



# Genomic Applications in the Clinical Management of Infectious Diseases

# 37

Martina I. Lefterova, Carlos J. Suarez, Niaz Banaei, and Benjamin A. Pinsky

## Introduction

Next-generation sequencing (NGS) methods, also referred to as deep, ultra-deep, high-throughput, or massively parallel sequencing, comprise a number of sequencing technologies that have succeeded the traditional dideoxynucleoside chain-termination (i.e., Sanger) method. Various platforms, which differ in their sequencing chemistries, read lengths, and throughput capabilities, are available (reviewed in [1]) (Table 37.1). As these platforms have become more accessible, they have become particularly attractive to clinical microbiology laboratories that already rely on molecular methods for pathogen identification and characterization.

NGS studies of microorganisms typically follow one of two general strategies: targeted sequencing or nontargeted sequencing (Fig. 37.1) [2, 3]. The first approach typically uses target-specific primers for PCR-mediated amplification, so that the genomic regions of interest are enriched and selectively sequenced. This approach is often performed to interrogate well-characterized genomic regions (e.g., identify known drug-resistant mutants). Sequencing for de novo assembly of whole genomes, on the other hand, frequently relies on nontargeted library preparation. Whole-genome sequencing or WGS is often performed on cultured isolates, when microorganisms are unknown or the goal is to define

the genomic content and functional potential of the organism under investigation. Nontargeted sequencing may also be applied to primary specimens for culture-independent pathogen identification or characterization of the microbial population. These nontargeted sequencing applications using primary specimens are termed metagenomic sequencing.

Examples of these approaches in infectious disease testing will be discussed with particular attention paid to the technical and bioinformatics challenges that arise with specific scenarios in virology and bacteriology. The use of NGS in clinical microbiology laboratories remains relatively limited, though its role in the diagnosis and management of infectious diseases continues to grow as standardized operational protocols, automation, and data analysis pipelines emerge.

## Specific Applications in Diagnostic Virology

### Viral Drug Resistance Mutation Testing

The emergence of drug resistance is an important factor in the management of several clinically significant viral infections. Genotypic drug resistance testing was originally performed using “population” or “bulk” sequencing, which involves amplification of specific viral genes followed by Sanger sequencing. However, Sanger methodology has limited sensitivity for minor variants when present at less than 15–20% of the viral population, while NGS methods can detect drug-resistant mutations (DRMs) present at ~1% [4, 5]. The prototypical virus for NGS-based genotypic resistance testing is HIV-1, and similar to Sanger-based methods, emerging NGS assays have used targeted sequencing of viral genomic regions known to develop resistance mutations [4]. Because it has been studied most extensively, HIV-1 will be used as a paradigm for a detailed discussion below of concepts related to NGS-based testing for viral drug resistance. Cytomegalovirus (CMV) will also be discussed. However, genotypic drug resistance testing is also utilized for the

M. I. Lefterova  
Department of Pathology, Stanford University School of Medicine,  
Stanford, CA, USA

Guardant Health, Redwood City, CA, USA

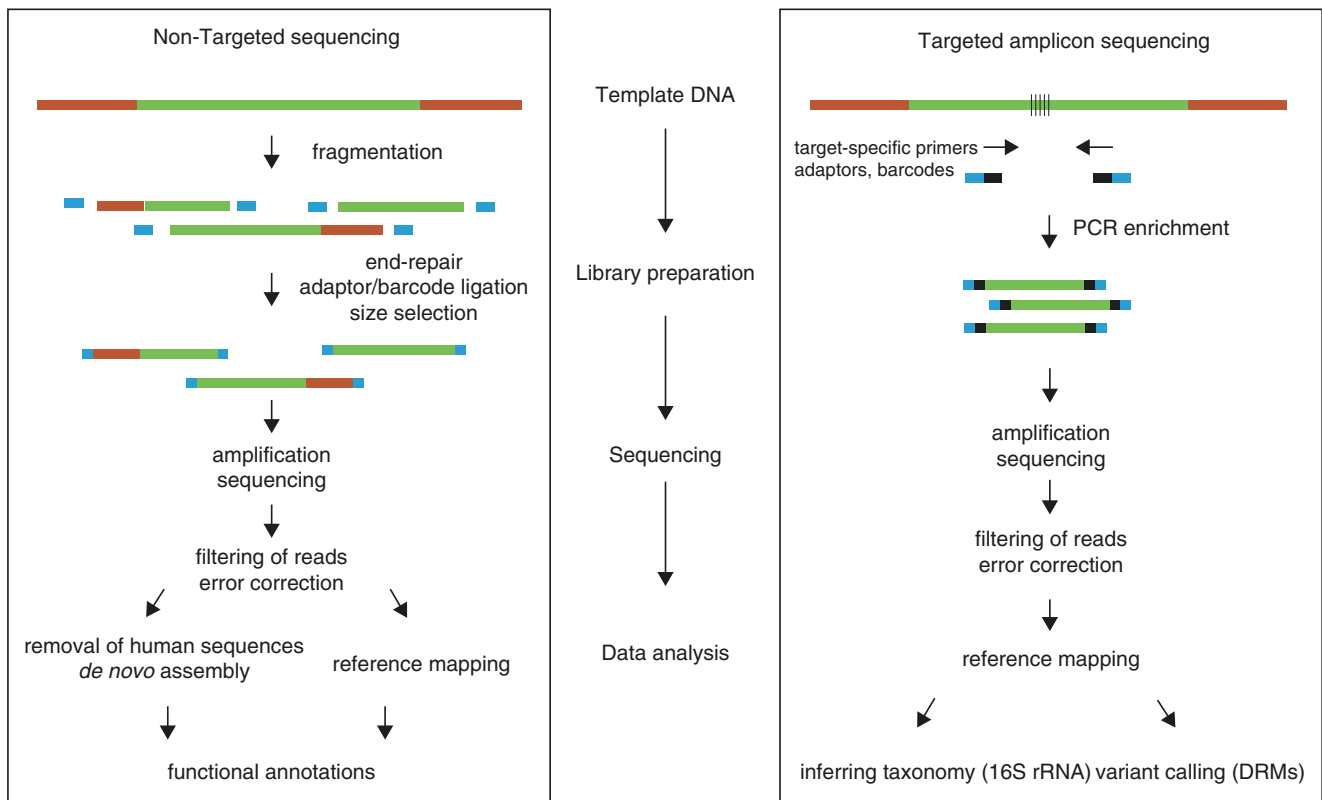
C. J. Suarez  
Department of Pathology, Stanford University School of Medicine,  
Stanford, CA, USA

N. Banaei · B. A. Pinsky (✉)  
Department of Pathology, Stanford University School of Medicine,  
Stanford, CA, USA

Department of Medicine, Division of Infectious Diseases and  
Geographic Medicine, Stanford University School of Medicine,  
Stanford, CA, USA  
e-mail: [bpinsky@stanford.edu](mailto:bpinsky@stanford.edu)

**Table 37.1** Characteristics of current NGS platforms. The specifications of the instruments were obtained from the manufacturers' websites and/or company's representatives and include up-to-date information as of April 2018

Platform	Manufacturer	Sequencing chemistry	Read length (bp)	Throughput	Run time
MiniSeq	Illumina	Sequencing by synthesis – reversible terminator	Up to 2 × 150	1.65–7.5 Gb	7–24 h
MiSeq/MiSeqDx	Illumina	Sequencing by synthesis –reversible terminator	Up to 2 × 300	540 Mb–15 Gb	4–56 h
NextSeq 550/ NextSeq 550Dx	Illumina	Sequencing by synthesis –reversible terminator	Up to 2 × 150	16.25–120 Gb	11–29 h
HiSeq 2500	Illumina	Sequencing by synthesis – reversible terminator	Up to 2 × 250	9 Gb–1 Tb	7 h–11 days
HiSeq 3000/4000	Illumina	Sequencing by synthesis – reversible terminator	2 × 150	105–1500 Gb	<1–3.5 days
HiSeq X	Illumina	Sequencing by synthesis – reversible terminator	2 × 150	1.6–1.8 Tb	< 3 days
NovaSeq 6000	Illumina	Sequencing by synthesis – reversible terminator	Up to 2 × 150	167 Gb–6 Tb	13–45 h
Ion PGM	Life Technologies <sup>a</sup>	Sequencing by synthesis –hydrogen ion detection	200–400	30 Mb–2 Gb	2–7 h
Ion Proton	Life Technologies <sup>a</sup>	Sequencing by synthesis –hydrogen ion detection	200	10 Gb	2–4 h
Ion S5	Life Technologies <sup>a</sup>	Sequencing by synthesis –hydrogen ion detection	200–600	15–50 Gb	6.5–19 h
PacBio RS II	Pacific Biosciences	Single-molecule real-time sequencing	Mean 14 kb	500 Mb–1 Gb	30 min–6 h
PacBio Sequel	Pacific Biosciences	Single-molecule real-time sequencing	Mean 10 kb	5–10 Gb	30 min–20 h

<sup>a</sup>Thermo Fisher Scientific**Fig. 37.1** Illustration of sequencing approaches for diagnosis and monitoring of infectious diseases. Targeted amplicon sequencing (left panel) utilizes target-specific primers for template enrichment, followed by primers that are partially complementary to the target-specific primers (black bars), and contains sequencing adaptors and bar codes (blue bars). Nontargeted or metagenomic sequencing (right panel) utilizes enzymatic or mechanical fragmentation, followed by end repair to

allow ligation of primers that contain sequencing adaptors and bar codes (blue bars). Size selection allows only fragments of a predefined length to be used for sequencing. Bioinformatics removal of human sequences is required since the nucleic acids of the organism of interest frequently constitute less than 1% of the nucleic acid pool. Note that fragmentation libraries may also be made from PCR-enriched amplicons

clinical management of other viral infections, including hepatitis B, hepatitis C, and influenza, and NGS methods are applicable for these viruses as well.

### Human Immunodeficiency Virus Type 1 (HIV-1)

Epidemiologic studies in HIV-1-positive patients have shown that the presence of mutations conferring resistance to highly active antiretroviral therapy (HAART) can predict treatment outcomes [6]. Therefore, genotypic testing for DRMs is currently recommended for therapy-naïve patients when they enter into clinical care and for therapy-experienced patients when they show evidence of virologic failure [7]. A number of studies have compared NGS and Sanger sequencing methods for capturing minority resistant variants, demonstrating that at least half of the DRMs identified by NGS are missed by Sanger sequencing [8, 9]. The presence of such variants has been shown to predict an increased risk for therapy failure [10].

A major consideration when assessing minor variants is distinguishing true mutations from artifacts generated during PCR amplification, library preparation, or sequencing. These include mismatches, insertions/deletions, and PCR-mediated recombination products, known as chimeric sequences [11, 12]. This is particularly problematic for clinical specimens with low virus loads because the numbers of viral copies that are used for library preparation are small and a mixed viral population may not be accurately represented, even with the use of high-fidelity polymerases. Differential amplification of some variants can skew the final PCR product mixture because of stochastic events in early PCR cycles or differences in the efficiency of primer annealing [4, 13]. One possible solution is to estimate empirical error rates for a given NGS assay and for different viral concentrations and then to set thresholds for minor variant detection safely above the empirical error rates. For instance, a plasmid of known genotype can be subjected to NGS and Sanger sequencing, with the assumption that all NGS calls not validated by the Sanger “truth” are due to library preparation and/or NGS errors [14]. Alternatively, the library preparation step can employ primers tagged with a random sequence, such that each template receives a unique identifier. This allows a consensus sequence to be generated for each original template molecule, thus correcting for random errors during library preparation and sequencing [15, 16]. Another approach for addressing PCR bias has been to perform multiple independent amplifications from the same clinical specimen and pool the products to serve as a template for library preparation [17, 18]. Novel bioinformatics tools have also been used to process NGS data in ways that reduce error rates and call authentic low-abundance viral variants [4].

A large number of HIV-1 research studies demonstrating the superior performance of NGS methods compared to Sanger sequencing [4] have resulted in the introduction of several clinical NGS HIV-1 drug resistance assays. The most comprehensive is the DEEPGEN<sup>TM</sup>HIV (developed by University Hospitals Case Medical Center, Cleveland, OH), which assesses for resistance mutations in the protease, reverse transcriptase, and integrase genes, in addition to predicting HIV-1 co-receptor tropism, with mean error rates of 0.37–0.39%, sensitivity for minor variants of 5%, and capacity to multiplex up to 96 samples in a single run [5]. Though not yet available in the USA, Vela Diagnostics have obtained CE marking for their ion PGM-based Sentosa SQ HIV-1 Genotyping Assay for the automated detection of drug resistance mutations in the protease, reverse transcriptase, and integrase genes at a level of 5% [19].

Two other assays that use deep sequencing of the HIV-1 *env* V3 loop for HIV co-receptor tropism are available clinically: the HIV-1 CCR5 tropism test (V3) offered by the British Columbia Centre for Excellence in HIV/AIDS (Vancouver, Canada) [20] and the HIV-1 co-receptor tropism with reflex to ultra-deep sequencing offered by Quest Diagnostics [21]. These assays and DEEPGEN<sup>TM</sup>HIV have been shown to predict non-CCR5 tropism as accurately as the phenotypic gold standard (Trofile, Monogram Biosciences) and to exhibit a higher sensitivity than Sanger sequencing for detecting minor CXCR4-tropic variants.

Importantly, the clinical significance of low-abundance HIV-1 drug resistance variants detected by NGS remains to be fully characterized. Several studies have retrospectively evaluated the impact of low-abundance resistance variants detected by NGS in treatment-naïve patients [8, 9], as well as in treatment-experienced patients with virologic failure [22, 23]. Although patients with low-abundance DRMs detected by NGS alone appear to have a modestly increased risk of failing therapy, in general, the risk of failure is substantially higher with high-abundance mutants that can be demonstrated both by NGS and Sanger sequencing [8].

### Cytomegalovirus (CMV)

CMV is another virus for which genotypic drug resistance testing is clinically useful, particularly in transplant recipients [24, 25]. Rates of CMV drug resistance vary based on patient populations: 5–12.5% in solid organ transplant (SOT) recipients and 2–5% in hematopoietic stem cell transplant (HSCT) recipients [26]. Timely detection of CMV drug resistance is critical because DRMs can accumulate with continued exposure to a drug [26, 27], potentially leading to shortened graft survival and increased morbidity [28, 29]. Furthermore, rational change of therapy following

identification of drug resistance has been shown to lead to more rapid clearance of virus [30]. Mutations conferring resistance to the CMV therapeutics, ganciclovir, foscarnet, and cidofovir have been characterized in two CMV genes, the DNA polymerase *UL54* and the phosphotransferase *UL97*, together representing <6 kb of coding sequence, which makes CMV well suited for an amplicon sequencing NGS-based approach analogous to assays targeting HIV protease and reverse transcriptase. In fact, an NGS assay for CMV *UL54* and *UL97* is demonstrating low overall empirical error rate (0.189%) and reliable detection of CMV DRMs in clinical plasma specimens with a wide range of viral loads [18]. Mutations conferring resistance to the terminase inhibitor, Letemovir, FDA-cleared in 2017, have been identified in several genes encoding members of the terminase complex, primarily *UL56* and less commonly *UL89* and *UL51* [31–33]. Subsequent assays for CMV genotypic resistance will likely include *UL56* and may include other genes important for the development of Letemovir resistance.

The impact of minor-population resistant variants on clinical outcomes in CMV-positive patients has not yet been assessed in large clinical trials. However, there is emerging evidence that NGS can facilitate the detection of impending drug resistance and assist in therapy optimization [27]. NGS studies of viral drug resistance are also expected to identify novel putative DRMs, which, after appropriate phenotypic validation [27], can be incorporated into CMV DRM databases and genotypic interpretation systems, similar to those that exist for HIV-1 [34]. Such automated tools have been shown to improve sequence analysis in addition to expediting and standardizing workflow when compared to manual sequence curation [35].

## Virus Identification in Clinical Specimens

Proof-of-concept studies have demonstrated the ability of nontargeted, metagenomic sequencing to identify common, clinically relevant viruses from a variety of specimen types previously shown to be positive by routine molecular testing [36]. Another area of diagnostic virology where NGS is being successfully applied is for the identification of viral pathogens in clinical scenarios where a viral agent is suspected but not detected by conventional diagnostic methods [37]. Many viruses cannot be cultured or identified by traditional molecular techniques, while other methods such as cloning and Sanger sequencing are laborious, time-consuming, and mainly applicable to sterile samples like cerebrospinal fluid [2]. Microarrays targeting highly conserved regions within viral families are capable of detecting known viruses, but they cannot identify novel pathogens without sequence similarity to oligonucleotides on the array

[37]. In contrast, NGS offers an efficient, highly sensitive, and unbiased alternative for the detection of viruses in clinical specimens [2, 37]. The general approach in such studies is fundamentally different from that used in targeted sequencing. First, the virus of interest is usually not known and therefore cannot be selectively amplified with target-specific primers. Thus, specialized laboratory and bioinformatics strategies are needed to enrich viral RNA or DNA from the predominantly human nucleic acids. Second, a reference sequence may not be available for mapping of sequencing reads if the virus is novel or largely divergent from known related viruses. This necessitates de novo assembly of the viral genome.

As the nucleic acids in clinical specimens are predominantly of host origin, the enrichment of viral and/or depletion of host sequences is an important step for sensitive NGS discovery of viruses in clinical specimens. Laboratory methods for viral particle purification and enrichment include viral culture, ultracentrifugation, density gradient centrifugation, and pretreatment of the sample with nucleases in order to remove host nucleic acids, while preserving capsid-protected viral particles [2, 38]. Nucleic acid amplification methods for enrichment of viral genomes include rolling circle amplification for viruses with a circular genome [39] and use of restriction enzyme sites that are more frequently encountered in viral nucleic acids than human, followed by ligation of adaptors and PCR amplification [40, 41]. Other methods have incorporated hybridization approaches to capture viral nucleic acids with antisense oligonucleotides as baits, although bait design requires at least some prior knowledge of the pathogen [42, 43]. For example, both ViroCap [44] and VirCapSeq-VERT [45] contain probes for capture of all viruses known to infect vertebrates. Similarly, hybridization methods have been designed to deplete human nucleic acids, including methods utilizing CRISPR-based depletion [46]. Furthermore, computational tools have been developed for “subtracting” host sequences from the initial read pool containing mixed human and microbial sequences [47–49]. This filtering step is crucial because viral sequences may comprise <1% of the initial aligned reads [37, 49] (Table 37.2).

Additionally, it is frequently unknown whether a putative viral pathogen contains a DNA or RNA genome, which necessitates processing for total nucleic acid extraction. Amplification with random primers may also be necessary to generate sufficient template for library preparation. An interesting approach to this problem for RNA viruses involves reverse transcription with random primers and cDNA amplification using Phi29 bacteriophage polymerase-based multiple displacement amplification [50]. The choice of sequencing platform (Table 37.1) also requires consideration, as read length and sequence depth may impact virus detection and genome assembly [51].

**Table 37.2** Select studies describing culture-independent NGS pathogen identification from primary human clinical specimens

Pathogen name	Clinical information	Specimen(s)	Method	Pathogen aligned reads	Citation(s)
<b>Viruses</b>					
Merkel cell polyomavirus	Rare but aggressive neuroectodermal tumor of the immune compromised	Total RNA from Merkel cell carcinoma tissues	454	0.000003% (1/382,747)	[96]
LCMV-related arenavirus	Fatal febrile illness with sepsis and encephalopathy in three solid organ transplant recipients with the same donor	Total RNA from pooled tissue	454	0.0001% (14/103,632)	[97]
Lujovirus	Hemorrhagic fever outbreak in Southern Africa	Total RNA from liver tissue and serum	454	Not Provided <sup>a</sup>	[98]
Enterovirus 109	Influenza-like illness	Total RNA from nasopharyngeal swab	Illumina	0.00001% (119/10,400,000)	[99]
Influenza A	Upper respiratory illness	Total RNA from nasopharyngeal swabs	Illumina	0.001–0.0001% <sup>b</sup>	[100]
Yellow fever virus	Fatal hemorrhagic fever	Total RNA from serum	454	0.5% (3229/599,158)	[101, 102]
Rubella virus	Anterior and intermediate uveitis	Total RNA from intraocular fluid	Illumina	0.41% (585/1,684,220)	[93]
West Nile virus	Acute meningoencephalitis in a renal transplant recipient	Total RNA from CSF	Illumina	0.001% (101/7,777,470)	[103]
Cache Valley virus	Chronic meningoencephalitis in an agammaglobulinemic patient	Total RNA from CSF	Illumina	0.00002% (5/25,069,677)	[104]
<b>Bacteria</b>					
<i>Francisella tularensis</i>	Abscess	Total DNA from abscess drainage	Illumina	0.002% (833/38,285,502)	[105]
<i>Leptospira santarosai</i>	Recurrent meningoencephalitis in a pediatric patient with SCID	Total DNA from CSF	Illumina	0.016% (475/3,063,784)	[76]
<i>Brucella melitensis</i>	Headache, nausea, and myoclonus in a pediatric patient	Total DNA from CSF	Illumina	0.0012% (277/23,638,587)	[106]
<b>Parasites</b>					
<i>Balamuthia mandrillaris</i>	Endophthalmitis and meningoencephalitis	Total RNA from cerebrospinal fluid	Illumina	0.00002% (5/19,642,962)	[107]

LCMV lymphocytic choriomeningitis virus, HSCT hematopoietic stem cell transplant, FFPE formalin-fixed, paraffin embedded, SCID severe combined immune deficiency

<sup>a</sup>Obtained coverage of 5.6 kb of a 10.4 kb genome

<sup>b</sup>The number of influenza A-specific and total reads varied between samples

Perhaps the most critical aspect of successful viral discovery is the choice of bioinformatics tools. When the reference genome is known, as in amplicon sequencing experiments, read alignment software typically applies stringent mismatch rules in order to minimize errors. In contrast, with unknown pathogens it may be impossible to map reads to publicly available viral databases if the target virus is highly divergent. Instead bioinformatics tools must assemble reads into contiguous sequences (contigs) by identifying overlapping sequences between reads, followed by contig assembly into genomes [52]. Sequencing methods that produce long reads and therefore linkage information facilitate contig assembly (Table 37.1). Sequences assembled this way can be compared to public databases by using algorithms with relaxed stringency in order to identify related viruses. Repetitive sequences pose a significant challenge in de novo assembly, because they can interfere with PCR amplification as well as accurate genomic mapping. Computational and experimental strategies are being developed to address such issues [2, 52]. Table 37.2 summarizes several representative

studies in which the NGS approaches described above have been used to identify viral pathogens in patients with infectious syndromes of unclear etiology.

An important caveat to viral discovery is that demonstrating the presence of a virus in a patient with disease does not automatically imply pathogenicity. Traditionally, proving that a microorganism is the causative agent of disease has depended on fulfilling Koch's postulates: a putative etiologic agent is found in affected hosts but not healthy controls; it is propagated in culture and can reproduce the disease when a healthy host is inoculated. However, it is increasingly evident that many viruses cannot be cultured, which has prompted the revision of traditional approaches to prove causality for a microorganism in a disease [53]. Such guidelines eliminate the requirement for microorganism isolation but expand on the rigor with which the association between microorganism and disease is established. For example, it may be necessary to demonstrate the presence of virus in affected tissues using immunostaining or molecular methods, to establish a correlation between viral copy number and

disease severity or to show seroconversion from acute to convalescent plasma specimens.

## Specific Applications in Clinical Bacteriology, Mycobacteriology, and Mycology

### Identification by Targeted or Nontargeted Sequencing

Genomic approaches are also likely to assist in the diagnosis and management of bacterial, mycobacterial, and fungal infections, including pathogen identification, as well as characterization of virulence factors, strain typing, and antibiotic resistance markers. Clinical microbiology has traditionally relied on isolation of pathogens by culture followed by biochemical tests and more recently matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) to identify the genus or species of the infecting organism [54]. These assays are inexpensive, at least on a per test basis, have a rapid turnaround time, and are therefore appropriate as first-line diagnostics. However, proof-of-concept studies have shown that NGS holds significant potential for microbial identification from primary human specimens, both by targeted amplicon sequencing of ribosomal RNA (rRNA) genes and nontargeted, metagenomics [3].

### Ribosomal RNA Sequencing

rRNA gene sequencing by Sanger is routinely used for bacterial and fungal identification in clinical microbiology laboratories [55, 56]. Furthermore, rRNA sequencing by NGS is the basis for studies of the microbiome. For bacteria, these approaches employ primers targeting conserved 16S rRNA sequences, with variable intervening regions that provide sufficient sequence diversity for taxonomic assignment, often to the species level. The choice of primers is important because certain areas of 16S rRNA genes may allow amplification of a broader spectrum of bacteria than others [57]. Additionally, in some cases, classification may only be possible to the family or genus level because the amount of sequence variation may be insufficient for a species-level identification [58].

Informatics is critical for the interpretation of 16S rRNA sequencing data that is obtained by NGS. The length, number, and quality of sequencing reads, as well as possible bacterial contamination of reagents, are all factors that can impact pathogen identification and introduce bias in microbiome diversity assessments [59]. For example, commonly used DNA extraction kits are frequently contaminated with environmental bacteria, leading to overestimation of bacterial diversity in specimens with low starting bacterial loads such as cerebrospinal fluid, blood, or tissue biopsies [60].

Data analysis tools include error correction methods [58, 61] and removal of amplification-derived chimeric sequences [62, 63]. Processed data or raw sequences can be analyzed in dedicated pipelines such as QIIME [64], the Ribosomal Database Project [65] and mothur [62], which cluster similar sequences into operational taxonomic units (OTUs) based on at least 97% sequence identity [3], followed by phylogenetic analyses. In addition, efforts are under way to standardize the use of these pipelines by establishing quality-filtering parameters based on the sequencing platform and the quality of sequencing data [66]. Importantly, the accuracy of bacterial identification largely depends on the scope and completeness of reference databases used for analysis. A number of extensive databases have been created for 16S rRNA sequences: for example, SILVA ([www.arb-silva.de](http://www.arb-silva.de)) containing >three million small subunit and >250,000 large subunit bacterial rRNA gene sequences [67] or Greengenes (<http://www.greengenes.secondgenome.com/downloads>) which can calculate taxonomic relationships based on >400,000 16S rRNA sequences [68].

Targeted 16S rRNA sequencing by NGS may have immediate clinical application for the characterization of mixed infections, particularly those containing uncultivable or nonviable organisms. This approach has been successfully applied directly to brain abscess material, lymph node biopsy tissue, cystic fibrosis (CF) sputa, and mastoid abscess material [69, 70].

Sequencing of 16S rRNA targets by NGS has also been used to study the genomic diversity of bacterial communities, or the microbiome, in health and various disease states. For example, sequencing of the 16S rRNA hypervariable region was used to study bacterial vaginosis, revealing increased bacterial heterogeneity compared to the healthy state [71]. On the other hand, studies of microbiome in the lower airway of cystic fibrosis patients [72] and stool of patients with *Clostridium difficile* infections (CDI) [73] or inflammatory bowel disease (IBD) [74] have shown that disease progression is marked by decreasing bacterial diversity, which may be related to escalating antibiotic exposures. Such results indicate that certain disease states may be driven by disturbances in the normal structure and diversity of a microbial community rather than the action of individual pathogens. Although genomic approaches may elucidate the mechanisms by which changes in the microbiome contribute to disease, the diagnostic utility of characterizing the microbiome in patient management remains to be established.

### Metagenomic Sequencing

In contrast to rRNA-based NGS approaches, nontargeted approaches allow more detailed functional and taxonomic analyses, either when a cultured isolate is tested or an entire microbial community is being characterized, a field termed

metagenomics [54, 58]. WGS methods can also be helpful for bacterial pathogen discovery in patients with suspected infections where culture and other standard diagnostic methods have failed. In such scenarios, a direct patient specimen can be sequenced in a relatively unbiased way, similar to what is described above for viral discovery. The potential diagnostic utility of this approach was demonstrated in a pediatric case of severe combined immunodeficiency and recurrent meningoencephalitis, in which WGS coupled with a rapid, dedicated bioinformatics pipeline [75] detected *Leptospira santarosai* sequences in cerebrospinal fluid (CSF) within 48 h of specimen receipt [76]. Table 37.2 shows representative studies in which WGS was used for the culture-independent identification of bacterial pathogens in patients with infectious syndromes of unclear etiology.

Analysis of metagenomic data poses even more challenges than those discussed for 16S rRNA sequencing [58]. When sequencing direct clinical specimens, data need to be filtered for human sequences and sequencing errors. In addition, the putative bacterial reads have to be aligned to reference genomes or subjected to de novo assembly of contigs in order for gene predictions to be made and biological functions to be assigned [58, 75]. Examples of pipelines that have been used for clinical pathogen identification include SURPI (sequence-based ultrarapid pathogen identification) [75, 76] and Taxonomer [77], among many others. However, both taxonomic and functional annotations may be limited by the availability of reference genomes. In that respect, large endeavors exploring bacterial metagenomics in the human host, such as the Human Microbiome Project [78] and Metagenomics of the Human Intestinal Tract (MetaHit) project [79], are actively expanding bacterial genomic databases.

Several groups have assessed the feasibility of nontargeted sequencing for bacterial identification and characterization in a clinical microbiology laboratory. One study tested the feasibility of this approach for routine use by sequencing 130 cultured isolates, including aerobic and anaerobic bacteria, mycobacteria, and fungi [80]. The steps from colony harvest to acquisition of analyzable data took ~55 h, with most of the time attributable to the sequencing run (39 h). Comparison of these sequencing results to identification by MALDI-TOF-MS, in addition to conventional culture and biochemical methods, demonstrated good correlation: 115/130 samples (88.5%) showed concordant results, while 15/130 could not be identified due to insufficient coverage or absence of applicable reference genomes in publicly available databases (mainly for sterile molds). Thus nontargeted sequencing was able to identify the majority of organisms identified by conventional methods; however, the turnaround time was substantially slower, and a cost analysis was not performed.

## Genotypic Pathogen Characterization

In addition to organism identification, NGS methods can be used to identify genotypic markers of drug resistance and virulence, as well as strain typing [54, 81]. Although phenotypic antimicrobial resistance testing is relatively well standardized, it is available for a limited number of organisms and can take up to several weeks for slow-growing organisms like *Mycobacterium tuberculosis* [82]. Molecular assays with improved sensitivity and turnaround times already exist for some resistance markers; however, resistance to an antimicrobial class can be mediated by several molecular mechanisms, necessitating multiple individual tests or panel testing [82]. Whole-genome-based genotyping, therefore, could simplify workflow and eliminate the need for individual PCR-based assays by simultaneously interrogating all possible genotypic resistance mechanisms, especially if sequencing is being performed for other purposes, such as identification, strain typing, or to detect toxin genes [83]. Use of this approach is likely to expand as new drug resistance mechanisms are characterized and catalogued in publicly available databases such as ResFinder [84] and ARG-ANNOT [85], which use BLAST to query a user-supplied sequence against a curated list of bacterial antimicrobial resistance genes. A number of proof-of-concept studies have assessed the ability of NGS to predict bacterial drug resistance patterns and have been reviewed by Koser et al. [83]. Though genotypic resistance prediction appears feasible, susceptibility determination is nuanced and challenging. Notably, a drawback of genotypic assays is that they do not provide a quantitative measure of antimicrobial susceptibility. In particular, if the presence of a resistance gene or mutation confers variable or inducible resistance, phenotypic assays will still be required [82]. Similarly, when used alone, sequencing may fail to predict a resistance pattern if it has not been characterized genetically or is absent from a given database. For these reasons, whole-genome sequencing is unlikely to replace existing cost-effective resistance assays (phenotypic or molecular) for fast-growing organisms even as cost and turnaround time for sequencing assays continue to decrease. The greatest utility of whole-genome-based drug resistance testing may be for slow-growing organisms such as *M. tuberculosis*, where multidrug regimens are used, phenotypic testing is complex and available for a limited number of drugs, and the number of genes and intergenic regions that need to be targeted for a comprehensive molecular assay is prohibitively large for a targeted approach [86, 87].

Another area where bacterial whole-genome sequencing is being implemented for clinical purposes is for strain typing in hospital outbreak investigations. Traditionally strain typing has been performed either by fragment analysis methods, e.g., pulsed field gel electrophoresis, or by sequence-

based techniques, such as multilocus sequence typing [88]. However, typing schemes exist only for a limited number of organisms, and currently typing is performed primarily in reference and public health laboratories, which means that results are frequently not available within a clinically actionable time frame. In contrast, analysis of single nucleotide polymorphisms (SNPs) based on bacterial whole-genome data can be performed during ongoing outbreaks, does not depend on the availability of established typing schemes, and has higher resolution than most existing sequence-based typing Schemes [54]. In this approach, whole-genome sequencing data are aligned to a reference genome, SNPs are identified and filtered based on preestablished quality metrics, and then phylogenetic analysis is performed to assess the relatedness of bacterial isolates. The feasibility of this approach for reconstructing transmission pathways in hospital outbreak investigations has been demonstrated in a number of studies [54]. However, it remains to be shown whether the use of whole-genome sequencing in the setting of hospital outbreaks will be cost-effective and will be associated with prevention of transmission events.

### Validation, Quality Control, and Maintenance of Proficiency

The use of any diagnostic test in the clinical laboratory requires analytical and clinical validation, as well as the careful monitoring and documentation of quality control and proficiency testing (Table 37.3). In that regard, NGS performed in the clinical laboratory for patient care differs from

**Table 37.3** Assessment of the performance characteristics of NGS-based tests for clinical microbiology

Performance characteristic	Approach to evaluation
Accuracy	Use of specimens with known findings and confirmation of additional findings detected during validation by an orthogonal method
Precision	Reproducibility (between-run precision): sequencing of the same samples on different runs Repeatability (within-run precision): Sequencing of the same samples in replicates within a run  Between library precision: sequencing different library preparations of the same samples on the same sequencing run [108]
Analytical sensitivity	Microbial variant detection: mixes of known variant strains and wild-type strains at different percentages and at low, medium, and high levels (e.g., viral loads) Microbial identification: serial dilutions of samples in an appropriate matrix containing a known pathogen(s) coupled with an estimation of the minimum coverage needed to detect the pathogen
Analytical specificity	Microbial variant detection and microbial identification: estimation of the false-positive rate at various read depths

NGS performed in the research setting, even though the sequencing methods may be the same. As such, the American College of Medical Genetics and Genomics (ACMG) has published detailed clinical laboratory standards for NGS [89]. Furthermore, the College of American Pathologists (CAP) has developed an NGS checklist for accreditation of molecular pathology laboratories performing clinical NGS testing [90]. The molecular pathology NGS checklist details requirements for documentation, validation, quality control, and quality monitoring for both the wet bench work and bioinformatics and includes guidelines for data storage, as well as the assessment and implementation of new technology and software releases. Though this checklist has been updated to include examples relevant to NGS for infectious diseases, it is anticipated that in the future, the microbiology checklist will contain a separate section for NGS tailored specifically for microbiology. To further assist in the validation of NGS-based assays for infectious diseases, the American Society for Microbiology and CAP published a manuscript describing the challenges and potential solutions for validating metagenomic pathogen detection tests in clinical laboratories [91].

However, the application of NGS in clinical infectious disease testing poses unique challenges that are distinct from the diagnostic settings of human inherited diseases or cancer. For example, as NGS is increasingly adopted for clinical microbiology, well-characterized and extensively sequenced reference microbial organisms will be required for use as controls and proficiency material. In order to supplement reference strains, mock sequence data may also be necessary to ensure adequate bioinformatics pipelines. These *in silico* controls and proficiency challenges will be particularly important for the clinical characterization of the microbial metagenome, low-level DRM detection, and the identification of organisms that are unculturable or difficult to culture.

NGS technologies that are being used for clinical infectious disease testing are currently being performed as laboratory-developed tests, as no clinical microbiology NGS tests have yet been approved by the United States Food and Drug Administration (FDA). Nevertheless, the FDA is keenly interested in the regulatory oversight of NGS in clinical microbiology, particularly for microbial identification and the detection of antimicrobial resistance markers. As such, the FDA has published a discussion paper detailing clinical applications and validation approaches for the regulatory approval/clearance of NGS diagnostic devices for clinical microbiology [92]. Of note, this document reports that the FDA is engaged in the development of a database (FDA MicroDB) comprised of >550 high-quality, “regulatory-grade” sequences from clinically relevant bacterial microorganisms to be used in the pathway for regulatory approval. The availability of FDA-approved infectious dis-



ease NGS in vitro diagnostics will likely aid in the standardization of specimen handling, library preparation and sequencing, as well as data interpretation, in order to ensure the accuracy and reproducibility of NGS-derived genotypic results. This standardization and quality assurance may be particularly important given that contaminating microbial DNA is ubiquitously found in commonly used extraction kits and reagents used for NGS, as well as “sterile” specimen transport containers [60].

## Conclusions

In this chapter we have reviewed areas of clinical microbiology in which next-generation sequencing approaches have been used to identify and characterize medically important pathogens. While many of these studies have been conducted as proof-of-concept experiments or research investigations, NGS-based testing has already been adopted in select diagnostic microbiology laboratories, including academic clinical laboratories, large commercial reference laboratories, and startup companies. Routine applications are likely to increase as cost, turnaround time, and complexity decrease sufficiently to make NGS complementary to existing affordable, standardized, and considerably simpler methods. As technologies like this one are developed and evaluated, the use of NGS for infectious diseases testing may become more widespread.

Targeted NGS assays relying on amplicon sequencing, such as HIV drug resistance testing, were the first to be introduced clinically given the sensitivity advantages over Sanger sequencing and the accumulating data supporting the clinical relevance of low-abundance resistance mutations. NGS-based amplicon sequencing of ribosomal RNA genes may also become more commonly used for identification of pathogenic bacteria and fungi when there is high suspicion for infection and culture is negative or not available or when mixed infections are suspected. Metagenomic strategies may also be useful for pathogen identification in sterile specimens if testing can be optimized to provide clinically actionable data faster than culture or currently available molecular methods. Importantly, the ability of NGS methods and bioinformatics pipelines to accurately identify and characterize pathogens will need to be rigorously validated and compared with traditional diagnostic techniques [93, 94].

The greatest attraction of genomic approaches is that metagenomics sequencing could provide all relevant information about a pathogen in a single assay, including species identification, strain typing, virulence determination, and antimicrobial resistance. In practice, widespread implementation of NGS in clinical microbiology laboratories will require acquisition of costly new equipment and, in particular, the training of personnel in methods that are reliant on

bioinformatics. Bioinformatics pipelines will need to provide user-friendly interfaces that allow the user to input data directly from the sequencing instrument and receive best-hit matches to comprehensive and well-curated reference genome databases [54].

Thus, at this point in time, NGS methods are expected to supplement, rather than replace, conventional diagnostic testing. An important hurdle, even in the most sophisticated of clinical laboratories, is that genotype-phenotype correlations for many clinically relevant microorganisms are unknown, although large-scale metagenomic efforts like the Human Microbiome Project will undoubtedly define numerous new associations between sequence and function. Ultimately, the tremendous promise of NGS methods for diagnostic infectious disease testing will require the successful development of clinical microbiologists capable of interpreting and evaluating NGS data and placing these data in the appropriate clinical context.

**Acknowledgments** Material and references were used from Lefterova MI, Banaei N, and Pinsky BA. Genomic Applications in the Clinical Management of Infectious Diseases. In: Netto GJ, Schrijver I, eds. *Genomic Applications in Pathology*. 1st Ed. Springer. 2015. The authors thank the AMP NGS in ID Work Group, a joint project of the AMP Infectious Diseases Subdivision Leadership and Clinical Practice Committee for their contributions to the previous version of this manuscript [95].

## References

1. Loman NJ, Constantinidou C, Chan JZ, et al. High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. *Nat Rev Microbiol*. 2012;10(9):599–606.
2. Radford AD, Chapman D, Dixon L, Chantrey J, Darby AC, Hall N. Application of next-generation sequencing technologies in virology. *J Gen Virol*. 2012;93(Pt 9):1853–68.
3. Weinstock GM. Genomic approaches to studying the human microbiota. *Nature*. 2012;489(7415):250–6.
4. Vrancken B, Lequime S, Theys K, Lemey P. Covering all bases in HIV research: unveiling a hidden world of viral evolution. *AIDS Rev*. 2010;12(2):89–102.
5. Gibson RM, Schmotzer CL, Quinones-Mateu ME. Next-generation sequencing to help monitor patients infected with HIV: ready for clinical use? *Curr Infect Dis Rep*. 2014;16(4):401.
6. Dunn DT, Coughlin K, Cane PA. Genotypic resistance testing in routine clinical care. *Curr Opin HIV AIDS*. 2011;6(4):251–7.
7. Cortez KJ, Maldarelli F. Clinical management of HIV drug resistance. *Viruses*. 2011;3(4):347–78.
8. Simen BB, Simons JF, Hullsiek KH, et al. Low-abundance drug-resistant viral variants in chronically HIV-infected, antiretroviral treatment-naïve patients significantly impact treatment outcomes. *J Infect Dis*. 2009;199(5):693–701.
9. Lataillade M, Chiarella J, Yang R, et al. Prevalence and clinical significance of HIV drug resistance mutations by ultra-deep sequencing in antiretroviral-naïve subjects in the CASTLE study. *PLoS One*. 2010;5(6):e10952.
10. Li JZ, Paredes R, Ribaudo HJ, et al. Low-frequency HIV-1 drug resistance mutations and risk of NNRTI-based antiretroviral treatment failure: a systematic review and pooled analysis. *JAMA*. 2011;305(13):1327–35.

11. Gorzer I, Guelly C, Trajanoski S, Puchhammer-Stockl E. The impact of PCR-generated recombination on diversity estimation of mixed viral populations by deep sequencing. *J Virol Methods*. 2010;169(1):248–52.
12. Harismendy O, Ng PC, Strausberg RL, et al. Evaluation of next generation sequencing platforms for population targeted sequencing studies. *Genome Biol*. 2009;10(3):R32.
13. Mild M, Hedskog C, Jernberg J, Albert J. Performance of ultra-deep pyrosequencing in analysis of HIV-1 pol gene variation. *PLoS One*. 2011;6(7):e22741.
14. Wang C, Mitsuya Y, Gharizadeh B, Ronaghi M, Shafer RW. Characterization of mutation spectra with ultra-deep pyrosequencing: application to HIV-1 drug resistance. *Genome Res*. 2007;17(8):1195–201.
15. Jabara CB, Jones CD, Roach J, Anderson JA, Swanson R. Accurate sampling and deep sequencing of the HIV-1 protease gene using a Primer ID. *Proc Natl Acad Sci U S A*. 2011;108(50):20166–71.
16. Clutter DS, Zhou S, Varghese V, et al. Prevalence of drug-resistant minority variants in untreated HIV-1-infected individuals with and those without transmitted drug resistance detected by Sanger sequencing. *J Infect Dis*. 2017;216(3):387–91.
17. Vandembroucke I, Van Marck H, Mostmans W, et al. HIV-1 V3 envelope deep sequencing for clinical plasma specimens failing in phenotypic tropism assays. *AIDS Res Ther*. 2010;7:4.
18. Sahoo MK, Lefterova MI, Yamamoto F, et al. Detection of cytomegalovirus drug resistance mutations by next-generation sequencing. *J Clin Microbiol*. 2013;51(11):3700–10.
19. Tzou PL, Ariyaratne P, Varghese V, et al. Comparison of an in vitro diagnostic next-generation sequencing assay with Sanger sequencing for HIV-1 genotypic resistance testing. *J Clin Microbiol*. 2018;56(6):pii: e00105-18.
20. Swenson LC, Mo T, Dong WW, et al. Deep sequencing to infer HIV-1 co-receptor usage: application to three clinical trials of maraviroc in treatment-experienced patients. *J Infect Dis*. 2011;203(2):237–45.
21. Kagan RM, Johnson EP, Siaw M, et al. A genotypic test for HIV-1 tropism combining Sanger sequencing with ultra-deep sequencing predicts virologic response in treatment-experienced patients. *PLoS One*. 2012;7(9):e46334.
22. Le T, Chiarella J, Simen BB, et al. Low-abundance HIV drug-resistant viral variants in treatment-experienced persons correlate with historical antiretroviral use. *PLoS One*. 2009;4(6):e6079.
23. Codoner FM, Pou C, Thielen A, et al. Added value of deep sequencing relative to population sequencing in heavily pre-treated HIV-1-infected subjects. *PLoS One*. 2011;6(5):e19461.
24. Kotton CN, Kumar D, Caliendo AM, et al. International consensus guidelines on the management of cytomegalovirus in solid organ transplantation. *Transplantation*. 2010;89(7):779–95.
25. Ljungman P, Reusser P, de la Camara R, et al. Management of CMV infections: recommendations from the infectious diseases working party of the EBMT. *Bone Marrow Transplant*. 2004;33(11):1075–81.
26. Lurain NS, Chou S. Antiviral drug resistance of human cytomegalovirus. *Clin Microbiol Rev*. 2010;23(4):689–712.
27. Chou S, Ercolani RJ, Sahoo MK, Lefterova MI, Strasfeld LM, Pinsky BA. Improved detection of emerging drug-resistant mutant cytomegalovirus subpopulations by deep sequencing. *Antimicrob Agents Chemother*. 2014;58(8):4697–702.
28. Li F, Kenyon KW, Kirby KA, Fishbein DP, Boeckh M, Limaye AP. Incidence and clinical features of ganciclovir-resistant cytomegalovirus disease in heart transplant recipients. *Clin Infect Dis*. 2007;45(4):439–47.
29. Limaye AP, Corey L, Koelle DM, Davis CL, Boeckh M. Emergence of ganciclovir-resistant cytomegalovirus disease among recipients of solid-organ transplants. *Lancet*. 2000;356(9230):645–9.
30. Myhre HA, Haug Dorenberg D, Kristiansen KI, et al. Incidence and outcomes of ganciclovir-resistant cytomegalovirus infections in 1244 kidney transplant recipients. *Transplantation*. 2011;92(2):217–23.
31. Chou S. Comparison of cytomegalovirus terminase gene mutations selected after exposure to three distinct inhibitor compounds. *Antimicrob Agents Chemother*. 2017;61(11)
32. Chou S. Rapid in vitro evolution of human cytomegalovirus UL56 mutations that confer letermovir resistance. *Antimicrob Agents Chemother*. 2015;59(10):6588–93.
33. Chou S. A third component of the human cytomegalovirus terminase complex is involved in letermovir resistance. *Antivir Res*. 2017;148:1–4.
34. Tang MW, Liu TF, Shafer RW. The HIVdb system for HIV-1 genotypic resistance interpretation. *Intervirology*. 2012;55(2):98–101.
35. Woods CK, Brumme CJ, Liu TF, et al. Automating HIV drug resistance genotyping with RECALL, a freely accessible sequence analysis tool. *J Clin Microbiol*. 2012;50(6):1936–42.
36. Petty TJ, Cordey S, Padioulet I, et al. Comprehensive human virus screening using high-throughput sequencing with a user-friendly representation of bioinformatics analysis: a pilot study. *J Clin Microbiol*. 2014;52(9):3351–61.
37. Chiu CY. Viral pathogen discovery. *Curr Opin Microbiol*. 2013;16(4):468–78.
38. Barzon L, Lavezzo E, Militello V, Toppo S, Palu G. Applications of next-generation sequencing technologies to diagnostic virology. *Int J Mol Sci*. 2011;12(11):7861–84.
39. Rector A, Tachezy R, Van Ranst M. A sequence-independent strategy for detection and cloning of circular DNA virus genomes by using multiply primed rolling-circle amplification. *J Virol*. 2004;78(10):4993–8.
40. de Vries M, Deijns M, Canuti M, et al. A sensitive assay for virus discovery in respiratory clinical samples. *PLoS One*. 2011;6(1):e16118.
41. Pyrc K, Jebbink MF, Berkhout B, van der Hoek L. Detection of new viruses by VIDISCA. Virus discovery based on cDNA-amplified fragment length polymorphism. *Methods Mol Biol*. 2008;454:73–89.
42. Depledge DP, Palser AL, Watson SJ, et al. Specific capture and whole-genome sequencing of viruses from clinical samples. *PLoS One*. 2011;6(11):e27805.
43. Duncavage EJ, Magrini V, Becker N, et al. Hybrid capture and next-generation sequencing identify viral integration sites from formalin-fixed, paraffin-embedded tissue. *J Mol Diagn*. 2011;13(3):325–33.
44. Wylie TN, Wylie KM, Herter BN, Storch GA. Enhanced virome sequencing using targeted sequence capture. *Genome Res*. 2015;25(12):1910–20.
45. Briese T, Kapoor A, Mishra N, et al. Virome capture sequencing enables sensitive viral diagnosis and comprehensive virome analysis. *MBio*. 2015;6(5):e01491-15.
46. Gu W, Crawford ED, O'Donovan BD, et al. 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.
47. Bhaduri A, Qu K, Lee CS, Ungewickell A, Khavari PA. Rapid identification of non-human sequences in high-throughput sequencing datasets. *Bioinformatics*. 2012;28(8):1174–5.
48. Kostic AD, Ojesina AI, Pedamallu CS, et al. PathSeq: software to identify or discover microbes by deep sequencing of human tissue. *Nat Biotechnol*. 2011;29(5):393–6.
49. Yozwiak NL, Skewes-Cox P, Stenglein MD, Balmaseda A, Harris E, DeRisi JL. Virus identification in unknown tropical febrile illness cases using deep sequencing. *PLoS Negl Trop Dis*. 2012;6(2):e1485.

50. Berthet N, Reinhardt AK, Leclercq I, et al. Phi29 polymerase based random amplification of viral RNA as an alternative to random RT-PCR. *BMC Mol Biol.* 2008;9:77.
51. Cheval J, Sauvage V, Frangeul L, et al. Evaluation of high-throughput sequencing for identifying known and unknown viruses in biological samples. *J Clin Microbiol.* 2011;49(9):3268–75.
52. Miller JR, Koren S, Sutton G. Assembly algorithms for next-generation sequencing data. *Genomics.* 2010;95(6):315–27.
53. Falkow S. Molecular Koch's postulates applied to bacterial pathogenicity – a personal recollection 15 years later. *Nat Rev Microbiol.* 2004;2(1):67–72.
54. Koser CU, Ellington MJ, Cartwright EJ, et al. Routine use of microbial whole genome sequencing in diagnostic and public health microbiology. *PLoS Pathog.* 2012;8(8):e1002824.
55. Clarridge JE 3rd. Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. *Clin Microbiol Rev.* 2004;17(4):840–62. table of contents
56. Balajee SA, Sigler L, Brandt ME. DNA and the classical way: identification of medically important molds in the 21st century. *Med Mycol.* 2007;45(6):475–90.
57. Klindworth A, Pruesse E, Schweer T, et al. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* 2013;41(1):e1.
58. Kuczynski J, Lauber CL, Walters WA, et al. Experimental and analytical tools for studying the human microbiome. *Nat Rev Genet.* 2012;13(1):47–58.
59. Cox MJ, Cookson WO, Moffatt MF. Sequencing the human microbiome in health and disease. *Hum Mol Genet.* 2013;22(R1):R88–94.
60. Salter SJ, Cox MJ, Turek EM, et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol.* 2014;12:87.
61. Schloss PD, Gevers D, Westcott SL. Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. *PLoS One.* 2011;6(12):e27310.
62. Schloss PD, Westcott SL, Ryabin T, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol.* 2009;75(23):7537–41.
63. Haas BJ, Gevers D, Earl AM, et al. Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Res.* 2011;21(3):494–504.
64. Caporaso JG, Kuczynski J, Stombaugh J, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods.* 2010;7(5):335–6.
65. Cole JR, Wang Q, Cardenas E, et al. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res.* 2009;37(Database issue):D141–5.
66. Bokulich NA, Subramanian S, Faith JJ, et al. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nat Methods.* 2013;10(1):57–9.
67. Quast C, Pruesse E, Yilmaz P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 2013;41(Database issue):D590–6.
68. McDonald D, Price MN, Goodrich J, et al. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J.* 2012;6(3):610–8.
69. Salipante SJ, Sengupta DJ, Rosenthal C, et al. Rapid 16S rRNA next-generation sequencing of polymicrobial clinical samples for diagnosis of complex bacterial infections. *PLoS One.* 2013;8(5):e65226.
70. Salipante SJ, Hoogstraat DR, Abbott AN, et al. Coinfection of *Fusobacterium nucleatum* and *Actinomyces israelii* in mastoiditis diagnosed by next-generation DNA sequencing. *J Clin Microbiol.* 2014;52(5):1789–92.
71. Srinivasan S, Hoffman NG, Morgan MT, et al. Bacterial communities in women with bacterial vaginosis: high resolution phylogenetic analyses reveal relationships of microbiota to clinical criteria. *PLoS One.* 2012;7(6):e37818.
72. Zhao J, Schloss PD, Kalikin LM, et al. Decade-long bacterial community dynamics in cystic fibrosis airways. *Proc Natl Acad Sci U S A.* 2012;109(15):5809–14.
73. Dethlefsen L, Huse S, Sogin ML, Relman DA. The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing. *PLoS Biol.* 2008;6(11):e280.
74. Morgan XC, Tickle TL, Sokol H, et al. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biol.* 2012;13(9):R79.
75. Naccache SN, Federman S, Veeraghavan N, et al. A cloud-compatible bioinformatics pipeline for ultrarapid pathogen identification from next-generation sequencing of clinical samples. *Genome Res.* 2014;24(7):1180–92.
76. Wilson MR, Naccache SN, Samayoa E, et al. Actionable diagnosis of neuroleptospirosis by next-generation sequencing. *N Engl J Med.* 2014;370(25):2408–17.
77. Flygare S, Simmon K, Miller C, et al. Taxonomer: an interactive metagenomics analysis portal for universal pathogen detection and host mRNA expression profiling. *Genome Biol.* 2016;17(1):111.
78. Group NHW, Peterson J, Garges S, et al. The NIH Human Microbiome Project. *Genome Res.* 2009;19(12):2317–23.
79. Qin J, Li R, Raes J, et al. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature.* 2010;464(7285):59–65.
80. Long SW, Williams D, Valson C, et al. A genomic day in the life of a clinical microbiology laboratory. *J Clin Microbiol.* 2013;51(4):1272–7.
81. Didelot X, Bowden R, Wilson DJ, Peto TEA, Crook DW. Transforming clinical microbiology with bacterial genome sequencing. *Nat Rev Genet.* 2012;13(9):601–12.
82. Jenkins SG, Schuetz AN. Current concepts in laboratory testing to guide antimicrobial therapy. *Mayo Clin Proc.* 2012;87(3):290–308.
83. Koser CU, Ellington MJ, Peacock SJ. Whole-genome sequencing to control antimicrobial resistance. *Trends Genet.* 2014;30(9):401–7.
84. Zankari E, Hasman H, Cosentino S, et al. Identification of acquired antimicrobial resistance genes. *J Antimicrob Chemother.* 2012;67(11):2640–4.
85. Gupta SK, Padmanabhan BR, Diene SM, et al. ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. *Antimicrob Agents Chemother.* 2014;58(1):212–20.
86. Warner DF, Mizrahi V. Complex genetics of drug resistance in *Mycobacterium tuberculosis*. *Nat Genet.* 2013;45(10):1107–8.
87. Heysell SK, Houpt ER. The future of molecular diagnostics for drug-resistant tuberculosis. *Expert Rev Mol Diagn.* 2012;12(4):395–405.
88. MacCannell D. Bacterial strain typing. *Clin Lab Med.* 2013;33(3):629–50.
89. Rehm HL, Bale SJ, Bayrak-Toydemir P, et al. ACMG clinical laboratory standards for next-generation sequencing. *Genet Med.* 2013;15(9):733–47.
90. College of American Pathologists. Molecular pathology checklist. Next generation sequencing. Northfield: College of American Pathologists; 2014. p. 8.
91. Schlager R, Chiu CY, Miller S, et al. Validation of metagenomic next-generation sequencing tests for universal pathogen detection. *Arch Pathol Lab Med.* 2017;141(6):776–86.

92. Food and Drug Administration. High-throughput sequencing technologies for microbial identification and detection of antimicrobial resistance markers. 2014.
93. Doan T, Acharya NR, Pinsky BA, et al. Metagenomic DNA sequencing for the diagnosis of intraocular infections. *Ophthalmology*. 2017;124(8):1247–8.
94. Graf EH, Simmon KE, Tardif KD, et al. 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(4):1000–7.
95. Lefterova MI, Suarez CJ, Banaei N, Pinsky BA. Next-generation sequencing for infectious disease diagnosis and management: a report of the Association for Molecular Pathology. *J Mol Diagn*. 2015;17(6):623–34.
96. Feng H, Shuda M, Chang Y, Moore PS. Clonal integration of a polyomavirus in human Merkel cell carcinoma. *Science*. 2008;319(5866):1096–100.
97. Palacios G, Druce J, Du L, et al. A new arenavirus in a cluster of fatal transplant-associated diseases. *N Engl J Med*. 2008;358(10):991–8.
98. Briese T, Paweska JT, McMullan LK, et al. Genetic detection and characterization of Lujo virus, a new hemorrhagic fever-associated arenavirus from southern Africa. *PLoS Pathog*. 2009;5(5):e1000455.
99. Yozwiak NL, Skewes-Cox P, Gordon A, et al. Human enterovirus 109: a novel interspecies recombinant enterovirus isolated from a case of acute pediatric respiratory illness in Nicaragua. *J Virol*. 2010;84(18):9047–58.
100. Yongfeng H, Fan Y, Jie D, et al. Direct pathogen detection from swab samples using a new high-throughput sequencing technology. *Clin Microbiol Infect*. 2011;17(2):241–4.
101. McMullan LK, Frace M, Sammons SA, et al. Using next generation sequencing to identify yellow fever virus in Uganda. *Virology*. 2012;422(1):1–5.
102. Wamala JF, Malimbo M, Okot CL, et al. Epidemiological and laboratory characterization of a yellow fever outbreak in northern Uganda, October 2010–January 2011. *Int J Infect Dis*. 2012;16(7):e536–42.
103. Wilson MR, Zimmermann LL, Crawford ED, et al. Acute West Nile virus meningoencephalitis diagnosed via metagenomic deep sequencing of cerebrospinal fluid in a renal transplant patient. *Am J Transplant*. 2017;17(3):803–8.
104. Wilson MR, Suan D, Duggins A, et al. A novel cause of chronic viral meningoencephalitis: Cache Valley virus. *Ann Neurol*. 2017;82(1):105–14.
105. Kuroda M, Sekizuka T, Shinya F, et al. Detection of a possible bioterrorism agent, *Francisella* sp., in a clinical specimen by use of next-generation direct DNA sequencing. *J Clin Microbiol*. 2012;50(5):1810–2.
106. Mongkolrattanothai K, Naccache SN, Bender JM, et al. Neurobrucellosis: unexpected answer from metagenomic next-generation sequencing. *J Pediatric Infect Dis Soc*. 2017;6(4):393–8.
107. Wilson MR, Shanbhag NM, Reid MJ, et al. Diagnosing Balamuthia mandrillaris encephalitis with metagenomic deep sequencing. *Ann Neurol*. 2015;78(5):722–30.
108. Frey KG, Herrera-Galeano JE, Redden CL, et al. Comparison of three next-generation sequencing platforms for metagenomic sequencing and identification of pathogens in blood. *BMC Genomics*. 2014;15:96.