro, F.; Cirone, F.; Zaccaria, G.; et al. Diagnosis and characterization of canine distemper virus through sequencing by MinION nanopore technology. *Sci. Rep.* **2019**, *9*, 1714. [CrossRef]

204. Theuns, S.; Vanmechelen, B.; Bernaert, Q.; Deboutte, W.; Vandenhole, M.; Beller, L.; Matthijnssens, J.; Maes, P.; Nauwynck, H.J. Nanopore sequencing as a revolutionary diagnostic tool for porcine viral enteric disease complexes identifies porcine kobuvirus as an important enteric virus. *Sci. Rep.* **2018**, *8*, 9830. [CrossRef]

205. Bronzato Badial, A.; Sherman, D.; Stone, A.; Gopakumar, A.; Wilson, V.; Schneider, W.; King, J. Nanopore Sequencing as a Surveillance Tool for Plant Pathogens in Plant and Insect Tissues. *Plant Dis.* **2018**, *102*, 1648–1652. [CrossRef]

206. Gallagher, M.D.; Matejusova, I.; Nguyen, L.; Ruane, N.M.; Falk, K.; Macqueen, D.J. Nanopore sequencing for rapid diagnostics of salmonid RNA viruses. *Sci. Rep.* **2018**, *8*, 16307. [CrossRef]

207. Ameur, A.; Kloosterman, W.P.; Hestand, M.S. Single-Molecule Sequencing: Towards Clinical Applications. *Trends Biotechnol.* **2019**, *37*, 72–85. [CrossRef]

208. Houldcroft, C.J.; Breuer, J. Tales from the crypt and coral reef: The successes and challenges of identifying new herpesviruses using metagenomics. *Front. Microbiol.* **2015**, *6*. [CrossRef] [PubMed]

209. Munro, A.C.; Houldcroft, C. Human cancers and mammalian retroviruses: Should we worry about bovine leukemia virus? *Future Virol.* **2016**, *11*, 163–166. [CrossRef]

210. Hue, S.; Gray, E.R.; Gall, A.; Katzourakis, A.; Tan, C.P.; Houldcroft, C.J.; McLaren, S.; Pillay, D.; Futreal, A.; Garson, J.A.; et al. Disease-associated XMRV sequences are consistent with laboratory contamination. *Retrovirology* **2010**, *7*, 111. [CrossRef] [PubMed]

211. Naccache, S.N.; Greninger, A.L.; Lee, D.; Coffey, L.L.; Phan, T.; Rein-Weston, A.; Aronsohn, A.; Hackett, J.; Delwart, E.L.; Chiu, C.Y. The Perils of Pathogen Discovery: Origin of a Novel Parvovirus-Like Hybrid Genome Traced to Nucleic Acid Extraction Spin Columns. *J. Virol.* **2013**, *87*, 11966–11977. [CrossRef]

212. Salter, S.J.; Cox, M.J.; Turek, E.M.; Calus, S.T.; Cookson, W.O.; Moffatt, M.F.; Turner, P.; Parkhill, J.; Loman, N.J.; Walker, A.W. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol.* **2014**, *12*, 87. [CrossRef] [PubMed]

213. Matsvay, A.D.; Alborova, I.E.; Pimkina, E.V.; Markelov, M.L.; Khafizov, K.; Mustafin, K.K. Experimental approaches for ancient DNA extraction and sample preparation for next generation sequencing in ultra-clean conditions. *Conserv. Genet. Resour.* **2019**, *11*, 345–353. [CrossRef]

214. Morfopoulou, S.; Brown, J.R.; Davies, E.G.; Anderson, G.; Virasami, A.; Qasim, W.; Chong, W.K.; Hubank, M.; Plagnol, V.; Desforges, M.; et al. Human Coronavirus OC43 Associated with Fatal Encephalitis. *N. Engl. J. Med.* **2016**, *375*, 497–498. [CrossRef]

215. Lipkin, W.I. A Vision for Investigating the Microbiology of Health and Disease. *J. Infect. Dis.* **2015**, *212*, S26–S30. [CrossRef]

216. Hall, R.J.; Draper, J.L.; Nielsen, F.G.G.; Dutilh, B.E. Beyond research: A primer for considerations on using viral metagenomics in the field and clinic. *Front. Microbiol.* **2015**, *6*. [CrossRef]

217. Shafer, R.W. Rationale and Uses of a Public HIV Drug-Resistance Database. *J. Infect. Dis.* **2006**, *194*, S51–S58. [CrossRef]

218. Gnaneshan, S.; Ijaz, S.; Moran, J.; Ramsay, M.; Green, J. HepSEQ: International Public Health Repository for Hepatitis B. *Nucleic Acids Res.* **2007**, *35*, D367–D370. [CrossRef] [PubMed]

219. Rhee, S.-Y.; Margeridon-Thermet, S.; Nguyen, M.H.; Liu, T.F.; Kagan, R.M.; Beggel, B.; Verheyen, J.; Kaiser, R.; Shafer, R.W. Hepatitis B virus reverse transcriptase sequence variant database for sequence analysis and mutation discovery. *Antivir. Res.* **2010**, *88*, 269–275. [CrossRef] [PubMed]

220. Kuiken, C.; Yusim, K.; Boykin, L.; Richardson, R. The Los Alamos hepatitis C sequence database. *Bioinformatics* **2005**, *21*, 379–384. [CrossRef] [PubMed]