Application of next-generation sequencing to identify different pathogens

Aljuboori M. Nafea[1,2†], Yuer Wang[1†], Duanyang Wang[1], Ahmed M. Salama[3,4], Manal A. Aziz[2], Shan Xu[1]* and Yigang Tong[1]*

[1]College of Life Science and Technology, Beijing University of Chemical Technology, Beijing, China, [2]College of Medicine, Department of Microbiology, Ibn Sina University of Medical and Pharmaceutical Science, Baghdad, Iraq, [3]State Key Laboratory of Chemical Resource Engineering, Beijing University of Chemical Technology, Beijing, China, [4]Medical Laboratory at Sharkia Health Directorate, Ministry of Health, Sharkia, Egypt

Early and precise detection and identification of various pathogens are essential for epidemiological monitoring, disease management, and reducing the prevalence of clinical infectious diseases. Traditional pathogen detection techniques, which include mass spectrometry, biochemical tests, molecular testing, and culture-based methods, are limited in application and are time-consuming. Next generation sequencing (NGS) has emerged as an essential technology for identifying pathogens. NGS is a cutting-edge sequencing method with high throughput that can create massive volumes of sequences with a broad application prospects in the field of pathogen identification and diagnosis. In this review, we introduce NGS technology in detail, summarizes the application of NGS in that identification of different pathogens, including bacteria, fungi, and viruses, and analyze the challenges and outlook for using NGS to identify clinical pathogens. Thus, this work provides a theoretical basis for NGS studies and provides evidence to support the application of NGS in distinguishing various clinical pathogens.

KEYWORDS

Sanger, next generation sequencing, pathogens, bacteria, fungi

Introduction

For patients to be treated effectively, identification of the microorganisms that cause the infection is essential. The traditional pathogen identification methods include culture method, serological detection, and molecular biology methods (such as nucleic acid amplification). Nevertheless, not all bacterial species can be effectively cultured in the diagnostic laboratory, and novel pathogens cannot be detected by nucleic acid amplification. Some unknown pathogens can swiftly spark hospital epidemics that put patients in danger while they are being treated. DNA sequencing has made great advances in various disciplines, and specific progress has been made in personalized therapies (Jiang et al., 2022). The process of DNA sequencing can be used to determine the exact order of the nucleotide bases (adenine, cytosine, guanine, and thymine). Since the emergence of Frederick Sanger's pioneering work in the 1970s, when he used the "plus and minus method" to sequence the first complete genome, DNA sequencing technology

has been progressing quickly. Eventually, the Sanger chain termination or dideoxy technique, which was first described in 1977, laid the groundwork for the swift advancement of DNA sequencing technologies and made it possible to sequence the human genome for the first time in 2001 (Sanger and Coulson Nicklen, 1977). Sanger sequencing continues to be a popular technique, particularly for examining a small number of DNA sequences, as it offers high-quality DNA sequencing information for areas up to 1,000 bases. Sanger sequencing laid the foundation for the development of next-generation sequencing (NGS) technology.

The need for large-scale sequencing has quickly led to the development of NGS. Pyrosequencing, reversible-dye terminator, and proton detection are examples of the various NGS systems and methods that are created over time and based on unique chemistries and detection techniques (Garrido-Cardenas et al., 2017). NGS can be used to detect several infections. In contrast to conventional sequencing techniques, which allowed for the sequencing of one or a few very short DNA fragments that had previously been amplified by polymerase chain reaction (PCR), this cutting-edge technology has been a true revolution. NGS applications are increasingly prevalent, and they have evolved from study tools to diagnostic techniques. This review first presents detailed descriptions of Sanger and second-generation sequencing, which are the most used sequencing methods, in addition to a brief overview of the rapidly developing third-generation sequencing. Then, the application of NGS, including whole-genome sequencing (WGS), targeted next-generation sequencing (tNGS), and metagenomic next-generation sequencing (mNGS) (Mitchell and Simner, 2019), in clinical disease diagnosis is introduced. In addition, we focus on the clinical application of NGS for the identification of different pathogens, such as bacteria, fungi, and viruses. Finally, we discuss the challenges and the outlook of NGS in pathogen identification.

Development of sequencing technologies

Sequencing technologies are categorized as first-generation sequencing (for example, Sanger), second-generation sequencing (for example, NGS), and third-generation sequencing (for example, nanopore sequencing).

The first generation consists of the Sanger and Maxam Gilbert procedures, which are two distinct techniques. Based on the chain-termination method, Sanger sequencing is more widely used. The developing chain in the chain termination method ends when dideoxynucleotides (ddNTPs) are incorporated. After being run on conventional slab gels, fragments of DNA varying in length (by a single nucleotide) were recovered, and the pattern of bands was used to determine the sequence. Fluorescently labeled ddNTPs (an automated sequencing approach) are used in place of radiolabeling, and the sequence is determined by varying the wavelength of the laser light (Smith et al., 1986). A maximum read length of 800–1,000 bp can be produced with this technique. The length of the sequenced fragment is the result of one run as using this technology only one fragment in a single capillary can be sequenced (Gupta and Verma, 2019).

The second-generation sequencing allows millions of sequencing reactions to occur simultaneously on a single solid surface, such as a glass slide or beads. This approach only requires the reaction to be spatially isolated rather than physically separated in a different well, lane or tube. As a result, thousands of millions of distinct reactions take place at the same time, leading to a significant reduction in both the overall cost and manpower when compared to other conventional approaches. Many commercial NGS systems that are based on various technologies but generally follow a basic pattern or steps have been developed. The most significant benefit of NGS is the capacity to extract sequence information from individual DNA fragments in a library, which does not require large amounts of DNA/RNA. Moreover, NGS allows for *de novo* assembly that does not rely on references or amplification (Sohn and Nam, 2018). Therefore, NGS can be used to identify unknown pathogens. However, NGS reads are short and require different computational methods to analyze the data. The price of NGS in China ranges from several thousand RMB to tens of thousands of RMB. In the United States, NGS costs approximately $99 for non-invasive prenatal testing and $2,500 for exome sequencing (Phillips et al., 2020). The use of NGS in clinical patients will become more common as the cost of NGS decreases and health insurance reimbursement increases. These techniques have significantly contributed to the study in several areas of life science and are being introduced in clinical laboratories more frequently, with several diagnostic uses in the fields of pathogen identification, oncology, and human genetics (Rhoads and Kin Fai, 2015; Yamada and Nomura, 2020).

Third-generation sequencing methods can be used to sequence individual DNA molecules in real time without the need for an amplification step. These methods have made sample preparation simpler and can provide single runs. Moreover, third-generation methods often yield larger reads, roughly a few kilobases in length, which addresses the challenges in read assembly. Moreover, significantly, this method can be used to identify different pathogens. The first nanopore sequencing instrument was called MinION, was licensed in 2007 by Oxford Nanopore Technologies, United Kingdom, and went on sale in May 2014. The core of this device contains a flow cell with 2048 individual nanopores that are organized into four groups, with 512 nanopores per group, and the nanopores are managed by an application-specific integrated circuit (ASIC). The following is a quick summary of the sequencing process: the adapters are ligated to either end of the fragments and adapters facilitate polymerase binding at the 5′ ends of the fragments and allow fragment capture. Furthermore, by concentrating the DNA fragments near the nanopore, these adapters increase the fragment capture rate one thousandfold. Additionally, by covalently binding the complementary strands to one another, these hairpin-like adapters enable the adjoined sequencing of two strands. The polymerase moves along the template strand upon fragment translocation via the nanopore, and the procedure is repeated for the complementary strand. As pieces pass through the nanopore, the sensor detects the shift in ionic charge. To ensure the corresponding duration, mean amplitude, and variation, the change in ionic charge or characteristic disruption in current is split into discrete occurrences. Ultimately, the sequence of events is deciphered by computer tools and graphical models (such as MinKNOW) to determine the nucleotide sequence. The data gathered from the complimentary and template strands are combined to

provide the "2D read"(Jain et al., 2018). The potential of this method to advance pathogen detection and comprehension in many contexts is demonstrated by its capacity to identify specific viruses. For example, the effective handheld MinION sequencer enables the nontargeted detection of Ross River virus (RRV) using a metagenomic technique within a few hours (Batovska et al., 2017). The reads spanning up to 2.5 kb helped identify the virus, with an accuracy rate of more than 98%.

Here, we compare the first, second, and third generation sequencing (Table 1).

Sanger method

The Sanger sequencing method has progressed via automation and commercialization. Although it has a slower sequencing speed than the NGS method, the Sanger method continues to be the best sequencing technique for many applications. The discovery of fluorescent dyes, the use of thermal cycle sequencing, which requires less input DNA, and the creation of thermostable polymerases to accurately and effectively insert terminator colors into the developing DNA strands are the three most important developments in Sanger sequencing (Figure 1) (Crossley et al., 2020). Specific nucleotides with end chains (dideoxy nucleotides) are used in Sanger sequencing as they do not include a 3′-OH group. As a result, DNA polymerase cannot create a phosphodiester bond, which causes the developing DNA chain to
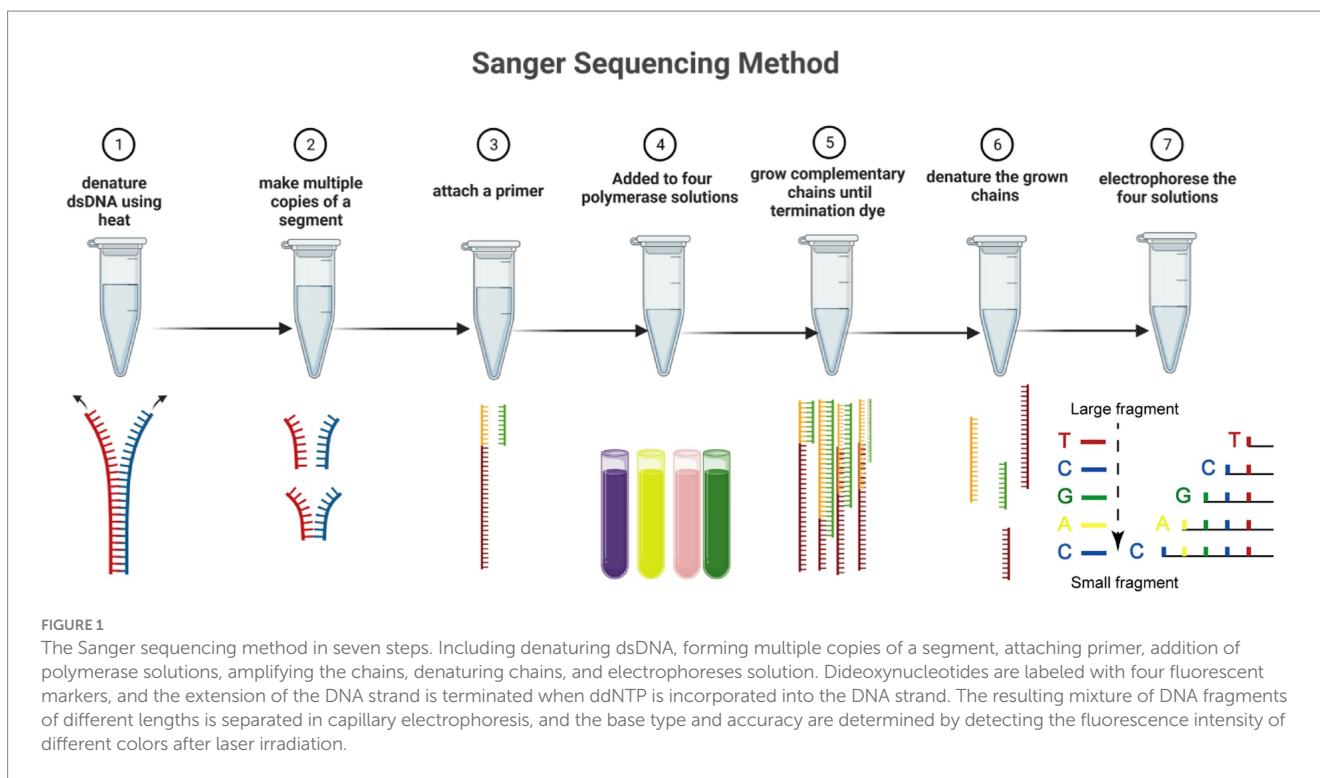
stop at that location. The ddNTPs are fluorescently or radioactively tagged for detection in automated sequencing instruments (Benner et al., 2007).

The most common use of Sanger sequencing technology is single-reaction sequencing operations that employ a particular DNA primer to a given template, such as those used to validate plasmid constructs or polymerase chain reaction products (Crossley et al., 2020). The popular Sanger sequencing technique makes use of time-consuming and low-cost molecular biological products, such as DNA purification reagents and kits, as well as inexpensive, high-quality manufactured primers (Sander et al., 1976). In addition, this technique can also be applied to determine the function of specific enzymes on fluorescently labeled DNA substrates by examining DNA fragment size. Various fluorescent labels, substrates, products, and reaction intermediates can be assessed using capillary electrophoresis in a single experiment. For instance, the kinetics of DNA polymerase and DNA ligase and several coupled enzyme processes, such as the processing of Okazaki fragments and ribonucleotide excision repair, can be detected by these methods. Moreover, researchers have performed high-throughput experiments using capillary electrophoresis (Greenough et al., 2015).

This technique is the most sophisticated for sequencing isolated genes and short tandem repeats. researchers have determined the source of many illness-related genetic mutations with the help of Sanger sequencing. However, the major drawback of this technique is that it is time-consuming because of the low conductivity. This

TABLE 1 Comparison of three sequencing technologies.

	First generation sequencing (Sanger method)	Second generation sequencing (NGS method)	Third generation sequencing (nanopore sequencing)	References
DNA amount	Usually requires a higher quantity of DNA (ranging from micrograms to milligrams), depending on the specific method and application.	Requires a comparatively little quantity of DNA, often ranging from nanograms to micrograms.	Without the need for an amplification step.	Gupta and Verma (2019), Deamer et al. (2016)
Quantification	While correct quantification is vital, the criteria are not as strict as for NGS.	Accurate measurement is essential because of the increased sensitivity and decreased input demands.	Quantitative PCR techniques were used at Harvard to quantify the number of DNA molecules that passed to the trans side of the pore.	Kiehn and Car (2017), Gupta and Verma (2019), Deamer et al. (2016)
Read Length	Produces extended read lengths, frequently reaching up to 1,000 bases or more.	Typically produces shorter read lengths than Sanger sequencing, although modern technologies have made some improvements.	Up to 2.273 megabases or more.	Nowrousian (2010), Gupta and Verma (2019), Wang et al. (2021)
Cost	Increased sequencing expenses.	Reduced expense per sequenced.	Higher costs, about $1,000.	Gupta and Verma (2019), Faulk (2023)
Speed	Lengthier procedure, particularly for large-scale projects.	Parallel processing capabilities result in faster turnaround times, facilitating rapid data creation.	Less time required, about a few hours.	Gupta and Verma (2019), Koch et al. (2023)
Throughput	The throughput is limited because to the sequential processing of individual DNA fragments.	Utilizing high throughput technology, this process allows for the simultaneous sequencing of millions of DNA fragments in parallel.	Increased to ~10–15 gigabases.	Slatko et al. (2018), Wang et al. (2021)

FIGURE 1
The Sanger sequencing method in seven steps. Including denaturing dsDNA, forming multiple copies of a segment, attaching primer, addition of polymerase solutions, amplifying the chains, denaturing chains, and electrophoreses solution. Dideoxynucleotides are labeled with four fluorescent markers, and the extension of the DNA strand is terminated when ddNTP is incorporated into the DNA strand. The resulting mixture of DNA fragments of different lengths is separated in capillary electrophoresis, and the base type and accuracy are determined by detecting the fluorescence intensity of different colors after laser irradiation.

technique only recognizes single-stranded mutations and can process short DNA sequences (300-1,000 base pairs) simultaneously. As a result, the demand for new technology to enable faster throughput sequencing of larger genomes at cheaper costs led to the creation of NGS technologies by using a wide range of inventive approaches (Rehm, 2013).

Next-generation sequencing technology

NGS is distinguished by its fast and high-throughput properties. NGS can be used to obtain general data by deciphering millions of different DNA sequences at the same time (Petersen et al., 2019), making it possible to qualitatively investigate multiple types of genetic alterations (Manuscript and Malignancies, 2014). NGS is parallel sequencing, which can be used to simultaneously assesses multiple genes. NGS offers excellent throughput and speed and produces many sequences in a single run at a relatively low cost. NGS sequencing methods include the 454 Roche method, sequencing by oligonucleotideligation and detection (SOLiD), and Illumina. Here, we will focus on Illumina. The Illumina Solexa DNA sequencing system uses an eight-lane flow cell with oligonucleotide anchors to perform end repair, adenylation, and fragmentation of template DNA. Hybridization is facilitated by ligating adapters to the complementary anchors of the flow cell. "Bridge amplification" creates clusters, and sequencing is accomplished by incorporating fluorescently labeled reversible terminators (Adessi et al., 2000; Guo et al., 2008). Illumina's methods, such as the MiSeq and HiSeq series, are industry leaders in NGS. The most recent models, the HiSeq 3,000 and HiSeq 4,000, use patterned flow cell technology and fall between the HiSeq X Ten and HiSeq 2,500 in terms of data output

and run time (Reuterl et al., 2015). The smallest and most economical sequencer is the HiSeq 100.

Illumina's sequencers are extensively utilized in the advancement of large-scale sequencing endeavors. The reasons for the widespread use of NGS technologies are multifaceted. First, these technologies offer exceptional accuracy in sequencing, ensuring reliable results. Additionally, the cost per gigabyte (Gb) of data obtained through these methods is quite low. Furthermore, the market offers a diverse range of equipment options, thereby allowing researchers to select the most suitable tools for their specific project requirements. The range of sequencing equipment varies from compact bench-top machines with moderate performance, such as MiniSeq, to large-scale instruments utilized for sequencing entire genomes in population-based initiatives, such as HiSeqX (Garrido-Cardenas et al., 2017).

Since NGS is capable of quickly identifying pathogens that threaten public health, it allows for health care workers to take urgent measures according to the type of pathogen to prevent or control large-scale spread. Therefore, NGS is of great significance in the clinical diagnosis of pathogens. For example, Köser et al. compared single nucleotide polymorphisms in clinical outbreak isolates of methicillin-resistant *Staphylococcus aureus* (*S. aureus*) using whole-genome sequencing, thereby providing health care professionals with rapid access to validated clinically relevant data (Köser et al., 2012). In addition, Greninger et al. detected the complete genome of *Balamuthia mandrillaris* in cerebrospinal fluid samples from patients with primary amoebic meningoencephalitis (PAM) using NGS, and the diagnosis was confirmed by the Centers for Disease Control and Prevention (CDC) using PCR (Greninger et al., 2015). NGS can also serve as a powerful tool for identifying virulence factors of pathogens, which can improve public health responses (Gilchrist et al., 2015; Li et al., 2016). In addition, NGS can play a role in infection prevention, with phylogenetic analysis of

isolates facilitating the prevention of large-scale outbreaks of pathogens (Zhou et al., 2016).

Library construction for NGS has been developed and distributed in commercial kits, such as the NEBNext Ultra II Directional RNA Library Prep Kit (Illumina, San Diego, CA, United States), so that health care professionals can build the library according to the manufacturer's instructions, which reduces the difficulty of the process. Taking RNA library building as an example, this process is simply divided into the steps of RNA fragmentation and priming, first-strand cDNA synthesis, second-strand cDNA synthesis, end prep of cDNA library, adaptor ligation, and PCR enrichment of adaptor-ligated DNA. The libraries are sequenced to obtain raw reads stored in fastaq file format, which contains the sequence information of the reads as well as the sequencing quality information. Not all reads are meaningful for analysis, and some short reads without overlap cannot be assembled into contigs. After removing low-quality and short reads, reads with overlapping regions are assembled into contigs and compared to known sequences by blastn and blastx. Subsequently, the complete sequence can be obtained through PCR detection and Sanger sequencing. The obtained sequences are aligned with known sequences, and the maximal likelihood tree or neighbor-joining tree is constructed to determine which branch it clusters with (Podnar et al., 2014; Tian et al., 2023).

Sometimes, only a few reads from pathogen signature sequences are sufficient to identify the pathogen. For example, bacterial species can be identified based on the analysis of 16S rRNA-related reads (Church et al., 2020). 16S rRNA-based analysis facilitates the study of complex microbial communities (Tourlousse et al., 2017; Dai et al., 2022). Jin et al. used single-base accurate cellular barcoded 16S rRNA sequences to identify individual bacteria to study the microbiota (Jin et al., 2022). In addition, Mukherjee et al. showed that 16-23S-based intergenic spacer regions (ISRs) improved the accuracy of bacterial community analysis at the subspecies level compared to 16S rRNA (Mukherjee et al., 2018). Analysis of the sequence of the 18S rRNA gene based on related reads can be used to identify fungi (Petri et al., 2019). For example, Zahedi et al. identified multiple fungi from 49 eukaryotic phyla in wastewater based on 18S rRNA NGS (Zahedi et al., 2019; Figure 2).

The core principle of most NGS technology is sequencing by synthesis (SBS). The DNA molecule to be sequenced is replicated (synthesized) by using DNA polymerases and dNTPs. To determine the base type of the incorporated nucleotide as the DNA molecule extends, nucleotides are either modified with distinguishing tags, such as fluorophores, or via the stimulated fluorescence of other substrates. In addition, NGS can also apply the principle of sequencing by hybridization (SBH). SBH is a method that assembles several overlapping oligonucleotide sequences to identify the DNA sequence of an organism. The foundation of SBH is the renaturing of complementary DNA strands following melting. As a consequence, the oligonucleotide probes hybridize in a way that makes it possible to identify the complementary sequences in the DNA target.

The advantages of NGS technologies promote the development of molecular biology by facilitating large-scale whole-genome sequencing (Pmg et al., 2019). NGS technology is usually used in multilevel studies for genomics, transcriptomics, and epigenomics. Other methods of pathogen identification include PCR-based and microarray assays. PCR-based assays can only be used to identify known pathogens. Microarray technology for pathogen detection differs in methodology depending on the specific pathogens being targeted, the design of the probes, and the method used for the array. The advantages of this approach include its affordability and multiplex analysis (Green and Pass, 2005). Limitations include insufficient probe density, data noise, and substantial initial expenses (Wang et al., 2002, 2003). NGS data is considered more quantifiable than microarray data (Radovich and Ragoussis, 2014; Gaston et al., 2022). Microarrays can be practical for recognizing DNA methylation (Han et al., 2019) and can also be used for non-model organisms. The microarray, however, is limited to classic organisms. Because of the large number of genes, efficient bioinformatics approaches need to be established to assess the significant amount of sequencing information gained in these experiment (Schmieder and Edwards, 2011). At the same time, not all reads are meaningful for analysis, and some short reads without overlap cannot be assembled into contigs (Hung and Weng, 2017). Here, we compare the different processes of the Sanger method and NGS in detecting different pathogens in Figure 2.

Whole-genome sequencing

WGS is the assembly and sequencing of an organism's whole genome and is currently the most frequently used technology to identify unknown organisms. In addition, NGS can be utilized to detect known microorganisms, and other approaches for detecting known microorganisms include mass spectrometry, culture, etc. (Bauermeister et al., 2022). The technological process of WGS differs depending on the type of organism. WGS of viral genomes is usually carried out directly from patient samples and does not call for the culture or isolation of the virus. In contrast, when attempting to detect infectious bacteria in a clinic, the presence of additional bacteria in clinical specimens that are either clinically insignificant, such as normal skin flora, or represent polymicrobial infections may complicate WGS results. Therefore, the bacteria must first be isolated and cultured to prepare for the extraction and sequencing of bacterial nucleic acids using WGS. Thus, WGS has evident limitations when the microorganism is impossible or difficult to cultivate. By monitoring outbreaks, WGS has proven particularly helpful in hospital and public health epidemiology research (Brown et al., 2015). For instance, WGS of clinical *Mycobacterium tuberculosis* (*M. tuberculosis*) has been performed to diagnose tuberculosis. In addition, WGS was able to identify and track the spread of *Klebsiella pneumoniae* isolates that produced CTX-M-15 or carbapenemase to resist colistin, thereby guiding infection control measures and halting the spread of these multidrug-resistant pathogens. Unlike Sanger sequencing of the hexon gene, WGS of adenovirus genomes isolated from patients in the neonatal intensive care unit can be used to identify the adenovirus species and genome characteristics to develop better therapeutic strategies (Zhou et al., 2015).

In addition to pathogen identification, WGS can offer information regarding pathogen virulence and novel resistance mechanisms. Identifying virulence factor genes, which are not detected by clinical laboratories or used in patient treatment, allows for the further study of virulence. One study, for instance, highlighted the potential use of WGS in identifying and classifying specific virulence genes in *S. aureus*, such as *S. aureus* toxins and panton valentine leucocidin (PVL) (Leopold et al., 2014). WGS can also be used for the early detection of novel resistance mechanisms that conventional molecular
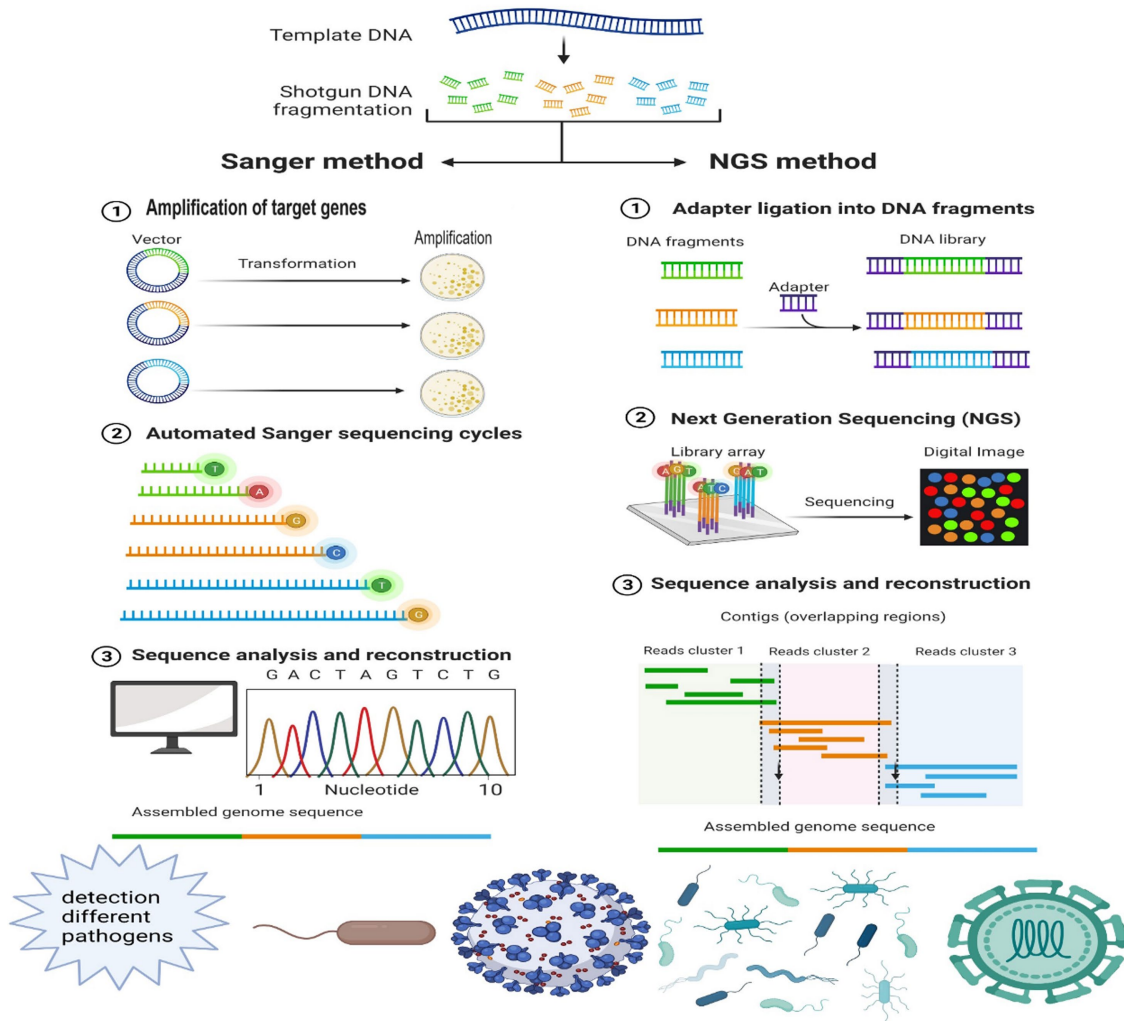
FIGURE 2
Comparing the different processes of the Sanger method and NGS in detecting different pathogens. Sanger method's methodology comprises amplification, automated cycling, sequence analysis, and reconstruction. This technique generates several copies of the target DNA region. The workflow of NGS (take Illumina for example) in clinical setup, including sample separation and preparation, NGS based on the addressed request, Illumina process, related bioinformatics data processing, and retrieval of the final results. The different procedures of two methods result in different application in detecting various pathogens, including viruses, bacteria, and fungi.

detection techniques may overlook, such as PCR of a particular gene or locus. The ability to identify resistant subpopulations is one of the biggest potential benefits of the viral WGS approach, which the Sanger method cannot achieve. According to one study, WGS of HIV improves the ability to find low-frequency drug-resistant mutations in HIV-1 (Tzou et al., 2018). Antimicrobial resistance (AMR) prediction is one of the most intriguing potential uses of WGS, and it may offer preliminary results more rapidly than conventional phenotypic methods. Furthermore, virulence factors and AMR genes can be detected by metagenomics (Sanabria et al., 2021).

With the correlation between antimicrobial resistance genes (ARGs) and phenotypic outcomes, numerous published reports have shown the potential of using WGS for identifying resistance in a range of bacteria. Nevertheless, comprehending the complex relationship between ARGs and phenotypic resistance is a complex task. Finding

an ARG does not always imply that the gene is expressed; for example, regulatory processes and environmental circumstances might affect gene translation (Nielsen and Browne, 2022). Furthermore, using databases to identify ARGs is difficult. These databases might not always be complete or current, which could result in inadequate knowledge of the resistome. Genetic changes can confer resistance without the direct participation of recognized ARGs, which further complicates issues. This emphasizes the necessity of thorough genomic analysis that putative variations resulting from mutational events in addition to known resistance genes (Mao and Zhang, 2023). Extrachromosomal DNA fragments called plasmids are essential for the spread of ARGs. The link between genotypic and phenotypic resistance can become increasingly complicated when bacteria acquire plasmids that confer resistance. A more comprehensive understanding of the variables driving the spread of antibiotic resistance among

bacterial populations can be obtained by examining plasmid-mediated resistance mechanisms (Dewan and Uecker, 2023).

In conclusion, improving our understanding of antimicrobial resistance will require determining the complex interactions between ARGs and phenotypic resistance. A multifaceted strategy that integrates genotypic and phenotypic investigations is necessary because of the constraints related to gene expression, mutations, plasmid-mediated resistance, database accuracy, and mutations. By combining sophisticated molecular approaches with strategies such As The Kirby-Bauer method, antimicrobial resistance dynamics can be explored in greater detail and open the door for focused interventions and efficient resistance management (Hughes and Andersson, 2017).

Next-generation targeted sequencing

Before creating libraries and sequencing, in tNGS, known target sequences are amplified and enriched. The benefit of using tNGS over a metagenomic method is avoiding the needle-in-the-haystack problem of amplifying sparse microbial sequences in samples with a high proportion of host cells (Schlaberg et al., 2017). However, selection procedures, such as multiplex PCR for particular genes, could result in target bias. Although assays may also target genes associated with antibiotic resistance, the primary objective of tNGS in clinical applications is to identify microbial pathogens or pathogens in patient samples. PCR amplifies the 16S ribosomal RNA (rRNA) gene before NGS, which is the most popular enrichment technique for clinical applications and microbiome research (Salipante et al., 2014). Alternative enrichment and sequencing techniques are also being created. The development of a bacterial tNGS assay by amplifying and sequencing the complete 16S-23S rRNA region in urine or blood samples from patients with suspected urinary tract infections was reported by Sabat et al. (2017). Interpreting 16S data frequently only yields genus-level identifications, and the addition of 23S regions improves the specificity and sensitivity. In addition, 16S amplicon sequencing is also considered metagenomic. Although rRNA microarrays can also detect 16S rRNA, distinguishing between closely related strains is difficult with this data compared to NGS data (Zabarovsky et al., 2003; Pozhitkov et al., 2006). In addition, the range of species that can be detected using microarrays is limited, and novel or unknown organisms cannot be identified. tNGS was used to accurately discover the viral resistance of cytomegalovirus (CMV) in clinical samples, especially minor variants (Briese et al., 2015).

Metagenomic next-generation sequencing

Compared to tNGS, which amplifies specific gene fragments, mNGS can be used to detect all genetic material. Therefore, in comparison to tNGS, mNGS has the advantage of detecting potential pathogens and has a wide detection range (Gaston et al., 2022; Gauthier et al., 2023). However, there are the disadvantages of unstable detection, higher cost and long detection time. All microbial groups (including bacterial, viral, and fungal agents), resistance markers, virulence factors, or even host biomarkers associated with various disease states can be sequenced simultaneously with unbiased detection. A straightforward diagnosis from patient samples can

be made according to mNGS data. mNGS can detect cell-free DNA (cfDNA) or cell-free RNA (cfRNA) to identify pathogens. Importantly, whether the pathogens discovered based on mNGS are responsible for the disease needs to be carefully determined. Because these novel pathogens have been less researched in the past, they may be overlooked.

A few current studies indicate that mNGS has improved sensitivity when compared with current diagnostic methods, and these result indicate the potential of mNGS. A retrospective chart review revealed that the total clinical sensitivity and specificity for mNGS were 50.7 and 85.7%, respectively, while those for standard diagnostics were 35.2 and 89.1%, respectively (Mitchell and Simner, 2019). mNGS performs particularly well for fungi, viruses, anaerobes, and *M. tuberculosis*. For example, when the microorganism culture is negative, mNGS of sonicated fluids or synovial fluid from prosthetic joint infection gives an incremental 25 and 18.3% yield, respectively. One HIV-1, two Taenia solium, four fungi, and one other pathogen were found in the cerebral fluid of 94 patients who had subacute or chronic meningitis by using mNGS (Mitchell and Simner, 2019).

Application of NGS for the detection of different pathogens

Infectious diseases are still a leading cause of human morbidity and mortality globally. The rapid and precise diagnosis of aetiologic microorganisms can promote the therapeutic process. The diversity of detectable microorganisms is relatively narrow when using culture methodology (Tomblyn et al., 2009; Miao et al., 2018), including inaccurate and time-consuming identification techniques involving pathogen isolation, selective culture, and pathological inspection. Clinical specimens may not yield conclusive results for cultivating pathogenic bacteria for days to weeks (Arastehfar et al., 2019).

Furthermore, the reported turnaround time for NGS requires substantially less time from receiving clinical samples to data analysis than traditional methods (Breitwieser et al., 2018). There are several reported practical uses of NGS in clinical settings (Ondov et al., 2011; Maabar et al., 2019). The results of NGS provide useful information for diagnoses, controlled treatment, the evaluation of efficiency, and the prognosis of infectious diseases (Murkey et al., 2017; Kufner et al., 2019). We summarize the applications of NGS in identifying bacteria, fungi, and viruses in detail to further promote its clinical use. Table 2 shows the advantages and disadvantages of NGS in the detection of different pathogens.

Bacteria

NGS shows high resolution regarding the bacterial genotypes and is a significant and powerful method in infectious disease epidemiology (Tweedy et al., 2018). The composition of causative microorganisms, such as gram-negative bacteria, gram-positive bacteria, anaerobes, and fungi, differs in septic patients in dissimilar clinical circumstances. In approximately half of sepsis patients, the causative organisms remain unidentified, specifically, culture-negative sepsis. Several investigations have revealed that genomic RNA or DNA fragments are linked with specific microorganisms. As a consequence, NGS of cfRNA or cfDNA in cleansed plasma can be applied to identify

TABLE 2 Advantages and disadvantages of sequencing technology in the detection of different pathogens.

Pathogen	Advantage	Disadvantage	Sequencing approach	References
Bacteria	• Simplified bioinformatics • Amplifies the amount of microorganisms present and targets • Reduced price	• Only Bacteria identification • Particular microbes • Can still miss some species due to primer mismatches • Scarce copy of pathogen sequences may be harder to detect.	tNGS (16S rRNA)	Petersen et al. (2019)
Fungi	• Reduced price • Better sensitivity and massive information • More speed • Simplified bioinformatics • The ability of amplified most of the fungus is available.	• Still miss some species due to primer mismatches • Particular to a subgroup of parasites and fungus • Scarce copy of pathogen sequences may be harder to detect	tNGS (18S rRNA gene/ITS gene sequence)	Xing et al. (2022), Arastehfar et al. (2019)
Virus	• Higher sensitivity and amount of data allows *de novo* assembly and even reads level detect • Potential to detect the full spectrum of viruses, including unknown and unexpected viruses • Allowing for the thorough identification of minority variants, which represents a clear advantage over direct (Sanger) sequencing • Detection of pathogens that do not rely on references or amplification	• Assembly and characterization of complex/highly repetitive genomic regions • Reconstruction of complete "real" viral haplotypes	WGS; mNGS	Barzon et al. (2011), Quer et al. (2022) Houldcroft et al. (2017)

pathogens in sepsis samples, alongside information concerning genetic relatedness (Tassios and Moran-Gilad, 2018).

Globally, WGS may be beneficial to track the spread of bacterial pathogens, which have shown how quickly infectious diseases spread throughout the world (Harris et al., 2010; Jain et al., 2013; Katz et al., 2014). Using this technology, researchers observed a genomic change in a sensitive bacterial strain with resistance to the international disease Cholera (Mutreja et al., 2013). In addition, genomic mutations were observed in *Streptococcus pneumoniae* in a large group of specimens (Croucher et al., 2013). WGS can also help clarify the bacterial transmission between population groups by sampling bacteria from specific host groups. For instance, many studies investigate gonococcal genomic epidemiology (Grad et al., 2014). Transmission can occur between men who have sex with other individuals, which means that gonorrhea can be transmitted from one to another.

Additionally, WGS has shown advantages over alternative genotyping approaches for tracking and analyzing micro-epidemics; for example, WGS was used to compare 86 human *M. tuberculosis* isolates from a German outbreak (Roetzer et al., 2013; Bosch et al., 2016). WGS was employed in 2010 to analyze 63 methicillin-resistant *S. aureus* (MRSA) strains from diverse nations, and the results allowed for the reconstruction of intercontinental transmissions over a forty-year period, as well as the possibility for transmission within a hospital setting (Harris et al., 2010). The cholera outbreak in Haiti was also investigated using WGS, and it was discovered that the Haitian strains were closely related to those from Nepal (Jain et al., 2013; Katz et al., 2014). These groundbreaking experiments prove the utility of WGS for retroactive genotyping. Sequencing methods need to be improved

to make WGS a plausible genotyping tool during epidemics (Fournier et al., 2014).

Fungi

After years of focus on human-associated bacteria, human-associated fungi have progressively drawn attention. Clinical researchers have focused on fungal populations, especially Candida, Malassezia, Penicillium, etc., because of the generalization and rapidity of fungal diseases. The analysis and results of fungi detection can be influenced by the techniques utilized to analyze human fungal communities (Nilsson et al., 2019). The use of NGS in fungal diagnosis should be considered due to the lack of effective detection methods for clinical fungal infections and the seriousness of the fungal illness. NGS technology shows several advantages in detecting fungal pathogens. First, NGS technology is appropriate for microbial diseases caused by hostile cultures and slowly growing microbes, including fungi (Xing et al., 2022). In addition, NGS is a helpful tool for samples with low fungal loads (Sequencing et al., 2021). Second, NGS is considerably more specific than other approaches and offers more accurate identification of fungal species (Arastehfar et al., 2019). Finally, compared with first generation DNA sequencing, NGS exhibits better sensitivity and provides large amounts of information in addition to accuracy and speed.

However, the use of NGS to identify fungi has been relatively less researched during the last ten years. By the end of 2014, most bacterial and viral genomes had been sequenced with NGS, while many fungal genomes remained undiscovered (Thomma et al., 2016). Existing NGS

technologies are only partially believable due to technical issues and objective mistakes (Nilsson et al., 2019). Leho Tedersoo's group team focused on fungus and suggested that the existing NGS technique frequently conceals numerous significant and minor problems. Moreover, mycobiome sequencing requires both the repeatability of fungal sequencing data and the accessibility of public data. His research team also conducted a study assessing the impact of the Respiratory Pathogen ID/AMR (RPIP) kit on a specific NGS workflow. After thorough comparisons, they concluded that NGS workflows could not replace traditional culture and other methodologies, partially due to the complexity of bioinformatic analysis of NGS (Gaston et al., 2022). Additionally, various internal transcribed spacer (ITS) primer types could inadvertently result in the identification of several fungal species. For instance, although ITS1-F, ITS1, ITS5, etc., have a preference for basidiomycete amplification, ITS2, ITS3, ITS4, etc., have a preference for ascomycetes (Bellemain et al., 2010).

Further advancements should be made to improve the fungal genome database and next-generation detection techniques. Overall, the applications of NGS are expanding rapidly, from cutting-edge diagnostic techniques to common clinical detection. NGS may now be utilized for routine microbial detection because the speed of detection has increased. Genomic testing has become more popular, particularly since 2005, thanks to the development and advancement of NGS technology and the lower cost of the testing supplies (Fournier et al., 2014). The repeatability, quantification outcomes, and classification accuracy of NGS should be enhanced to differentiate a more comprehensive range of species. Consequently, there is still a long way to go before NGS can be a standard operating procedure for fungal detection in diagnostic laboratories.

Viruses

The application of NGS for virus identification has become increasingly popular. In addition, NGS offers a cutting-edge tool for massive the large-scale genomic sequencing of viruses such as Hantaan virus (HTNV), hepatitis C virus (HCV), and coronavirus. NGS has opened a new era of viral genomics for the surveillance, tracing, and risk management of viral diseases. Here, we will concentrate on the viruses mentioned above.

Complete genome sequencing and the isolation of infectious particles are essential to define and develop preventive measures for HTNV epidemics. Dong Hyun Song et al. used the lung tissues of striped field mice to isolate 12 HTNVs in highly Hemorrhagic Fever With Renal Syndrome (HFRS)-endemic regions. To obtain the genomic sequence of HTNV isolates, sequence-independent, single-primer amplification (SISPA) NGS was used. Based on the entire length of the prototype HTNV 76–118, the nucleotide sequences of the HTNV S, M, and L segments were covered to 99.4–100%, 97.5–100, and 95.6–99.8%, respectively (Ganzenmueller et al., 2013).

In the new world of direct-acting antiviral (DAA) medicines, NGS technologies for HCV can detect both viral genotypes and genetic resistance. In one study, the ability of NGS techniques to identify full-length, deep HCV sequences and their usefulness for clinical diagnosis were assessed. They examined the following three NGS techniques used in four UK centers: (i) metagenomics, (ii) pre-enrichment of HCV RNA by probe capture, and (iii) HCV preamplification by PCR. A panel of samples with various viral loads and genotypes was

used to compare the sequencing coverage, depth metrics, quasispecies diversity, detection of DAA resistance-associated variations (RAVs), mixed HCV genotypes, and other coinfections (Thomson et al., 2016). Nearly complete genome sequences were produced by each NGS technique from more than 90% of the samples. For samples with low virus loads, enrichment techniques and PCR preamplification provided better sequencing depth and efficiency. All NGS techniques correctly identified the mix of HCV genotypes in infections. Most RAVs were reliably discovered, and consensus sequences produced by various NGS techniques were generally consistent. However, the ability of various techniques to find identify RAV minor is varied. Human pestivirus coinfections have been discovered using metagenomic techniques.

In addition, NGS has also played a tremendously important role in the identification of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) variants and the traceability of SARS-CoV-2 (Chan et al., 2020). For example, Pan et al. used NGS to obtain 2,994 whole-genome sequences of SARS-CoV-2 and performed phylogenetic and population dynamics analyses to rapidly identify the variants and lineages (Pan et al., 2023). The broad use of NGS in diagnosis would significantly improve the capacity of researchers and governments to keep track of new strains of infectious diseases such as SARS-CoV-2. Massive efforts are made to coordinate extensive sequencing of the SARS-CoV-2 virus, as well as identifying SARS-CoV-2 and coinfections using amplicon and metagenomic-MinilON-based sequencing, respectively. By using primers targeting highly conserved regions of a genome, amplicon-based NGS technology is frequently used to offer detailed information of targeted region. mNGS, in contrast, employs a shotgun strategy to identify all genetic material in a sample, as opposed to the highly conserved genetic region.

Researchers are looking to improve NGS technology in SARS-CoV-2 detection. The amplicon-based sequencing method provided a significantly higher read depth than the metagenomic approach. For instance, SARS-CoV-2 was discovered in nasopharyngeal swab material from a patient in Feira de Santana-Bahia, Brazil, using meta-transcriptomic NGS (Campos et al., 2020). To eliminate rRNA from one sample, they used Thermo Fisher's Low Input RiboMinus TM Eukaryote System v2 for the ion-semiconductor sequencing utilized in this procedure. Human transcripts comprised 77.29% of all reads in the rRNA-depleted sample, while 84.49% of all reads in the whole RNA library were human transcripts. Despite the inefficiency of host genome removal, the genome coverage of contigs in the rRNA-depleted library increased 30% compared to nondepleted samples. How to effectively remove host contamination is also one of the problems that needs to be solved (Song and Xie, 2020). These results imply that rRNA depletion methods may improve the diagnostic capabilities of NGS (Liu et al., 2020).

Conclusion and prospects

Outcomes following NGS application for diagnosing infectious illnesses are heartening. The use of NGS in pathogen identification has several benefits over conventional approaches, including the capacity to study numerous pathogens at once, high sensitivity and specificity, and the capacity to identify new or emerging pathogens. NGS can be used to accurately identify and detect pathogens such as viruses, bacteria, fungi, and parasites in a variety of sample types, including

clinical specimens, environmental samples, and vectors. The rapid and accurate identification of the pathogen responsible for an outbreak, the tracking of transmission patterns, and the monitoring of genomic alterations as an outbreak develops have all been made possible by NGS-based pathogen identification. Through the unbiased and thorough sequencing of microbial genomes made possible by NGS, new or genetically varied strains of bacteria can be found, in addition to well-known diseases. Additionally, NGS has proven crucial for the detection and tracking of resistance, the identification of zoonotic infections, and the understanding of the genetic diversity and evolution of pathogens in the surveillance and monitoring of infectious diseases.

However, there are also issues with using NGS for pathogen detection, such as the requirement for strong bioinformatics pipelines, the standardization of methods, and workflow optimization for various sample types. Data interpretation and analysis, which call on knowledge in bioinformatics and genetics, continue to be significant challenges. The application of NGS in recognizing clinical pathogens remains in the primary phase, and no mNGS procedures are approved by the Food and Drug Management Administration (Kim et al., 2016). Consequently, the current NGS procedures are personalized, and there is a long way to go to accomplish standardized detection (Kong et al., 2022). Moreover, the restricted capacity of NGS to detect low-frequency variations is one of its major drawbacks (Cohen et al., 2021). This restriction becomes important when there are several virus variants coexisting in quasispecies or when there are genetic differences in an organism known as heteroplasmy. Additionally, there are challenges with the clinical identification of pathogens. Currently, high diagnostic costs and a lack of genomics competence are the main barriers that prevent the adoption of NGS in clinics (Stoddard et al., 2014).

In conclusion, NGS has enormous potential in pathogen identification as it allows for more precise, quicker, and more thorough detection of infections. The application of NGS for identifying various pathogens will be further enhanced by ongoing developments in sequencing technologies, data analysis tools, and collaboration between researchers, clinicians, and public health agencies. These developments will also contribute to improve the diagnostics, surveillance, and control of infectious diseases.

Author contributions

AN: Writing – original draft. YW: Writing – original draft. DW: Writing – original draft, Writing – review & editing. AS: Writing – original draft, Writing – review & editing. MA: Writing – original draft, Writing – review & editing. SX: Funding acquisition, Project administration, Writing – original draft, Writing – review & editing. YT: Funding acquisition, Project administration, Supervision, Writing – original draft, Writing – review & editing. All authors read and approved the final manuscript.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Adessi, C., Matton, G., Ayala, G., Turcatti, G., Mermod, J. J., Mayer, P., et al. (2000). Solid phase DNA amplification: characterisation of primer attachment and amplification mechanisms. *Nucleic Acids Res.* 28:87. doi: 10.1093/nar/28.20.e87

Arastehfar, A., Boekhout, T., Butler, G., Buda De Cesare, G., Dolk, E., Toni Gabaldón, A., et al. (2019). Recent trends in molecular diagnostics of yeast infections: from PCR to NGS. *FEMS Microbiol. Rev.* 43, 517–547. doi: 10.1093/femsre/fuz015

Barzon, L., Lavezzo, E., Militello, V., Toppo, S., and Palù, G. (2011). Applications of next-generation Sequencing technologies to diagnostic virology. *Int. J. Mol. Sci.* 12, 7861–7884. doi: 10.3390/ijms12117861

Batovska, J., Cogan, N. O. I., Lynch, S. E., and Blacket, M. J. (2017). Using next-generation sequencing for DNA barcoding: capturing allelic variation in ITS2. G3: genes, genomes. *Genetics* 7, 19–29. doi: 10.1534/g3.116.036145

Bauermeister, A., Mannochio-Russo, H., Costa-Lotufo, L. V., Jarmusch, A. K., and Dorrestein, P. C. (2022). Mass spectrometry-based metabolomics in microbiome investigations. *Nat. Rev. Microbiol.* 20, 143–160. doi: 10.1038/s41579-021-00621-9

Bellemain, E., Carlsen, T., Brochmann, C., Coissac, E., Taberlet, P., and Kauserud, H. (2010). ITS as an environmental DNA barcode for Fungi: an in silico approach reveals potential PCR biases. *BMC Microbiol.* 10, 1–9. doi: 10.1186/1471-2180-10-189

Benner, S., Chen, R. J. A., Wilson, N. A., Abu-shumays, R., Hurt, N., Lieberman, K. R., et al. (2007). Sequence-specific detection of individual DNA polymerase complexes in real time using a nanopore. *LETTERS* 2, 718–724. doi: 10.1038/nnano.2007.344

Bosch, T., Witteveen, S., Haenen, A., Landman, F., and Schouls, L. M. (2016). Next-generation Sequencing confirms presumed nosocomial transmission of livestock-associated methicillin-resistant *Staphylococcus Aureus* in the Netherlands. *Appl. Environ. Microbiol.* 82, 4081–4089. doi: 10.1128/AEM.00773-16

Breitwieser, F. P., Baker, D. N., and Salzberg, S. L. (2018). Kraken Uniq: confident and fast metagenomics classification using unique k-Mer counts. *Genome Biol.* 19, 1–10. doi: 10.1186/s13059-018-1568-0

Briese, T., Kapoor, A., Mishra, N., Jain, K., Kumar, A., Jabado, O. J., et al. (2015). Virome capture Sequencing enables sensitive viral diagnosis and comprehensive Virome analysis. *MBio* 6, 1–11. doi: 10.1128/mBio.01491-15

Brown, A. C., Bryant, J. M., Einer-Jensen, K., Holdstock, J., Houniet, D. T., Chan, J. Z. M., et al. (2015). Rapid whole-genome Sequencing of *Mycobacterium Tuberculosis* isolates directly from clinical samples. *J. Clin. Microbiol.* 53, 2230–2237. doi: 10.1128/JCM.00486-15

Campos, G. S., Sardi, S. I., Falcao, M. B., Belitardo, E. M. M. A., Rocha, D. J. P. G., Rolo, C. A., et al. (2020). Since January 2020 Elsevier has created a COVID-19 resource Centre with free information in English and mandarin on the novel coronavirus COVID-19. The COVID-19 resource Centre is hosted on Elsevier connect, the company's public news and information. *J. Virol. Methods J.* 6, 1–12. doi: 10.1093/nar/gkn151

Chan, J. F., Yuan, S., Kok, K. H., To, K. K., Chu, H., Yang, J., et al. (2020). A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *Lancet* 395, 514–523. doi: 10.1016/S0140-6736(20)30154-9

Church, D. L., Cerutti, L., Gürtler, A., Griener, T., Zelazny, A., and Emler, S. (2020). Performance and application of 16S rRNA gene cycle Sequencing for routine identification of Bacteria in the clinical microbiology laboratory. *Clin. Microbiol. Rev.* 33, 1–24. doi: 10.1128/CMR.00053-19

Cohen, J. D., Douville, C., Dudley, J. C., Mog, B. J., Popoli, M., Ptak, J., et al. (2021). Detection of low-frequency DNA variants by targeted sequencing of the Watson and Crick strands. *Nat. Biotechnol.* 39, 1220–1227. doi: 10.1038/s41587-021-00900-z

Crossley, B. M., Bai, J., Glaser, A., Maes, R., Porter, E., Killian, M. L., et al. (2020). Guidelines for Sanger Sequencing and molecular assay monitoring. *J. Vet. Diagn. Investig.* 32, 767–775. doi: 10.1177/1040638720905833

Croucher, N. J., Finkelstein, J. A., Pelton, S. I., Mitchell, P. K., Lee, G. M., Parkhill, J., et al. (2013). Epidemiology. *Nat. Genet.* 45, 656–663. doi: 10.1038/ng.2625

Dai, D., Zhu, J., Sun, C., Li, M., Liu, J., Wu, S., et al. (2022). GMrepo v2: a curated human gut microbiome database with special focus on disease markers and cross-dataset comparison. *Nucleic Acids Res.* 50, D777–D784. doi: 10.1093/nar/gkab1019

Deamer, D., Akeson, M., and Branton, D. (2016). Three decades of nanopore sequencing. *Nat. Biotechnol.* 34, 518–524. doi: 10.1038/nbt.3423

Dewan, I., and Uecker, H. (2023). A mathematician's guide to plasmids: an introduction to plasmid biology for modellers. *Microbiol. Soc.* 1–23. doi: 10.1099/mic.0.001362

Faulk, C. (2023). De novo sequencing, diploid assembly, and annotation of the black carpenter ant, Camponotus pennsylvanicus, and its symbionts by one person for $1000, using nanopore sequencing. *Nucleic Acids Res.* 51, 17–28. doi: 10.1093/nar/gkac510

Fournier, P. E., Dubourg, G., and Raoult, D. (2014). Clinical detection and characterization of bacterial pathogens in the genomics era. *Genome Med.* 6, 1–15. doi: 10.1186/s13073-014-0114-2

Ganzenmueller, T., Hage, E., Yakushko, Y., Kluba, J., Woltemate, S., Schacht, V., et al. (2013). No human virus sequences detected by next-generation Sequencing in benign verrucous skin Tumors occurring in BRAF-inhibitor-treated patients. *Exp. Dermatol.* 22, 725–729. doi: 10.1111/exd.12249

Garrido-Cardenas, J. A., Garcia-Maroto, F., Alvarez-Bermejo, J. A., and Manzano-Agugliaro, F. (2017). DNA Sequencing sensors: an overview. *Sensors (Switzerland)* 17, 1–15. doi: 10.3390/s17030588

Gaston, D. C., Miller, H. B., Fissel, J. A., Jacobs, E., Gough, E., Jiajun, W., et al. (2022). Evaluation of metagenomic and targeted next-generation Sequencing workflows for detection of respiratory pathogens from bronchoalveolar lavage fluid specimens. *J. Clin. Microbiol.* 60, 1–15. doi: 10.1128/jcm.00526-22

Gauthier, N., Chorlton, S. D., Krajden, M., and Manges, A. R. (2023). Agnostic Sequencing for detection of viral pathogens. *Clin. Microbiol. Rev.* 36:e0011922. doi: 10.1128/cmr.00119-22

Gilchrist, C. A., Turner, S. D., Riley, M. F., Petri, W. A., and Hewlett, E. L. (2015). Whole-genome sequencing in outbreak analysis. *Clin. Microbiol. Rev.* 28, 541–563. doi: 10.1128/CMR.00075-13

Grad, Y. H., Kirkcaldy, R. D., Trees, D., Dordel, J., Harris, S. R., Goldstein, E., et al. (2014). Genomic epidemiology of *Neisseria Gonorrhoeae* with reduced susceptibility to Cefi Xime in the USA: a retrospective observational study. *Lancet Infect. Dis.* 3099, 1–7. doi: 10.1016/S1473-3099(13)70693-5

Green, N. S., and Pass, K. A. (2005). Neonatal screening by DNA microarray: spots and chips. *Nat. Rev. Genet.* 6, 147–151. doi: 10.1038/nrg1526

Greenough, L., Kelman, Z., and Gardner, A. F. (2015). The roles of family B and D DNA polymerases in Thermococcus species 9 ° N Okazaki fragment maturation *. *J. Biol. Chem.* 290, 12514–12522. doi: 10.1074/jbc.M115.638130

Greninger, A. L., Messacar, K., Dunnebacke, T., Naccache, S. N., Federman, S., Bouquet, J., et al. (2015). Clinical metagenomic identification of Balamuthia mandrillaris encephalitis and assembly of the draft genome: the continuing case for reference genome sequencing. *Genome Med.* 7:113. doi: 10.1186/s13073-015-0235-2

Guo, J., Xu, N., Li, Z., Zhang, S., Wu, J., Dae, H. K., et al. (2008). Four-color DNA sequencing with 3′-O-modified nucleotide reversible terminators and chemically cleavable fluorescent dideoxynucleotides. *Proc. Natl. Acad. Sci. U. S. A.* 105, 9145–9150. doi: 10.1073/pnas.0804023105

Gupta, N., and Verma, V. K. (2019). *Next-generation Sequencing and its application: empowering in public health beyond reality*. Singapore: Springer Nature, 313–341.

Han, D., Li, Z., Li, R., Tan, P., Zhang, R., and Li, J. (2019). Critical reviews in microbiology MNGS in clinical microbiology laboratories: on the Road to maturity. *Crit. Rev. Microbiol.* 1–18. doi: 10.1080/1040841X.2019.1681933

Harris, S. R., Feil, E. J., Holden, M. T. G., Quail, M. A., Nickerson, E. K., Chantratita, N., et al. (2010). Evolution of MRSA during hospital transmission and intercontinental spread. *Science* 327, 469–474. doi: 10.1126/science.1182395

Houldcroft, C. J., Beale, M. A., and Breuer, J. (2017). Clinical and biological insights from viral genome Sequencing. *Nat. Rev. Microbiol.* 15, 183–192. doi: 10.1038/nrmicro.2016.182

Hughes, D., and Andersson, D. I. (2017). Environmental and genetic modulation of the phenotypic expression of antibiotic resistance. *Invest. Sci.* 41, 374–391. doi: 10.1093/femsre/fux004

Hung, J.-h., and Weng, Z. (2017). Analysis of microarray and RNA-Seq expression profiling data. *Cold Spring Harbor* 1–7. doi: 10.1101/pdb.top093104

Jain, M., Koren, S., Miga, K. H., Quick, J., Rand, A. C., Sasani, T. A., et al. (2018). Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat. Biotechnol.* 36, 338–345. doi: 10.1038/nbt.4060

Jain, M., Kushwah, K. S., Kumar, P., and Goel, A. K. (2013). Molecular characterization of *Vibrio Cholerae* O1 reveals continuous evolution of its new variants in Indiaariants in India. *Indian J. Microbiol.* 53, 137–141. doi: 10.1007/s12088-013-0372-5

Jiang, S., Chen, Y., Han, S., Lv, L., and Li, L. (2022). Next-generation Sequencing applications for the study of fungal pathogens. *Microorganisms* 10, 1–13. doi: 10.3390/microorganisms10101882

Jin, J., Yamamoto, R., Takeuchi, T., Cui, G., Miyauchi, E., Hojo, N., et al. (2022). High-throughput identification and quantification of single bacterial cells in the microbiota. *Nat. Commun.* 13:863. doi: 10.1038/s41467-022-28426-1

Katz, L. S., Turnsek, M., Kahler, A., Hill, V. R., Fidelma Boyd, E., and Tarr, C. L. (2014). Draft genome sequence of environmental vibrio cholerae 2012EL-1759 with similarities to the *V. cholerae* O1 classical biotype. *Genome Announc.* 2, 4–5. doi: 10.1128/genomeA.00617-14

Kiehn, O., and Car. (2017). Enhancing the accuracy of next-generation sequencing for detecting rare and subclonal mutations. *Physiol. Behav.* 176, 139–148. doi: 10.1038/nrg.2017.117.Enhancing

Kim, D., Song, L., Breitwieser, F. P., and Salzberg, S. L. (2016). Centrifuge: rapid and sensitive classification of metagenomic sequences. *Genome Res.* 26, 1721–1729. doi: 10.1101/gr.210641.116

Koch, C., Reilly-O'Donnell, B., Gutierrez, R., Lucarelli, C., Ng, F. S., Gorelik, J., et al. (2023). Nanopore sequencing of DNA-barcoded probes for highly multiplexed detection of micro RNA, proteins and small biomarkers. *Nat. Nanotechnol.* 18, 1483–1491. doi: 10.1038/s41565-023-01479-z

Kong, M., Li, W., Kong, Q., Dong, H., Han, A., and Jiang, L. (2022). Application of metagenomic next-generation Sequencing in Cutaneo. *Front. Cell. Infect. Microbiol.* 12, 1–6. doi: 10.3389/fcimb.2022.942073

Köser, C. U., Holden, M. T., Ellington, M. J., Cartwright, E. J., Brown, N. M., Ogilvy-Stuart, A. L., et al. (2012). Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak. *N. Engl. J. Med.* 366, 2267–2275. doi: 10.1056/NEJMoa1109910

Kufner, V., Plate, A., Schmutz, S., Braun, D. L., Günthard, H. F., Capaul, R., et al. (2019). Two years of viral metagenomics in a tertiary diagnostics unit: evaluation of the first 105 cases. *Genes* 10, 1–15. doi: 10.3390/genes10090661

Leopold, S. R., Goering, R. V., Witten, A., Harmsen, D., and Mellmann, A. (2014). Bacterial whole-genome Sequencing revisited: portable, scalable, and standardized analysis for typing and detection of virulence and antibiotic resistance genes. *J. Clin. Microbiol.* 52, 2365–2370. doi: 10.1128/JCM.00262-14

Li, X. Q., Guo, B. L., Cai, W. Y., Zhang, J. M., Huang, H. Q., Zhan, P., et al. (2016). The role of melanin pathways in extremotolerance and virulence of Fonsecaea revealed by de novo assembly transcriptomics using illumina paired-end sequencing. *Stud. Mycol.* 83, 1–18. doi: 10.1016/j.simyco.2016.02.001

Liu, J., Zheng, X., Tong, Q., Li, W., Wang, B., Sutter, K., et al. (2020). Overlapping and discrete aspects of the pathology and pathogenesis of the emerging human pathogenic coronaviruses SARS-CoV, MERS-CoV, and 2019-NCoV. *J. Med. Virol.* 92, 491–494. doi: 10.1002/jmv.25709

Maabar, M., Davison, A. J., Vučak, M., Thorburn, F., Murcia, P. R., Gunson, R., et al. (2019). DisCVR: rapid viral diagnosis from high-throughput Sequencing data. *Virus Evol.* 5, 1–8. doi: 10.1093/ve/vez033

Manuscript, A., and Malignancies, H. (2014). Nihms582105. *Nat. Instit. Health* 6, 69–122. doi: 10.1126/scitranslmed.3007094.Detection

Mao, Y., and Zhang, N. (2023). iScience NCRD: a non-redundant comprehensive database for detecting antibiotic resistance genes. *ISCIENCE* 26:108141. doi: 10.1016/j.isci.2023.108141

Miao, Q., Ma, Y., Wang, Q., Pan, J., Zhang, Y., Jin, W., et al. (2018). Microbiological diagnostic performance of metagenomic next-generation Sequencing when applied to clinical practice. *Clin. Infect. Dis.* 67, S231–S240. doi: 10.1093/cid/ciy693

Mitchell, S. L., and Simner, P. J. (2019). Next-generation Sequencing in clinical microbiology: are we there yet? *Clin. Lab. Med.* 39, 405–418. doi: 10.1016/j.cll.2019.05.003

Moore, S. C., Penrice-Randal, R., Alruwaili, M., Randle, N., Armstrong, S., Hartley, C., et al. (2020). Amplicon-based detection and Sequencing of SARS-CoV-2 in nasopharyngeal swabs from patients with COVID-19 and identification of deletions in the viral genome that encode proteins involved in interferon antagonism. *Viruses* 12, 1–16. doi: 10.3390/v12101164

Mukherjee, C., Beall, C. J., Griffen, A. L., and Leys, E. J. (2018). High-resolution ISR amplicon sequencing reveals personalized oral microbiome. *Microbiome* 6:153. doi: 10.1186/s40168-018-0535-z

Murkey, J. A., Chew, K. W., Carlson, M., Shannon, C. L., Sirohi, D., Sample, H. A., et al. (2017). Hepatitis E virus-associated meningoencephalitis in a lung transplant recipient diagnosed by clinical metagenomic Sequencing. *Open Forum Infect. Dis.* 4, 1–4. doi: 10.1093/OFID/OFX121

Mutreja, A., Kim, D. W., Thomson, N., Connor, T. R., Hee, J., Lebens, M., et al. (2013). Evidence for multiple waves of global transmission within the seventh cholera pandemic. *Nature* 477, 462–465. doi: 10.1038/nature10392.Evidence

Nielsen, T. K., and Browne, P. D. (2022). Antibiotic resistance genes are differentially mobilized according to resistance mechanism. *Giga Science* 1–17. doi: 10.1093/gigascience/giac072

Nilsson, R. H., Anslan, S., Bahram, M., Wurzbacher, C., Baldrian, P., and Tedersoo, L. (2019). Mycobiome diversity: high-throughput Sequencing and identification of Fungi. *Nat. Rev. Microbiol.* 17, 95–109. doi: 10.1038/s41579-018-0116-y

Nowrousian, M. (2010). Next-generation sequencing techniques for eukaryotic microorganisms: Sequencing-based solutions to biological problems. *Eukaryot. Cell* 9, 1300–1310. doi: 10.1128/EC.00123-10

Ondov, B. D., Bergman, N. H., and Phillippy, A. M. (2011). Interactive metagenomic visualization in a web browser. *BMC Bioinform.* 12:385. doi: 10.1186/1471-2105-12-385

Pan, Y., Wang, L., Feng, Z., Xu, H., Li, F., Shen, Y., et al. (2023). Characterisation of SARS-CoV-2 variants in Beijing during 2022: an epidemiological and phylogenetic analysis. *Lancet (London, England)* 401, 664–672. doi: 10.1016/S0140-6736(23)00129-0

Petersen, L. M., Martin, I. W., Moschetti, W. E., Kershaw, C. M., and Tsongalis, G. J. (2019). Third-generation Sequencing in the clinical laboratory: Sequencing third-generation Sequencing in the clinical laboratory: exploring the advantages and challenges of nanopore Sequencing. *J. Clin. Microbiol.* 58, 1–10. doi: 10.1128/JCM.01315-19

Petri, B., Chaganti, S. R., Chan, P. S., and Heath, D. (2019). Phytoplankton growth characterization in short term MPN culture assays using 18S metabarcoding and qRT-PCR. *Water Res.* 164:114941. doi: 10.1016/j.watres.2019.114941

Phillips, K. A., Douglas, M. P., and Marshall, D. A. (2020). Expanding use of clinical genome Sequencing and the need for more data on implementation. *JAMA* 324, 2029–2030. doi: 10.1001/jama.2020.19933

Pmg, S., Pacbio, O. U., and Nextseq, I. (2019). Crossm first draft genome sequence of a pearl millet blast pathogen. *Genome Seq.* 1, 11–12.

Podnar, J., Deiderick, H., Huerta, G., and Hunicke-Smith, S. (2014). Next-generation Sequencing RNA-Seq library construction. *Curr. Protoc. Mol. Biol.* 106:4.21.1-4. 21.19. doi: 10.1002/0471142727.mb0421s106

Pozhitkov, A., Noble, P. A., Domazet-Loso, T., Nolte, A. W., Sonnenberg, R., Staehler, P., et al. (2006). Tests of rRNA hybridization to microarrays suggest that hybridization characteristics of oligonucleotide probes for species discrimination cannot be predicted. *Nucleic Acids Res.* 34:e66. doi: 10.1093/nar/gkl133

Quer, J., Colomer-Castell, S., Campos, C., Andrés, C., Piñana, M., Cortese, M. F., et al. (2022). Next-generation Sequencing for confronting virus pandemics. *Viruses* 14, 1–23. doi: 10.3390/v14030600

Radovich, M., and Ragoussis, J. (2014). Methods of quantifying micro RNAs for hypoxia research: classic and next generation. *Antioxid. Redox Signal.* 21, 1239–1248. doi: 10.1089/ars.2013.5716

Rehm, H. L. (2013). A cornerstone in the clinic disease-targeted Sequencing: a cornerstone in the clinic. *Perspectives* 12, 1–6. doi: 10.1038/nrg3463

Reuterl, J. A., Spacek, D., and Snyder, M. P. (2015). High-throughput Sequencing technologies. *Mol. Cell* 58, 586–597. doi: 10.1016/j.molcel.2015.05.004

Rhoads, A., and Kin Fai, A. (2015). PacBio Sequencing and its applications. *Genom. Proteom. Bioinform.* 13, 278–289. doi: 10.1016/j.gpb.2015.08.002

Roetzer, A., Diel, R., Kohl, T. A., Rückert, C., Nübel, U., Blom, J., et al. (2013). Whole genome Sequencing versus traditional genotyping for investigation of a *Mycobacterium Tuberculosis* outbreak: a longitudinal molecular epidemiological study. *PLoS Med.* 10, 1–12. doi: 10.1371/journal.pmed.1001387

Sabat, A. J., Van Zanten, E., Akkerboom, V., Wisselink, G., Van Slochteren, K., De Boer, R. F., et al. (2017). Targeted next-generation Sequencing of the 16S-23S RRNA region for culture-independent bacterial identification-increased discrimination of closely related species. *Sci. Rep.* 7, 1–12. doi: 10.1038/s41598-017-03458-6

Salipante, S. J., Sengupta, D. J., Cummings, L. A., Robinson, A., Kurosawa, K., Hoogestraat, D. R., et al. (2014). Whole genome Sequencing indicates *Corynebacterium Jeikeium* comprises 4 separate genomospecies and identifies a dominant genomospecies among clinical isolates. *Int. J. Med. Microbiol.* 304, 1001–1010. doi: 10.1016/j.ijmm.2014.07.003

Sanabria, A. M., Janice, J., Hjerde, E., Simonsen, G. S., and Hanssen, A. M. (2021). Shotgun-metagenomics based prediction of antibiotic resistance and virulence determinants in *Staphylococcus aureus* from periprosthetic tissue on blood culture bottles. *Sci. Rep.* 11:20848.

Sander, F., Goulson, A. R., and Road, H. (1976). A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J. Mol. Bid* 74, 441–48. doi: 10.1016/0022-2836(75)90213-2

Sanger, F., and Coulson Nicklen, S. (1977). DNA Sequencing with chain-terminating. *Biochemistry* 74, 5463–5467. doi: 10.1073/pnas.74.12.5463

Schlaberg, R., Chiu, C. Y., Miller, S., Procop, G. W., and Weinstock, G. (2017). Validation of metagenomic next-generation Sequencing tests for universal pathogen detection. *Arch. Pathol. Lab. Med.* 141, 776–786. doi: 10.5858/arpa.2016-0539-RA

Schmieder, R., and Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27, 863–864. doi: 10.1093/bioinformatics/btr026

Sequencing, N.-g., Tsang, C.-c., Teng, J. L. L., Lau, S. K. P., and Woo, P. C. Y. (2021). Rapid genomic diagnosis of fungal infections in the age of. *J. Fungi*

Slatko, B. E., Gardner, A. F., and Ausubel, F. M. (2018). Overview of next-generation Sequencing technologies. *Curr. Protoc. Mol. Biol.* 122, 1–11. doi: 10.1002/cpmb.59

Smith, L. M., Sanders, J. Z., Kaiser, R. J., Hughes, P., Dodd, C., Connell, C. R., et al. (1986). Fluorescence detection in automated DNA sequence analysis. *Nature* 321, 674–679. doi: 10.1038/321674a0

Sohn, J. I., and Nam, J. W. (2018). The present and future of de novo whole-genome assembly. *Brief. Bioinform.* 19, 23–40. doi: 10.1093/bib/bbw096

Song, L., and Xie, K. (2020). Engineering CRISPR/Cas9 to mitigate abundant host contamination for 16S rRNA gene-based amplicon sequencing. *Microbiome* 8:80. doi: 10.1186/s40168-020-00859-0

Stoddard, J. L., Niemela, J. E., Fleisher, T. A., and Rosenzweig, S. D. (2014). Targeted NGS: a cost-effective approach to molecular diagnosis of PIDs. *Front. Immunol.* 5, 1–7. doi: 10.3389/fimmu.2014.00531

Tassios, P. T., and Moran-Gilad, J. (2018). Bacterial next generation Sequencing (NGS) made easy. *Clin. Microbiol. Infect.* 24, 332–334. doi: 10.1016/j.cmi.2018.03.001

Thomma, B. P. H. J., Seidl, M. F., Shi-Kunne, X., Cook, D. E., Bolton, M. D., van Kan, J. A. L., et al. (2016). Mind the gap; seven reasons to close fragmented genome assemblies. *Fungal Genet. Biol.* 90, 24–30. doi: 10.1016/j.fgb.2015.08.010

Thomson, E., Camilla, L. C., Badhan, A., Christiansen, M. T., Walt Adamson, M., Ansari, A., et al. (2016). Comparison of next-generation Sequencing Technologies for Comprehensive Assessment of full-length hepatitis C viral genomes. *J. Clin. Microbiol.* 54, 2470–2484. doi: 10.1128/JCM.00330-16

Tian, F., He, J., Shang, S., Chen, Z., Tang, Y., Lu, M., et al. (2023). Survey of mosquito species and mosquito-borne viruses in residential areas along the Sino-Vietnam border in Yunnan Province in China. *Front. Microbiol.* 14:1105786. doi: 10.3389/fmicb.2023.1105786

Tomblyn, M., Chiller, T., Einsele, H., Gress, R., Sepkowitz, K., Storek, J., et al. (2009). Guidelines for preventing infectious complications among hematopoietic cell transplantation recipients: a global perspective. *Biol. Blood Marrow Transplant.* 15, 1143–1238. doi: 10.1016/j.bbmt.2009.06.019

Tourlousse, D. M., Yoshiike, S., Ohashi, A., Matsukura, S., Noda, N., and Sekiguchi, Y. (2017). Synthetic spike-in standards for high-throughput 16S rRNA gene amplicon sequencing. *Nucleic Acids Res.* 45:e23. doi: 10.1093/nar/gkw984

Tweedy, J. G., Escriva, E., Topf, M., and Gompels, U. A. (2018). Analyses of tissue culture adaptation of human herpesvirus-6a by whole genome deep Sequencing redefines the reference sequence and identifies virus entry complex changes. *Viruses* 10, 1–18. doi: 10.3390/v10010016

Tzou, P. L., Ariyaratne, P., Varghese, V., Lee, C., Rakhmanaliev, E., Villy, C., et al. (2018). Comparison of an in vitro diagnostic next-generation Sequencing assay with Sanger Sequencing for HIV-1 genotypic resistance testing. *J. Clin. Microbiol.* 56, 1–13. doi: 10.1128/JCM.00105-18

Wang, X., Mclachlan, J., Zamore, P. D., and Tanaka Hall, T. M. (2002). Modular recognition of RNA by a human Pumilio-homology domain. *Cells* 110, 501–512. doi: 10.1016/S0092-8674(02)00873-5

Wang, D., Urisman, A., Liu, Y. T., Springer, M., Ksiazek, T. G., Erdman, D. D., et al. (2003). Viral discovery and sequence recovery using DNA microarrays. *PLoS Biol.* 1, 257–260. doi: 10.1371/journal.pbio.0000002

Wang, Y., Zhao, Y., Bollas, A., Wang, Y., and Au, K. F. (2021). Nanopore sequencing technology, bioinformatics and applications. *Nat. Biotechnol.* 39, 1348–1365. doi: 10.1038/s41587-021-01108-x

Xing, F., Lo, S. K. F., Ma, Y., Ip, J. D., Chan, W. M., Zhou, M., et al. (2022). Rapid diagnosis of *Mycobacterium Marinum* infection by next-generation Sequencing: a case report. *Front. Med.* 9, 1–5. doi: 10.3389/fmed.2022.824122

Yamada, S., and Nomura, S. (2020). Review of single-cell RNA Sequencing in the heart. *Int. J. Mol. Sci.* 21, 1–15. doi: 10.3390/ijms21218345

Zabarovsky, E. R., Petrenko, L., Protopopov, A., Vorontsova, O., Kutsenko, A. S., Zhao, Y., et al. (2003). Restriction site tagged (RST) microarrays: a novel technique to study the species composition of complex microbial systems. *Nucleic Acids Res.* 31:e95. doi: 10.1093/nar/gng096

Zahedi, A., Greay, T. L., Paparini, A., Linge, K. L., Joll, C. A., and Ryan, U. M. (2019). Identification of eukaryotic microorganisms with 18S rRNA next-generation sequencing in wastewater treatment plants, with a more targeted NGS approach required for Cryptosporidium detection. *Water Res.* 158, 301–312. doi: 10.1016/j.watres.2019.04.041

Zhou, K., Lokate, M., Deurenberg, R. H., Arends, J., Ten Foe, J. L., Grundmann, H., et al. (2015). Characterization of a CTX-M-15 producing *Klebsiella Pneumoniae* outbreak strain assigned to a novel sequence type (1427). *Front. Microbiol.* 6, 1–10. doi: 10.3389/fmicb.2015.01250

Zhou, K., Lokate, M., Deurenberg, R. H., Tepper, M., Arends, J. P., Raangs, E. G., et al. (2016). Use of whole-genome sequencing to trace, control and characterize the regional expansion of extended-spectrum β-lactamase producing ST15 *Klebsiella pneumoniae*. *Sci. Rep.* 6:20840. doi: 10.1038/srep20840