CrossMark

SPECIAL ARTICLE

Next-Generation Sequencing for Infectious Disease Diagnosis and Management

A Report of the Association for Molecular Pathology

Martina I. Lefterova,*[†] Carlos J. Suarez,*[†] Niaz Banaei,*[†‡] and Benjamin A. Pinsky*[†‡]

From the Association for Molecular Pathology Next-Generation Sequencing in Infectious Disease Work Group, Bethesda, Maryland; and the Department of Pathology[†] and the Department of Medicine,[‡] Division of Infectious Diseases and Geographic Medicine, Stanford University School of Medicine, Stanford, California*

CME Accreditation Statement: This activity ("JMD 2015 CME Program in Molecular Diagnostics") has been planned and implemented in accordance with the Essential Areas and policies of the Accreditation Council for Continuing Medical Education (ACCME) through the joint providership of the American Society for Clinical Pathology (ASCP) and the American Society for Investigative Pathology (ASIP). ASCP is accredited by the ACCME to provide continuing medical education for physicians.

The ASCP designates this journal-based CME activity ("JMD 2015 CME Program in Molecular Diagnostics") for a maximum of 36 *AMA PRA Category 1 Credit(s)*™. Physicians should only claim credit commensurate with the extent of their participation in the activity.

CME Disclosures: The authors of this article and the planning committee members and staff have no relevant financial relationships with commercial interests to disclose.

Next-generation sequencing (NGS) technologies are increasingly being used for diagnosis and monitoring of infectious diseases. Herein, we review the application of NGS in clinical microbiology, focusing on genotypic resistance testing, direct detection of unknown disease-associated pathogens in clinical specimens, investigation of microbial population diversity in the human host, and strain typing. We have organized the review into three main sections: i) applications in clinical virology, ii) applications in clinical bacteriology, mycobacteriology, and mycology, and iii) validation, quality control, and maintenance of proficiency. Although NGS holds enormous promise for clinical infectious disease testing, many challenges remain, including automation, standardizing technical protocols and bioinformatics pipelines, improving reference databases, establishing proficiency testing and quality control measures, and reducing cost and turnaround time, all of which would be necessary for widespread adoption of NGS in clinical microbiology laboratories. *(J Mol Diagn 2015, 17: 623−634; http://dx.doi.org/10.1016/j.jmoldx.2015.07.004)*

Next-generation sequencing (NGS) methods, also referred to as deep, high-throughput, or massively parallel sequencing, comprise several sequencing technologies that have succeeded the traditional dideoxynucleoside chain termination (ie, Sanger) method. Various platforms, which differ in their sequencing chemistries, read lengths, and throughput capabilities, are available (Table 1).[1] Because these platforms have become more accessible, they have become particularly

Table 1 Characteristics of Current NGS Platforms

Platform	Manufacturer	Sequencing chemistry	Read length, bp	Throughput	Run time
MiSeq/MiSeqDx	Illumina (San Diego, CA)	Sequencing by synthesis-reversible terminator	Up to 2 × 300	13.2—15 Gb	Approximately 56 hours
NextSeq 500	Illumina	Sequencing by synthesis-reversible terminator	Mid output, 2 × 150; high output, 2 × 150	32.5—39 Gb; 100—120 Gb	26 Hours; approximately 29 hours
HiSeq 2500	Illumina	Sequencing by synthesis-reversible terminator	Rapid mode, 2 × 250; high output, 2 × 125	250—300 Gb; 900—1000 Gb	60 Hours; 6 days
HiSeq 3000	Illumina	Sequencing by synthesis-reversible terminator	2 × 150	650—750 Gb	<1—3.5 Days
HiSeq 4000	Illumina	Sequencing by synthesis-reversible terminator	2 × 150	1300—1500 Gb	<1—3.5 Days
HiSeq X	Illumina	Sequencing by synthesis-reversible terminator	2 × 150	1.6—1.8 Tb	<3 Days
Ion PGM	Life Technologies (Carlsbad, CA)*	Sequencing by synthesis—hydrogen ion detection	200—400	30 Mb—2 Gb	2—7 Hours
Ion Proton	Life Technologies*	Sequencing by synthesis—hydrogen ion detection	200	10 Gb	2—4 Hours
SOLiD 5500xl W System	Life Technologies*	Fluorescent probe ligation	Up to 50 × 50 paired end	Approximately 320 Gb	Approximately 5 days
PacBio RS II	Pacific Biosciences (Menlo Park, CA)	Single-molecule real-time sequencing	4200—8500	275—375 Mb	30—180 Minutes
454 GS Junior	Roche (Basel, Switzerland)	Pyrosequencing	600—700	35—53 Mb	10 Hours
454 GS FLX Titanium XL+	Roche	Pyrosequencing	Approximately 700	Approximately 600 Mb	23 Hours

The specifications of the instruments were obtained from the manufacturer's web sites and/or company's representatives, and include up-to-date information as of May 2015.

*Thermo Fisher Scientific.

NGS, next-generation sequencing.

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 amplicon sequencing or whole-genome sequencing (WGS) (Figure 1).[2,3] The first approach 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 (eg, identify known drug-resistant mutants). Sequencing for *de novo* assembly of whole genomes, on the other hand, relies on nontargeted library preparation. This is often performed when microorganisms are unknown or the goal is to define the genomic content and functional potential of the organism under investigation. WGS may also be applied to primary specimens for culture-independent pathogen identification or characterization of the microbial population.

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 is currently infrequent, although its

role in the diagnosis and management of infectious diseases is anticipated 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, the Sanger method has limited sensitivity for minor variants when present at <15% to 20% of the viral population, whereas NGS methods can detect drug-resistant mutations (DRMs) present at approximately 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
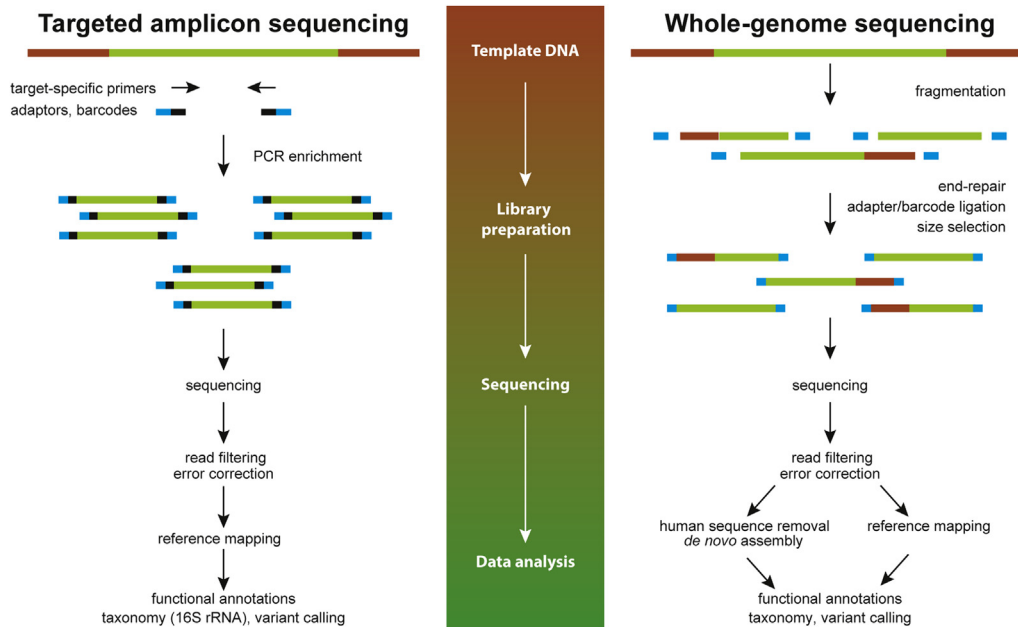
Figure 1   Illustration of sequencing approaches for diagnosis and monitoring of infectious diseases. Targeted amplicon sequencing uses target-specific primers for template (**green bars**) enrichment, followed by primers that are partially complementary to the target-specific primers (**black bars**) and contain sequencing adaptors and bar codes (**blue bars**). Whole-genome sequencing uses 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 because the nucleic acids of the organism of interest frequently constitute <1% of the nucleic acid pool. Fragmentation libraries may also be made from PCR-enriched amplicons.

viral drug resistance. However, genotypic drug resistance testing is also used for the clinical management of other viral infections, including cytomegalovirus (CMV), hepatitis B, hepatitis C, and influenza, and it is conceivable that NGS methods will be applicable for these viruses as well.

HIV-1

Epidemiological studies in HIV-1−positive patients have shown that the presence of mutations conferring resistance to highly active antiretroviral therapy 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] Several studies have compared NGS with 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 because of library preparation and/or NGS errors.[14] Alternatively, the library preparation step can use 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] 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.[16,17] 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]

Many HIV-1 research studies demonstrating the superior performance of NGS methods compared with Sanger sequencing[4] have resulted in the introduction of several clinical NGS HIV-1 drug resistance assays. The most comprehensive is the DEEPGEN 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 coreceptor tropism, with mean error rates

of 0.37% to 0.39%, sensitivity for minor variants of 5%, and capacity to multiplex up to 96 samples in a single run.[5] Two other assays that use deep sequencing of the HIV-1 *env* V3 loop for HIV coreceptor tropism are available clinically: the HIV-1 CCR5 Tropism Test (V3) offered by the British Columbia Center for Excellence in HIV/AIDS (Vancouver, BC, Canada)[18] and the HIV-1 Co-receptor Tropism with Reflex to Ultradeep Sequencing offered by Quest Diagnostics (Madison, NJ).[19] These assays and DEEPGEN HIV have been shown to predict non-CCR5 tropism as accurately as the phenotypic gold standard (Trofile; Monogram Biosciences, South San Francisco, CA) and to exhibit a higher sensitivity than Sanger sequencing for detecting minor CXCR4-tropic variants.

More important, 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.[20,21] 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 by both NGS and Sanger sequencing.[8]

CMV

CMV is another virus for which genotypic drug resistance testing is clinically useful, particularly in transplant recipients.[22,23] Rates of CMV drug resistance vary on the basis of patient populations: 5% to 12.5% in solid organ transplant recipients and 2% to 5% in hematopoietic stem cell transplant recipients.[24] Timely detection of CMV drug resistance is critical because DRMs can accumulate with continued exposure to a drug,[24,25] potentially leading to shortened graft survival and increased morbidity.[26,27] Furthermore, rational change of therapy after identification of drug resistance has been shown to lead to more rapid clearance of virus.[28] Resistance mutations to current CMV drugs 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 has recently been described, demonstrating a low overall empirical error rate (0.189%) and reliable detection of CMV DRMs in clinical plasma specimens with a wide range of viral loads.[17]

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.[25] NGS studies of viral drug resistance are also expected to identify novel putative DRMs, which, after appropriate phenotypic validation,[25] can be incorporated into CMV DRM databases

and genotypic interpretation systems, similar to those that exist for HIV-1.[29] Such automated tools have been shown to improve sequence analysis in addition to expediting and standardizing workflow when compared with manual sequence curation.[30]

Virus Identification in Clinical Specimens

Proof-of-concept studies have demonstrated the ability of WGS to identify common, clinically relevant viruses from a variety of specimen types previously shown to be positive by routine molecular testing.[31] 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.[32] Many viruses cannot be cultured or identified by traditional molecular techniques, whereas 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.[32] In contrast, NGS offers an efficient, highly sensitive, and unbiased alternative for the detection of viruses in clinical specimens.[2,32] 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.

Because 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 to remove host nucleic acids, while preserving capsid-protected viral particles.[2,33] Nucleic acid amplification methods for enrichment of viral genomes include rolling circle amplification for viruses with a circular genome[34]; use of restriction enzyme sites that are more frequently encountered in viral nucleic acids than human, followed by ligation of adaptors and PCR amplification[35,36]; and use of oligonucleotides that specifically target host nucleic acids (eg, rRNA) and interfere with their reverse transcription and amplification.[35] Newer 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.[37,38] Computational tools have been

Table 2 Select Studies Describing Culture-Independent NGS Pathogen Identification from Primary Human Clinical Specimens

Pathogen name	Clinical information	Specimen(s)	Method	Pathogen-aligned reads*	Reference(s)
Viruses					
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)	42
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)	43
Lujo virus	Hemorrhagic fever outbreak in southern Africa	Total RNA from liver tissue and serum	454	Not provided[†]	44
Enterovirus 109	Influenza-like illness in a child	Total RNA from nasopharyngeal swab	Illumina	0.00001 (119/10,400,000)	45
Influenza A	Upper respiratory tract illness	Total RNA from nasopharyngeal swabs	Illumina	0.001-0.0001%[‡]	46
Yellow fever virus	Fatal hemorrhagic fever in a 16-year-old male	Total RNA from serum	454	0.5 (3229/599,158)	47, 48
Bacteria					
Francisella tularensis	Abscess	Total DNA from abscess drainage	Illumina	0.002 (833/38,285,502)	49
Leptospira santarosai	Recurrent meningoencephalitis in a child with SCID	Total DNA from CSF	Illumina	0.016 (475/3,063,784)	50

*Data are given as percentage (number/total).
[†]Obtained coverage of 5.6 kb of a 10.4-kb genome.
[‡]The number of influenza A—specific and total reads varied between samples.
CSF, cerebrospinal fluid; LCMV, lymphocytic choriomeningitis virus; NGS, next-generation sequencing; SCID, severe combined immune deficiency.

developed for subtracting host sequences from the initial read pool containing mixed human and microbial sequences.[39–41] This filtering step is crucial because viral sequences may comprise <1% of the initial aligned reads[32,41] (Table 2).

In addition, 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.[51] The choice of sequencing platform (Table 1) also requires consideration, because read length and sequence depth may affect virus detection and genome assembly.[52]

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 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 by identifying overlapping sequences between reads, followed by

contiguous sequence assembly into genomes.[53] Sequencing methods that produce long reads and, therefore, linkage information facilitate contiguous sequence assembly (Table 1). Sequences assembled this way can be compared with public databases by using algorithms with relaxed stringency to identify related viruses. Repetitive sequences pose a significant challenge in *de novo* assembly, because they can interfere with PCR amplification and accurate genomic mapping. Computational and experimental strategies are being developed to address such issues.[2,53] Table 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 etiological agent is found in affected hosts but not healthy controls, it is propagated in culture, and it 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.[54] 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 serum specimens.

Specific Applications in Clinical Bacteriology, Mycobacteriology, and Mycology

Identification by Targeted Sequencing or WGS

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 to identify the genus or species of the infecting organism.[55] 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, by both targeted amplicon sequencing of rRNA genes and WGS.[3]

rRNA Sequencing

rRNA gene sequencing by Sanger is routinely used for bacterial and fungal identification in clinical microbiology laboratories.[56,57] Furthermore, rRNA sequencing by NGS is the basis for studies of the microbiome. For bacteria, these approaches use 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.[58] In addition, 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.[59]

Informatics is critical for the interpretation of 16S rRNA sequencing data that are obtained by NGS. The length, number, and quality of sequencing reads, as well as possible bacterial contamination of reagents, are all factors that can affect pathogen identification and introduce bias in microbiome diversity assessments.[60] 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 biopsy specimens.[61] Data analysis tools include error correction methods[59,62] and removal of amplification-derived chimeric sequences.[63,64] Processed data or raw sequences can be analyzed in dedicated pipelines, such as QIIME,[65] the Ribosomal Database Project,[66] and mothur,[63] which cluster similar sequences into operational taxonomic units on the basis of 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 on the basis of the sequencing platform and the quality of sequencing data.[67] More important, the accuracy of bacterial identification largely depends on the scope and completeness of reference databases used for analysis. Several extensive databases have been generated for 16S rRNA sequences, including the following: SILVA (http://www.arb-silva.de/documentation, last accessed August 4, 2015), containing >3 million small subunit and >250,000 large subunit bacterial rRNA gene sequences,[68] or Greengenes (http://www.greengenes.secondgenome.com/downloads, last accessed May 27, 2015), which can calculate taxonomic relationships on the basis of >400,000 16S rRNA sequences.[69]

Targeted 16S rRNA sequencing by NGS may have immediate clinical application for the characterization of mixed infections, particularly those containing uncultivatable or nonviable organisms. This approach has been successfully applied directly to brain abscess material, lymph node biopsy tissue, cystic fibrosis sputa, and mastoid abscess material.[70,71]

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 with the healthy state.[72] On the other hand, studies of microbiome in the lower airway of cystic fibrosis patients[73] and stool of patients with *Clostridium difficile* infections[74] or inflammatory bowel disease[75] have shown that disease progression is marked by decreasing bacterial diversity, which may be related to escalating antibiotic exposures. In particular, a recent microbiome study of pediatric Crohn's disease has shown a distinct pattern of gut microbial dysbiosis, consisting of increased abundance of *Enterobacteriaceae*, *Pasteurellaceae*, *Veillonellaceae*, and *Fusobacteriaceae* and decreased abundance of *Erysipelotrichales*, *Bacteroidales*, and *Clostridiales*. This pattern appears to correlate with disease severity and to be amplified by antibiotic use.[76] 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. In fact, studies in patients with recurrent *Clostridium difficile* infections have shown that administration of fecal microbiota from healthy donors is associated with restoration of intestinal bacterial diversity and repopulation by key phyla, such as *Bacteroides* and *Firmicutes*, and leads to resolution or reduction of symptoms in most recipients.[77] 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.

WGS

In contrast to rRNA-based NGS approaches, WGS approaches allow more detailed functional and taxonomic analyses, when either a cultured isolate is tested or an entire microbial community is being characterized, a field termed metagenomics.[55,59] 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[78] detected *Leptospira santarosai* sequences in cerebrospinal fluid within 48 hours of specimen receipt.[50] Table 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 WGS data poses even more challenges than those discussed for 16S rRNA sequencing.[59] 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 contiguous sequences for gene predictions to be made and biological functions to be assigned.[59,78] One of the most widely used tools for this purpose is the metagenomics RAST server (*http://metagenomics.anl.gov*, last accessed May 27, 2015).[79] An example of another tool that has been used for clinical pathogen identification is the SURPI pipeline (sequence-based ultrarapid pathogen identification), which can analyze $>10 \times 10^6$ reads in approximately 100 minutes,[50] and can be used on both cloud-based and standalone servers.[78] 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[80] and Metagenomics of the Human Intestinal Tract project,[81] are actively expanding bacterial genomic databases. A combination of these and other similar efforts have led to the recently reported integrated gene catalog of the human gut microbiome comprising 9,879,896 genes, derived from 1267 gut metagenomes of 1070 individuals, from three continents.[82]

Several groups have assessed the feasibility of WGS for bacterial identification and characterization in a clinical microbiology laboratory. One study tested the feasibility of WGS for routine use by sequencing 130 cultured isolates, including aerobic and anaerobic bacteria, mycobacteria, and fungi.[83] The steps from colony harvest to acquisition of analyzable data took approximately 55 hours, with most of the time attributable to the sequencing run (39 hours). Comparison of the WGS results with identification by matrix-assisted laser desorption ionization time-of-flight mass spectrometry, in addition to conventional culture and biochemical methods, demonstrated good correlation: 115 (88.5%) of 130 samples showed concordant results, whereas 15 of 130 could not be identified because of insufficient coverage or absence of applicable reference genomes in publicly available databases (mainly for sterile molds). Thus, WGS was able to identify most 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.[55,84] 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*.[85] 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.[85] WGS-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 detection of toxin genes.[86] Use of this approach is likely to expand because new drug resistance mechanisms are characterized and cataloged in publicly available databases, such as ResFinder[87] and ARG-ANNOT,[88] which use BLAST to query a user-supplied sequence against a curated list of bacterial antimicrobial resistance genes. Several proof-of-concept studies have assessed the ability of NGS to predict bacterial drug resistance patterns and have been reviewed by Koser et al.[86] Although 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.[85] 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, WGS 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 WGS-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.[89,90]

Table 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.[92]
Analytical sensitivity	Microbial variant detection: mixes of known variant strains and wild-type strains at different percentages and at low, medium, and high levels (eg, 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, next-generation sequencing.

Another area where bacterial WGS is being implemented for clinical purposes is for strain typing in hospital outbreak investigations. Traditionally, strain typing has been performed by either fragment analysis methods (eg, pulsed-field gel electrophoresis) or sequence-based techniques (eg, multilocus sequence typing).[91] 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 on the basis of bacterial WGS 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.[55] In this approach, WGS data are aligned to a reference genome, single-nucleotide polymorphisms are identified and filtered on the basis of pre-established 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 recently in several small studies.[55] However, it remains to be shown whether the use of WGS 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 3). In that regard, NGS performed in the clinical laboratory for patient care differs from NGS performed in the research setting, even though the sequencing methods may be the same. As such, the American College for Medical Genetics and Genomics has published detailed clinical laboratory standards for NGS.[93] Furthermore, the College of American Pathologists has developed an NGS checklist for accreditation of molecular pathology laboratories performing clinical NGS testing.[94] 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. Although not specifically developed for infectious diseases applications, this checklist provides the framework for the microbiology NGS checklist currently in development.

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, because 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. 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, because no clinical microbiology NGS tests have yet been approved by the US 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 article detailing clinical applications and validation approaches for the regulatory approval/clearance of NGS diagnostic devices for clinical microbiology.[95] This document reports that the FDA is engaged in the development of a database (FDA MicroDB) composed 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 disease NGS *in vitro* diagnostics will likely aid in the standardization of specimen handling, library preparation and sequencing, and data interpretation, 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.[61]

Conclusions

Herein, we have reviewed areas of clinical microbiology in which NGS approaches have been used to identify and characterize medically important pathogens. Although 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. 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. For example, GnuBIO, a BioRad company, has developed an integrated, bench-top sequencer that performs automated target enrichment, fluorescence resonance energy transfer—based digital sequencing, and sequence analysis from extracted nucleic acids in 3 hours.[96] 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 rRNA genes may also become more commonly used for identification of pathogenic bacteria and fungi when there is high suggestion of infection and culture is negative or not available, or when mixed infections are suspected. WGS 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. More important, 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.

The greatest attraction of genomic approaches is that WGS 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.[55]

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

Disclaimer

The AMP Clinical Practice Guidelines and Reports are developed to be of assistance to laboratory and other health care professionals by providing guidance and recommendations for particular areas of practice. The Guidelines or Reports should not be considered inclusive of all proper approaches or methods, or exclusive of others. The Guidelines or Reports cannot guarantee any specific outcome, nor do they establish a standard of care. The Guidelines or Reports are not intended to dictate the treatment of a particular patient. Treatment decisions must be made based on the basis of the independent judgment of health care providers and each patient's individual circumstances. AMP makes no warranty, expressed or implied, regarding the Guidelines or Reports and specifically excludes any warranties of merchantability and fitness for a particular use or purpose. AMP shall not be liable for direct, indirect, special, incidental, or consequential damages related to the use of the information contained herein.

References

1. Loman NJ, Constantinidou C, Chan JZ, Halachev M, Sergeant M, Penn CW, Robinson ER, Pallen MJ: High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. Nat Rev Microbiol 2012, 10:599—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):1853−1868

3. Weinstock GM: Genomic approaches to studying the human microbiota. Nature 2012, 489:250−256

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, 12:89−102

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, 16:401

6. Dunn DT, Coughlin K, Cane PA: Genotypic resistance testing in routine clinical care. Curr Opin HIV AIDS 2011, 6:251−257

7. Cortez KJ, Maldarelli F: Clinical management of HIV drug resistance. Viruses 2011, 3:347−378

8. Simen BB, Simons JF, Hullsiek KH, Novak RM, Macarthur RD, Baxter JD, Huang C, Lubeski C, Turenchalk GS, Braverman MS, Desany B, Rothberg JM, Egholm M, Kozal MJ; Terry Beirn Community Programs for Clinical Research on AIDS: Low-abundance drug-resistant viral variants in chronically HIV-infected, antiretroviral treatment-naive patients significantly impact treatment outcomes. J Infect Dis 2009, 199:693−701

9. Lataillade M, Chiarella J, Yang R, Schnittman S, Wirtz V, Uy J, Seekins D, Krystal M, Mancini M, McGrath D, Simen B, Egholm M, Kozal M: Prevalence and clinical significance of HIV drug resistance mutations by ultra-deep sequencing in antiretroviral-naive subjects in the CASTLE study. PLoS One 2010, 5:e10952

10. Li JZ, Paredes R, Ribaudo HJ, Svarovskaia ES, Metzner KJ, Kozal MJ, Hullsiek KH, Balduin M, Jakobsen MR, Geretti AM, Thiebaut R, Ostergaard L, Masquelier B, Johnson JA, Miller MD, Kuritzkes DR: Low-frequency HIV-1 drug resistance mutations and risk of NNRTI-based antiretroviral treatment failure: a systematic review and pooled analysis. JAMA 2011, 305:1327−1335

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: 248−252

12. Harismendy O, Ng PC, Strausberg RL, Wang X, Stockwell TB, Beeson KY, Schork NJ, Murray SS, Topol EJ, Levy S, Frazer KA: Evaluation of next generation sequencing platforms for population targeted sequencing studies. Genome Biol 2009, 10: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: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, 17:1195−1201

15. Jabara CB, Jones CD, Roach J, Anderson JA, Swanstrom R: Accurate sampling and deep sequencing of the HIV-1 protease gene using a Primer ID. Proc Natl Acad Sci U S A 2011, 108:20166−20171

16. Vandenbroucke I, Van Marck H, Mostmans W, Van Eygen V, Rondelez E, Thys K, Van Baelen K, Fransen K, Vaira D, Kabeya K, De Wit S, Florence E, Moutschen M, Vandekerckhove L, Verhofstede C, Stuyver LJ: HIV-1 V3 envelope deep sequencing for clinical plasma specimens failing in phenotypic tropism assays. AIDS Res Ther 2010, 7:4

17. Sahoo MK, Lefterova MI, Yamamoto F, Waggoner JJ, Chou S, Holmes SP, Anderson MW, Pinsky BA: Detection of cytomegalovirus drug resistance mutations by next-generation sequencing. J Clin Microbiol 2013, 51:3700−3710

18. Swenson LC, Mo T, Dong WW, Zhong X, Woods CK, Jensen MA, Thielen A, Chapman D, Lewis M, James I, Heera J, Valdez H, Harrigan PR: Deep sequencing to infer HIV-1 co-receptor usage: application to three clinical trials of maraviroc in treatment-experienced patients. J Infect Dis 2011, 203:237−245

19. Kagan RM, Johnson EP, Siaw M, Biswas P, Chapman DS, Su Z, Platt JL, Pesano RL: A genotypic test for HIV-1 tropism combining Sanger sequencing with ultradeep sequencing predicts virologic response in treatment-experienced patients. PLoS One 2012, 7:e46334

20. Le T, Chiarella J, Simen BB, Hanczaruk B, Egholm M, Landry ML, Dieckhaus K, Rosen MI, Kozal MJ: Low-abundance HIV drug-resistant viral variants in treatment-experienced persons correlate with historical antiretroviral use. PLoS One 2009, 4:e6079

21. Codoner FM, Pou C, Thielen A, Garcia F, Delgado R, Dalmau D, Alvarez-Tejado M, Ruiz L, Clotet B, Paredes R: Added value of deep sequencing relative to population sequencing in heavily pre-treated HIV-1-infected subjects. PLoS One 2011, 6:e19461

22. Kotton CN, Kumar D, Caliendo AM, Asberg A, Chou S, Snydman DR, Allen U, Humar A: Transplantation Society International CMV Consensus Group. International consensus guidelines on the management of cytomegalovirus in solid organ transplantation. Transplantation 2010, 89:779−795

23. Ljungman P, Reusser P, de la Camara R, Einsele H, Engelhard D, Ribaud P, Ward K; European Group for Blood and Marrow Transplantation: Management of CMV infections: recommendations from the infectious diseases working party of the EBMT. Bone Marrow Transplant 2004, 33:1075−1081

24. Lurain NS, Chou S: Antiviral drug resistance of human cytomegalovirus. Clin Microbiol Rev 2010, 23:689−712

25. Chou S, Ercolani RJ, Sahoo MK, Lefterova MI, Strasfeld LM, Pinsky BA: Improved detection of emerging drug-resistant mutant cytomegalovirus subpopulations by deep sequencing. Antimicrobial Agents Chemother 2014, 58:4697−4702

26. 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: 439−447

27. Limaye AP, Corey L, Koelle DM, Davis CL, Boeckh M: Emergence of ganciclovir-resistant cytomegalovirus disease among recipients of solid-organ transplants. Lancet 2000, 356:645−649

28. Myhre HA, Haug Dorenberg D, Kristiansen KI, Rollag H, Leivestad T, Asberg A, Hartmann A: Incidence and outcomes of ganciclovir-resistant cytomegalovirus infections in 1244 kidney transplant recipients. Transplantation 2011, 92:217−223

29. Tang MW, Liu TF, Shafer RW: The HIVdb system for HIV-1 genotypic resistance interpretation. Intervirology 2012, 55:98−101

30. Woods CK, Brumme CJ, Liu TF, Chui CK, Chu AL, Wynhoven B, Hall TA, Trevino C, Shafer RW, Harrigan PR: Automating HIV drug resistance genotyping with RECall, a freely accessible sequence analysis tool. J Clin Microbiol 2012, 50:1936−1942

31. Petty TJ, Cordey S, Padioleau I, Docquier M, Turin L, Preynat-Seauve O, Zdobnov EM, Kaiser L: Comprehensive human virus screening using high-throughput sequencing with a user-friendly representation of bioinformatics analysis: a pilot study. J Clin Microbiol 2014, 52:3351−3361

32. Chiu CY: Viral pathogen discovery. Curr Opin Microbiol 2013, 16: 468−478

33. Barzon L, Lavezzo E, Militello V, Toppo S, Palu G: Applications of next-generation sequencing technologies to diagnostic virology. Int J Mol Sci 2011, 12:7861−7884

34. 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, 78:4993−4998

35. de Vries M, Deijs M, Canuti M, van Schaik BD, Faria NR, van de Garde MD, Jachimowski LC, Jebbink MF, Jakobs M, Luyf AC, Coenjaerts FE, Claas EC, Molenkamp R, Koekkoek SM, Lammens C, Leus F, Goossens H, Ieven M, Baas F, van der Hoek L: A sensitive assay for virus discovery in respiratory clinical samples. PLoS One 2011, 6:e16118

36. 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

37. Depledge DP, Palser AL, Watson SJ, Lai IY, Gray ER, Grant P, Kanda RK, Leprout E, Kellam P, Breuer J: Specific capture and

whole-genome sequencing of viruses from clinical samples. PLoS One 2011, 6:e27805

38. Duncavage EJ, Magrini V, Becker N, Armstrong JR, Demeter RT, Wylie T, Abel HJ, Pfeifer JD: Hybrid capture and next-generation sequencing identify viral integration sites from formalin-fixed, paraffin-embedded tissue. J Mol Diagn 2011, 13:325−333

39. Bhaduri A, Qu K, Lee CS, Ungewickell A, Khavari PA: Rapid identification of non-human sequences in high-throughput sequencing datasets. Bioinformatics 2012, 28:1174−1175

40. Kostic AD, Ojesina AI, Pedamallu CS, Jung J, Verhaak RG, Getz G, Meyerson M: PathSeq: software to identify or discover microbes by deep sequencing of human tissue. Nat Biotechnol 2011, 29:393−396

41. 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:e1485

42. Feng H, Shuda M, Chang Y, Moore PS: Clonal integration of a polyomavirus in human Merkel cell carcinoma. Science 2008, 319:1096−1100

43. Palacios G, Druce J, Du L, Tran T, Birch C, Briese T, Conlan S, Quan PL, Hui J, Marshall J, Simons JF, Egholm M, Paddock CD, Shieh WJ, Goldsmith CS, Zaki SR, Catton M, Lipkin WI: A new arenavirus in a cluster of fatal transplant-associated diseases. N Engl J Med 2008, 358:991−998

44. Briese T, Paweska JT, McMullan LK, Hutchison SK, Street C, Palacios G, Khristova ML, Weyer J, Swanepoel R, Egholm M, Nichol ST, Lipkin WI: Genetic detection and characterization of Lujo virus, a new hemorrhagic fever-associated arenavirus from southern Africa. PLoS Pathog 2009, 5:e1000455

45. Yozwiak NL, Skewes-Cox P, Gordon A, Saborio S, Kuan G, Balmaseda A, Ganem D, Harris E, DeRisi JL: Human enterovirus 109: a novel interspecies recombinant enterovirus isolated from a case of acute pediatric respiratory illness in Nicaragua. J Virol 2010, 84:9047−9058

46. Yongfeng H, Fan Y, Jie D, Jian Y, Ting Z, Lilian S, Jin Q: Direct pathogen detection from swab samples using a new high-throughput sequencing technology. Clin Microbiol Infect 2011, 17:241−244

47. McMullan LK, Frace M, Sammons SA, Shoemaker T, Balinandi S, Wamala JF, Lutwama JJ, Downing RG, Stroeher U, MacNeil A, Nichol ST: Using next generation sequencing to identify yellow fever virus in Uganda. Virology 2012, 422:1−5

48. Wamala JF, Malimbo M, Okot CL, Atai-Omoruto AD, Tenywa E, Miller JR, Balinandi S, Shoemaker T, Oyoo C, Omony EO, Kagirita A, Musenero MM, Makumbi I, Nanyunja M, Lutwama JJ, Downing R, Mbonye AK: Epidemiological and laboratory characterization of a yellow fever outbreak in northern Uganda, October 2010-January 2011. Int J Infect Dis 2012, 16:e536−e542

49. Kuroda M, Sekizuka T, Shinya F, Takeuchi F, Kanno T, Sata T, Asano S: Detection of a possible bioterrorism agent, Francisella sp., in a clinical specimen by use of next-generation direct DNA sequencing. J Clin Microbiol 2012, 50:1810−1812

50. Wilson MR, Naccache SN, Samayoa E, Biagtan M, Bashir H, Yu G, Salamat SM, Somasekar S, Federman S, Miller S, Sokolic R, Garabedian E, Candotti F, Buckley RH, Reed KD, Meyer TL, Seroogy CM, Galloway R, Henderson SL, Gern JE, DeRisi JL, Chiu CY: Actionable diagnosis of neuroleptospirosis by next-generation sequencing. N Engl J Med 2014, 370:2408−2417

51. Berthet N, Reinhardt AK, Leclercq I, van Ooyen S, Batejat C, Dickinson P, Stamboliyska R, Old IG, Kong KA, Dacheux L, Bourhy H, Kennedy GC, Korfhage C, Cole ST, Manuguerra JC: Phi29 polymerase based random amplification of viral RNA as an alternative to random RT-PCR. BMC Mol Biol 2008, 9:77

52. Cheval J, Sauvage V, Frangeul L, Dacheux L, Guigon G, Dumey N, Pariente K, Rousseaux C, Dorange F, Berthet N, Brisse S, Moszer I, Bourhy H, Manuguerra CJ, Lecuit M, Burguiere A, Caro V, Eloit M: Evaluation of high-throughput sequencing for identifying known and unknown viruses in biological samples. J Clin Microbiol 2011, 49:3268−3275

53. Miller JR, Koren S, Sutton G: Assembly algorithms for next-generation sequencing data. Genomics 2010, 95:315−327

54. Falkow S: Molecular Koch's postulates applied to bacterial pathogenicity: a personal recollection 15 years later. Nat Rev Microbiol 2004, 2:67−72

55. Koser CU, Ellington MJ, Cartwright EJ, Gillespie SH, Brown NM, Farrington M, Holden MT, Dougan G, Bentley SD, Parkhill J, Peacock SJ: Routine use of microbial whole genome sequencing in diagnostic and public health microbiology. PLoS Pathog 2012, 8:e1002824

56. Clarridge JE 3rd: Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. Clin Microbiol Rev 2004, 17:840−862

57. Balajee SA, Sigler L, Brandt ME: DNA and the classical way: identification of medically important molds in the 21st century. Med Mycol 2007, 45:475−490

58. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, Glockner FO: Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Res 2013, 41:e1

59. Kuczynski J, Lauber CL, Walters WA, Parfrey LW, Clemente JC, Gevers D, Knight R: Experimental and analytical tools for studying the human microbiome. Nat Rev Genet 2012, 13:47−58

60. Cox MJ, Cookson WO, Moffatt MF: Sequencing the human microbiome in health and disease. Hum Mol Genet 2013, 22:R88−R94

61. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, Turner P, Parkhill J, Loman NJ, Walker AW: Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biol 2014, 12:87

62. Schloss PD, Gevers D, Westcott SL: Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. PLoS One 2011, 6:e27310

63. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF: Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol 2009, 75:7537−7541

64. Haas BJ, Gevers D, Earl AM, Feldgarden M, Ward DV, Giannoukos G, Ciulla D, Tabbaa D, Highlander SK, Sodergren E, Methe B, DeSantis TZ; Human Microbiome Consortium, Petrosino JF, Knight R, Birren BW: Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. Genome Res 2011, 21:494−504

65. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R: QIIME allows analysis of high-throughput community sequencing data. Nat Methods 2010, 7:335−336

66. Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ, Kulam-Syed-Mohideen AS, McGarrell DM, Marsh T, Garrity GM, Tiedje JM: The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucleic Acids Res 2009, 37(Database issue):D141−D145

67. Bokulich NA, Subramanian S, Faith JJ, Gevers D, Gordon JI, Knight R, Mills DA, Caporaso JG: Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. Nat Methods 2013, 10:57−59

68. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glockner FO: The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 2013, 41(Database issue):D590−D596

69. McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, Andersen GL, Knight R, Hugenholtz P: An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. ISME J 2012, 6:610–618

70. Salipante SJ, Sengupta DJ, Rosenthal C, Costa G, Spangler J, Sims EH, Jacobs MA, Miller SI, Hoogestraat DR, Cookson BT, McCoy C, Matsen FA, Shendure J, Lee CC, Harkins TT, Hoffman NG: Rapid 16S rRNA next-generation sequencing of polymicrobial clinical samples for diagnosis of complex bacterial infections. PLoS One 2013, 8:e65226

71. Salipante SJ, Hoogestraat DR, Abbott AN, SenGupta DJ, Cummings LA, Butler-Wu SM, Stephens K, Cookson BT, Hoffman NG: Coinfection of Fusobacterium nucleatum and Actinomyces israelii in mastoiditis diagnosed by next-generation DNA sequencing. J Clin Microbiol 2014, 52:1789–1792

72. Srinivasan S, Hoffman NG, Morgan MT, Matsen FA, Fiedler TL, Hall RW, Ross FJ, McCoy CO, Bumgarner R, Marrazzo JM, Fredricks DN: Bacterial communities in women with bacterial vaginosis: high resolution phylogenetic analyses reveal relationships of microbiota to clinical criteria. PLoS One 2012, 7:e37818

73. Zhao J, Schloss PD, Kalikin LM, Carmody LA, Foster BK, Petrosino JF, Cavalcoli JD, VanDevanter DR, Murray S, Li JZ, Young VB, LiPuma JJ: Decade-long bacterial community dynamics in cystic fibrosis airways. Proc Natl Acad Sci U S A 2012, 109:5809–5814

74. 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:e280

75. Morgan XC, Tickle TL, Sokol H, Gevers D, Devaney KL, Ward DV, Reyes JA, Shah SA, LeLeiko N, Snapper SB, Bousvaros A, Korzenik J, Sands BE, Xavier RJ, Huttenhower C: Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. Genome Biol 2012, 13:R79

76. Gevers D, Kugathasan S, Denson LA, Vazquez-Baeza Y, Van Treuren W, Ren B, Schwager E, Knights D, Song SJ, Yassour M, Morgan XC, Kostic AD, Luo C, Gonzalez A, McDonald D, Haberman Y, Walters T, Baker S, Rosh J, Stephens M, Heyman M, Markowitz J, Baldassano R, Griffiths A, Sylvester F, Mack D, Kim S, Crandall W, Hyams J, Huttenhower C, Knight R, Xavier RJ: The treatment-naive microbiome in new-onset Crohn's disease. Cell Host Microbe 2014, 15:382–392

77. van Nood E, Vrieze A, Nieuwdorp M, Fuentes S, Zoetendal EG, de Vos WM, Visser CE, Kuijper EJ, Bartelsman JF, Tijssen JG, Speelman P, Dijkgraaf MG, Keller JJ: Duodenal infusion of donor feces for recurrent Clostridium difficile. N Engl J Med 2013, 368:407–415

78. Naccache SN, Federman S, Veeraraghavan N, Zaharia M, Lee D, Samayoa E, Bouquet J, Greninger AL, Luk KC, Enge B, Wadford DA, Messenger SL, Genrich GL, Pellegrino K, Grard G, Leroy E, Schneider BS, Fair JN, Martinez MA, Isa P, Crump JA, DeRisi JL, Sittler T, Hackett J Jr, Miller S, Chiu CY: A cloud-compatible bioinformatics pipeline for ultrarapid pathogen identification from next-generation sequencing of clinical samples. Genome Res 2014, 24:1180–1192

79. Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M, Paczian T, Rodriguez A, Stevens R, Wilke A, Wilkening J, Edwards RA: The metagenomics RAST server: a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics 2008, 9:386

80. Group NHW, Peterson J, Garges S, Giovanni M, McInnes P, Wang L, et al: The NIH Human Microbiome Project. Genome Res 2009, 19:2317–2323

81. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, et al: A human gut microbial gene catalogue established by metagenomic sequencing. Nature 2010, 464:59–65

82. Li J, Jia H, Cai X, Zhong H, Feng Q, Sunagawa S, et al: An integrated catalog of reference genes in the human gut microbiome. Nat Biotechnol 2014, 32:834–841

83. Long SW, Williams D, Valson C, Cantu CC, Cernoch P, Musser JM, Olsen RJ: A genomic day in the life of a clinical microbiology laboratory. J Clin Microbiol 2013, 51:1272–1277

84. Didelot X, Bowden R, Wilson DJ, Peto TE, Crook DW: Transforming clinical microbiology with bacterial genome sequencing. Nat Rev Genet 2012, 13:601–612

85. Jenkins SG, Schuetz AN: Current concepts in laboratory testing to guide antimicrobial therapy. Mayo Clin Proc 2012, 87:290–308

86. Koser CU, Ellington MJ, Peacock SJ: Whole-genome sequencing to control antimicrobial resistance. Trends Genet 2014, 30:401–407

87. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, Aarestrup FM, Larsen MV: Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother 2012, 67:2640–2644

88. Gupta SK, Padmanabhan BR, Diene SM, Lopez-Rojas R, Kempf M, Landraud L, Rolain JM: ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. Antimicrobial Agents Chemother 2014, 58:212–220

89. Warner DF, Mizrahi V: Complex genetics of drug resistance in Mycobacterium tuberculosis. Nat Genet 2013, 45:1107–1108

90. Heysell SK, Houpt ER: The future of molecular diagnostics for drug-resistant tuberculosis. Expert Rev Mol Diagn 2012, 12:395–405

91. MacCannell D: Bacterial strain typing. Clin Lab Med 2013, 33:629–650

92. Frey KG, Herrera-Galeano JE, Redden CL, Luu TV, Servetas SL, Mateczun AJ, Mokashi VP, Bishop-Lilly KA: Comparison of three next-generation sequencing platforms for metagenomic sequencing and identification of pathogens in blood. BMC Genomics 2014, 15:96

93. Rehm HL, Bale SJ, Bayrak-Toydemir P, Berg JS, Brown KK, Deignan JL, Friez MJ, Funke BH, Hegde MR, Lyon E; Working Group of the American College of Medical Genetics and Genomics Laboratory Quality Assurance Committee: ACMG clinical laboratory standards for next-generation sequencing. Genet Med 2013, 15:733–747

94. College of American Pathologists: Molecular Pathology Checklist. Next Generation Sequencing. Northfield, IL, College of American Pathologists, 2014, pp 8

95. Food and Drug Administration. High-Throughput Sequencing Technologies for Microbial Identification and Detection of Antimicrobial Resistance Markers. [Internet], 2014. Available at http://www.fda.gov/downloads/MedicalDevices/NewsEvents/Workshops Conferences/UCM390380.pdf

96. Abate AR, Hung T, Sperling RA, Mary P, Rotem A, Agresti JJ, Weiner MA, Weitz DA: DNA sequence analysis with droplet-based microfluidics. Lab Chip 2013, 13:4864–4869