Clinical Infectious Diseases

REVIEW ARTICLE

Understanding the Promises and Hurdles of Metagenomic Next-Generation Sequencing as a Diagnostic Tool for Infectious Diseases

Patricia J. Simner,[1] Steven Miller,[2] and Karen C. Carroll[1]

[1]Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, Maryland; and [2]Department of Laboratory Medicine, University of California, San Francisco

Agnostic metagenomic next-generation sequencing (mNGS) has emerged as a promising single, universal pathogen detection method for infectious disease diagnostics. This methodology allows for identification and genomic characterization of bacteria, fungi, parasites, and viruses without the need for a priori knowledge of a specific pathogen directly from clinical specimens. Although there are increasing reports of mNGS successes, several hurdles need to be addressed, such as differentiation of colonization from infection, extraneous sources of nucleic acid, method standardization, and data storage, protection, analysis, and interpretation. As more commercial and clinical microbiology laboratories develop mNGS assays, it is important for treating practitioners to understand both the power and limitations of this method as a diagnostic tool for infectious diseases.

Keywords. metagenomics; diagnosis; next-generation sequencing; bioinformatics.

Infectious diseases remain leading causes of morbidity and mortality among all patient populations worldwide [1]. The spread of multidrug-resistant pathogens such as *Candida auris* and carbapenemase-producing gram-negative organisms and the reemergence of Ebola and Zika viruses highlight the ongoing challenges with diagnosis and management of patients with infectious diseases. Accurate diagnosis can be challenging due to a wide variety of pathogens causing clinically indistinguishable diseases. Current methods, such as culture, nucleic acid amplification tests, and serologic assays, generally require the use of a battery of tests to attempt to establish a diagnosis. Often times these methods still rely on a growth amplification step of viable microorganisms in culture for identification and antimicrobial susceptibility testing that take a minimum of 48 hours for commonly encountered pathogens and longer for more fastidious organisms (up to weeks for more insidious pathogens such as fungi and mycobacteria). Novel technologies such as syndromic multiplex polymerase chain reaction (PCR) panels, 16S ribosomal DNA (16S rDNA) Sanger sequencing, and matrix-assisted laser desorption/ionization–time-of-flight mass spectrometry (MALDI-TOF MS) have decreased turnaround times and subsequently beneficially impacted patient care [2]. Despite these advances, the etiology of infectious diseases remains unknown in up to 60% of cases depending on the clinical syndrome [3, 4]. Missed diagnoses due to the limitations of current microbiologic methodologies drive the use of empiric broad-spectrum antibiotics preventing the use of targeted, curative therapies.

Recent advances and lower costs of next-generation sequencing (NGS) technologies, more rapid, user- friendly data analysis tools, and the creation of accurate, comprehensive databases has allowed for NGS applications to cross the divide from microbial research to diagnostic microbiology [4]. One such application, unbiased metagenomic next-generation sequencing (mNGS), has the capability to overcome limitations of current diagnostic tests allowing for hypothesis-free, culture-independent, pathogen detection directly from clinical specimens. This method allows for universal pathogen detection regardless of the type of microbe (viruses, bacteria, fungi, and parasites) and can even be applied for novel organism discovery, potentially enabling replacement of many targeted pathogen tests with a single mNGS assay (Figure 1). As the availability of mNGS clinical diagnostic tests increases, it is important for treating practitioners to understand both its strengths and limitations as a tool for infectious disease diagnostics.

WHAT IS mNGS?

Next-generation sequencing technologies allow for sequencing of multiple individual DNA molecules in parallel regardless of composition, generating millions to billions of reads per instrument run. Reads are the basic element produced by DNA sequencing that are composed of a series of sequential bases (adenine, guanine, thymine, and cytosine) making up the DNA fragment, which can vary in size from small reads (75–600 bp) to long reads (1000–10 000s bp) depending on the sequencing technology. NGS overcomes many of the limitations of

Received 28 July 2017; editorial decision 27 September 2017; accepted 11 October 2017; published online October 12, 2017.

Correspondence: P. J. Simner, Johns Hopkins University School of Medicine, Meyer B1-193, 600 N Wolfe St, Baltimore, MD 21287 (psimner1@jhmi.edu).

Clinical Infectious Diseases® 2018;66(5):778–88

DOI: 10.1093/cid/cix881

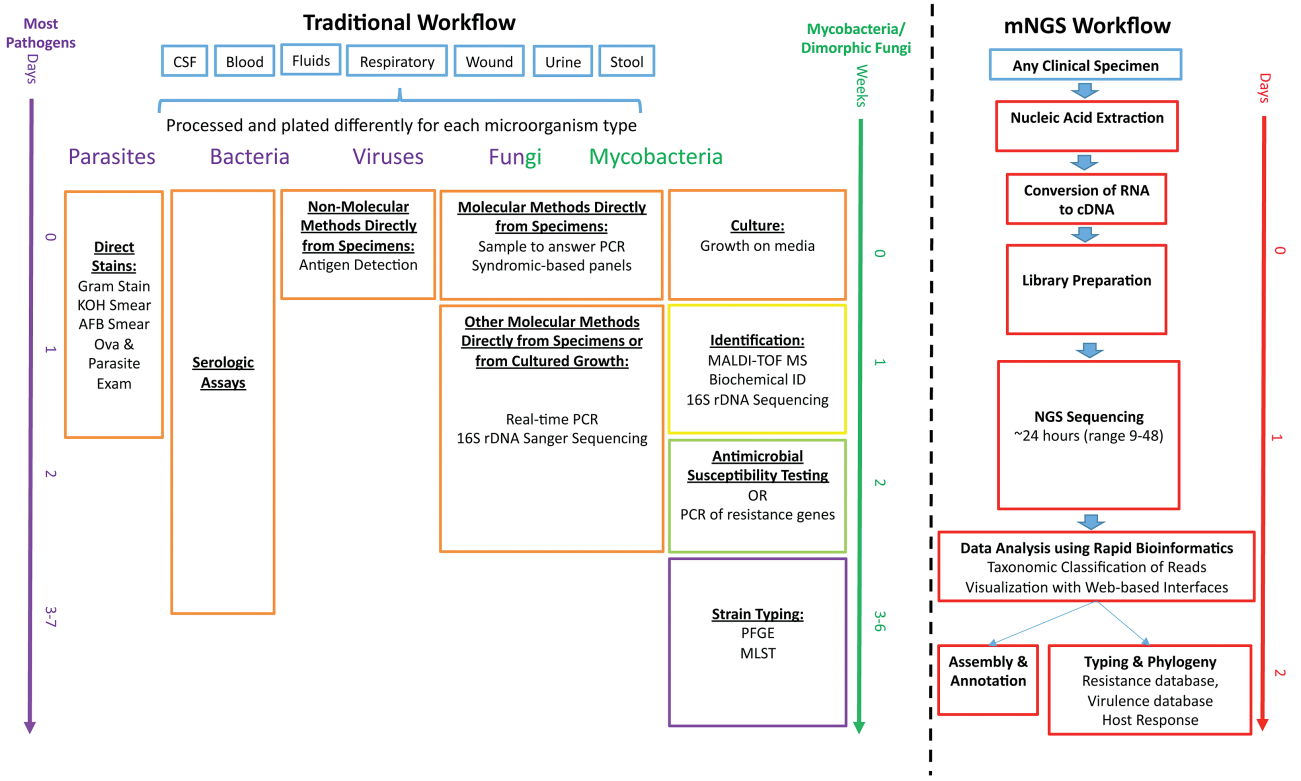


Figure 1. Summary of the traditional timeline and workflow in diagnostic medical microbiology laboratories and the future state with the incorporation of metagenomic next-generation sequencing (mNGS) methodologies. Current organism detection techniques (orange), identification (yellow), antimicrobial susceptibility testing (green), and strain typing (purple) can take up to a week or longer from specimen collection (blue) to strain typing results. mNGS has the capability of greatly reducing turnaround times and providing all the data summarized in current methods in a single modality (red) and could potentially provide all of these within 24–48 hours of specimen receipt. To date, the available evidence is poor to use antibiotic resistance gene detection to predict phenotypic antimicrobial susceptibility testing profiles for clinical care [16]. Abbreviations: AFB, acid-fast bacilli; cDNA, complementary DNA; CSF, cerebrospinal fluid; KOH, potassium hydroxide; MALDI-TOF MS, matrix-assisted laser desorption/ionization–time-of-flight mass spectrometry; MLST, multilocus sequence typing; mNGS, metagenomic next-generation sequencing; PCR, polymerase chain reaction; PFGE, pulsed-field gel electrophoresis; 16S rDNA, 16S ribosomal DNA.

traditional Sanger sequencing, which requires targeted sequencing (preamplification of a known target, such as 16S rDNA) of a uniform or low-diversity sample (the sample ideally is composed of a single organism or a maximum of 3 organisms).

The most common applications of NGS in diagnostic microbiology laboratories (Figure 2) include (1) whole-genome sequencing (sequencing and assembly of the genome of a pathogen of interest [eg, evaluating genetic relatedness during outbreak investigations, identification of new species]) [5–7]; (2) targeted NGS with different methods for enrichment including amplification or probe hybridization (ie, 16S rDNA bacterial profiling or PCR amplification of other specified targets followed by NGS) [7, 8]; or (3) mNGS (defined in more detail below) [9]. We refer the reader to other publications on this topic that summarize common terms and approaches used in DNA sequence analysis [10], the different NGS technologies [11–13], and the various NGS methods, bioinformatics tools [14] and clinical applications [10, 15]. This review will focus on mNGS for infectious disease diagnostic pathogen detection. Other applications of mNGS testing directly from specimens that are also not covered in this review include antimicrobial resistance gene and virulence gene detection, strain typing, and host immune response profiling [16].

mNGS, also known as unbiased or agnostic NGS or clinical metagenomics, is a method where all the nucleic acid (DNA and/or RNA) of a specimen is sequenced in parallel. This results in isolation and amplification of both host and pathogen nucleic acid sequenced from the specimen (Figure 2). Additional sources of nucleic acid that may be concurrently sequenced include nucleic acid introduced at the time of collection, within collection tubes, from the environment, and in sequencing reagents. Due to these complexities and potential breadth of detection, interpretation of mNGS results directly from clinical specimens can be difficult and requires careful interpretation and consideration.

HOST NUCLEIC ACID AMPLIFICATION: A BURDEN OR AN ADVANTAGE?

As the majority of a specimen is of human origin and the human genome is much larger than that of microorganisms (1000 times larger than bacterial genomes), the results from mNGS from clinical specimens typically generate >99% host

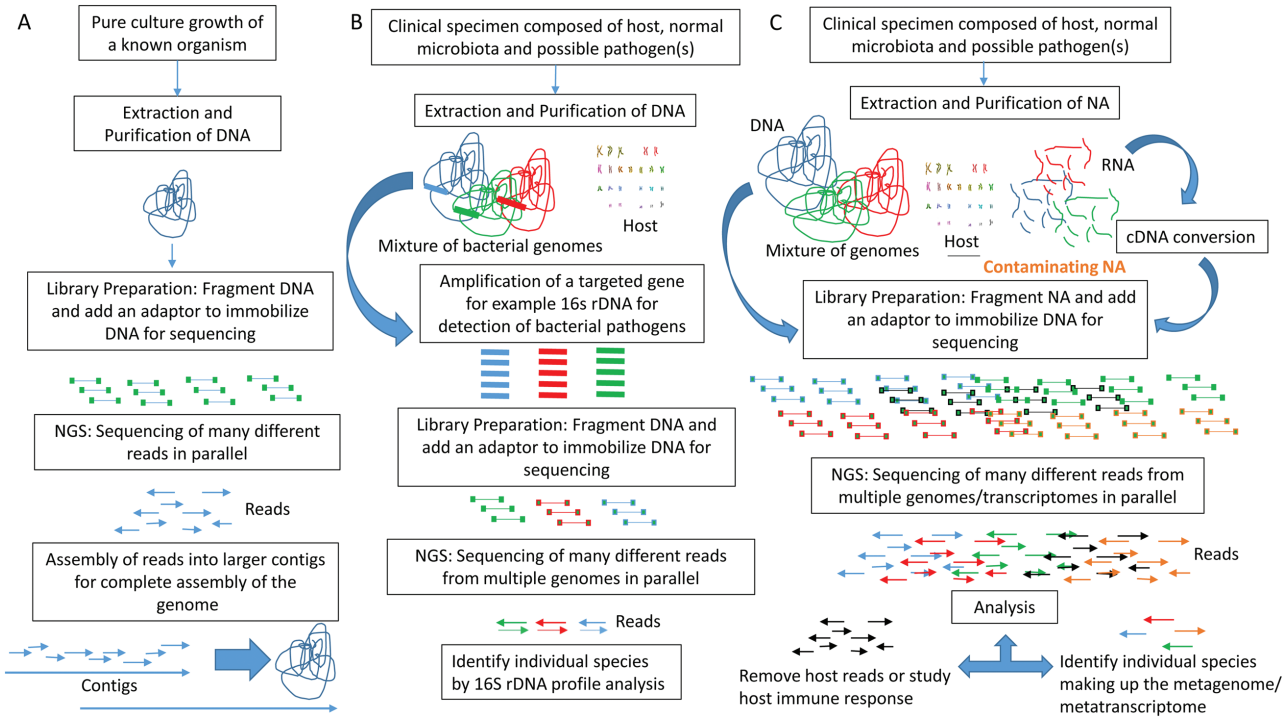


Figure 2. An overview of the different applications of next-generation sequencing analysis. *A*, Whole-genome sequencing of a pure organism from cultured growth. *B*, Targeted amplification of 16S rDNA from a clinical specimen for bacterial profiling. *C*, Metagenomic next-generation sequencing from clinical specimens. The nucleic acid composition of the specimens includes host (black), microbiome and pathogen detection (blue, green, and red), and last, the introduction of contaminating nucleic acid (orange). Analysis of reads generally involves removing host DNA from microbial DNA. The host DNA reads can be used to study the host immune response. The microbial reads are analyzed to identify the composition and abundance of reads of organisms present. The study of RNA can allow for transcriptome-based analysis to identify organisms that are transcriptionally active. Abbreviations: cDNA, complementary DNA; NGS, next-generation sequencing; NA, nucleic acid; 16S rDNA, 16S ribosomal DNA.

reads [17]. Successful clinical applications of mNGS have demonstrated that 0.00001%–0.7% of total reads (10 to tens of thousands of pathogen reads out of a total of millions of reads) have been used to make successful diagnoses (Table 1, mNGS results column). Pathogen reads make up a minute fraction of the sequencing results. Host nucleic acid can be depleted [18] or, if pathogen reads are enriched, more microorganism reads can be generated to more confidently provide a diagnosis and allow for increased analytical sensitivity (limit of detection). Additionally, if a robust immune response is generated in response to the pathogen, the specimen will contain a high concentration of leukocytes, further contributing to host nucleic acid in the sample and further limiting pathogen detection. Thus, the higher the sequencing depth (the higher number of reads generated/specimen) of the method, the higher the likelihood that the pathogen will be sequenced and detected (more reads equals higher sensitivity). For a conceptual illustration of how cellularity and overall nucleic acid affect analytic sensitivity, see Figure 1 in Schlaberg et al [4]. One potential advantage of sequencing amplified host nucleic acid is the opportunity to evaluate the host immunologic response associated with the presence and type of infection. This can include the presence of gene biomarkers associated with an antiviral or antibacterial response (specific expression of cytokines, chemokines, interferons, etc), genes involved in signaling pathways, and adaptive immunity. These biomarkers can provide insight into the pathogenesis of the microorganisms detected [19].

HOW TO DISCRIMINATE BETWEEN COLONIZATION AND INFECTION?

Detection of microorganisms in clinical samples by mNGS can reflect normal microbiota, transient colonizers, sample contamination, and/or infection. Initial applications of mNGS focused on detection of pathogens in normally sterile specimens, such as cerebrospinal fluid (CSF) and brain biopsies, simplifying the assignment of clinical significance for detected organisms [17, 20–23]. That being said, even body sites that are considered normally "sterile" may contain a microbiome, which can be defined and detected by deep sequencing [24]. Reports in the literature are now showing applications from other specimen types where normal microbiota can further complicate interpretation of results, such as respiratory specimens [19, 25]. These methods will likely require further studies to determine the best approach to distinguish colonization from infection. Currently, mNGS is only able to quantify pathogen reads as a percentage of the total number of sequence reads. Clinical microbiology laboratories have developed procedures throughout the years to distinguish true pathogens from colonizers in cultures, and similar

Table 1. Successful Clinical Application of Metagenomic Next-Generation Sequencing for Infectious Disease Diagnostics

Case(s)	mNGS Results and Specimen Types	Confirmatory Testing	Diagnosis	Outcome	Reference
Central nervous systems infections					
14-year-old boy with severe combined immunodeficiency presented 3 times for medical care over a 4-month period with fever and headache that progressed to hydrocephalus and status epilepticus necessitating a medically induced coma.	475 of 3 063 784 (0.016%) reads from the CSF specimen mapped to the Leptospiraceae family with the closest matched genome of *Leptospira borgpetersenii*.	PCR and serologic testing at the CDC subsequently confirmed the presence of *Leptospira santarosai* infection	Neuroleptospirosis	Treated with a course of penicillin G and discharged after 1 month with mild neurologic deficits.	[22]
42-year-old man with chronic lymphocytic leukemia presented with bilateral hearing loss. He developed neurological deterioration.	1612 of 134 068 968 (0.0012%) reads from brain tissue RNA aligned to astrovirus, no astrovirus reads were obtained from the 6 658 656 reads from CSF.	RT-PCR for astrovirus from the brain biopsy	Neuroinvasive astrovirus infection	Treated empirically with high-dose valacyclovir, broad-spectrum antibiotics, and IVIG. After the NGS analysis revealed astrovirus, the patient was treated with ribavirin and IVIG but eventually passed away.	[35]
3 male (aged 63, 63, and 72 years) breeders of variegated squirrels had progressive encephalitis	Variegated squirrel 1 bornavirus reads detected in brain samples from all 3 patients.	RT-qPCR	Variegated squirrel bornavirus encephalitis	All 3 patients passed away prior to mNGS results.	[43]
11-year-old girl from Mexico presented with 4 weeks of headache, back pain, nausea, and emesis. CSF was found to be positive for EBV and HHV7. She was treated with acyclovir and discharged. Two weeks later she presented with back pain and worsening headache. She was rehospitalized with a diagnosis of tuberculosis.	277 of 23 638 587 (0.0012%) reads from CSF DNA aligned to the *Brucella* genus with most aligning to *B. melitensis*, no *Brucella* reads were founds in the 9 161 626 CSF RNA reads.	Brucella PCR and serologies	Chronic neurobrucellosis	Treated with a course of doxycycline and rifampin. Two weeks after starting therapy she reported that her symptoms fully resolved.	[21]
15-year-old girl with diabetes presented to an emergency department with 7 days of progressive symptoms including right arm weakness, headache, vomiting, ataxia, and confusion. She clinically deteriorated over 8 days where she developed intracranial hypertension and cardiac arrest and died.	20 145 of 2 813 691 (0.7%) reads from CSF RNA, 13 of 3 714 322 (0.0004%) reads from brain biopsy DNA and 8 of 3 482 508 (0.0002%) reads from brain biopsy RNA aligned to *Balamuthia mandrillaris*.	PCR at the CDC confirmed the diagnosis of *B. mandrillaris* encephalitis.	Primary amoebic meningoencephalitis	Patient died prior to receiving therapy for *B. mandrillaris*.	[44]
A prospective series of 10 patients with neurologic problems indicating possible infection for whom conventional studies were negative or inconclusive. NGS successfully identified with high confidence the infectious agent in 3 of 10 cases: Patient 5: 52-year-old man was admitted for lower extremity weakness, right hemiparesis and a simple partial motor seizure. Patient 8: 67-year-old woman with osteomyelitis and lung disease presented with multifocal brain and spinal lesions. Patient 10: 44-year-old woman with a solid organ transplant presented with facial paralysis. Brain MRI demonstrated 3 enhancing lesions. In 2 cases the results were indeterminate and the other 5 were negative.	Patient 5: 8944 of 26, 919, 065 (0.03%) brain biopsy RNA reads aligned to JC virus. Patient 8: 15 of 13 990 253 (0.0001%) of a brain biopsy nodular lesion DNA aligned to *Mycobacterium tuberculosis*. Patient 10: 18 of 21 319 274 (0.00009%) reads from paraffin embedded brain biopsy aligned to EBV.	Patient 5: Pathology results showed marked astrogliosis and intranuclear inclusions in oligodendrocytes and positive immunostaining for SV40 T antigen (a surrogate for JC virus). Patient 8: Pathology studies of the sample showed necrotizing granulomas. Patient 10: In situ hydridization was positive for EBV-encoded RNA.	Patient 5: Progressive multifocal leukoencephalopathy. Patient 8: Granuloma, tuberculosis Patient 10: EBV encephalitis	Patient 5: Not described. Patient 8: The patient responded to antituberculous treatment. Patient 10: Not described.	[17]
34-year-old Australian man with X-linked agammaglobulinemia suffering from 3 years of meningoencephalitis that defied an etiologic disease despite extensive conventional testing, including brain biopsy.	5 of 25 069 677 (0.00002%) and 2 of 13 661 871 (0.00001%) reads of the CSF and brain biopsy aligned to Cache Valley virus.	CVV RT-PCR of the brain biopsy and immunohistochemistry straining of the FFPE brain tissue	Chronic viral meningoencephalitis: Cache Valley virus	Pooled IVIG was attempted but the patient passed away.	[23]
16-year-old boy from Vietnam suddenly became ill with fevers and rigors, followed by back pain, diarrhea, neck stiffness, limb weakness with eventual flaccid paralysis with convulsions and cognitive decline. mNGS was performed on CSF, urine, plasma, and rectal swab specimens.	The urine sample was positive for JEV RNA.	Seroconversion and PCR analysis confirmed findings.	JEV encephalitis	The patient regained consciousness on day 9 of hospitalization, the limb strength gradually returned.	[20]

Table 1. Continued

Case(s)	mNGS Results and Specimen Types	Confirmatory Testing	Diagnosis	Outcome	Reference
58-year-old female lung transplant recipient with meningoencephalitis and unexplained transaminitis.	CSF sample was positive for hepatitis E virus.	Positive serum IgM antibody and plasma hepatitis E virus level of 5960000 IU/mL.	Hepatitis E meningoencephalitis	Patient was treated with ribavirin. The lung transplant donor serum tested positive for hepatitis E virus IgG and IgM antibody, indicating possible transmission through organ donation.	[45]
68-year-old man with mantle cell lymphoma presenting with fever, chills, lethargy, and confusion.	CSF sample was positive for St Louis encephalitis virus.	CSF reverse-transcription PCR and viral culture positive for St Louis encephalitis virus.	St Louis encephalitis virus	Patient's condition deteriorated and was transferred to comfort care. He passed away the following day.	[46]
Respiratory tract infections					
A series of 22 hematopoietic stem cell transplant recipients with acute respiratory illnesses. mNGS was applied to study both the microbial composition and host response of BAL fluid specimens	mNGS confirmed all microbes identified by standard testing (human metapneumovirus, RSV, *Stenotrophomonas maltophilia*, HHV6, and CMV). mNGS identified previously unrecognized LRTI pathogens for which standard testing was negative (human coronavirus 229E, human rhinovirus A, *Corynebacterium propinquum*, and *Streptococcus mitis*)	6/22 confirmed by standard testing 6/22 negative by standard testing but confirmed mNGS findings by independent PCR testing. 10/22 mNGS identified microbes of uncertain or unlikely pathogenicity that were not confirmed by standard testing nor independent PCR.	mNGS confirmed the diagnosis of acute respiratory illness in 6 patients mNGS identified 6 previously unrecognized pathogens of acute respiratory illness.	Clinical outcomes not provided.	[19]
Patient 1: 41-year-old woman with connective tissue disorder- associated interstitial lung disease was admitted with hypoxic respiratory failure requiring mechanical ventilation. A chest radiograph revealed bilateral infiltrates and a mini-BAL was performed. Patient 2: 59-year-old man with abdominal sepsis developed hypoxic respiratory failure requiring mechanical ventilation. A mini-BAL was performed.	MinION real-time sequencing analysis: Patient 1: 9 hours after specimen receipt real-time sequencing analysis demonstrated 3217 bp aligned to *P. aeruginosa*. Patient 2: 6 high-quality DNA sequencing reads measuring between 909–8288 aligned to *Staphylococcus aureus*	Patient 1: Cultures yielded $>10^4$ CFU of *P. aeruginosa* 23 hours after lavage was collected. Patient 2: Cultures yielded $>10^4$ CFU *S. aureus* 24 hours after specimen collection.	Real-time sequencing analysis on the MinION confirmed culture results of respiratory pathogens.	Clinical outcomes not provided.	[25]
Comparison of RNA-seq based metagenomic analysis to a respiratory viral PCR panel for 109 pediatric nasopharyngeal swabs.	Untargeted mNGS detected 86% of known respiratory virus infections and detected an additional 12 viruses.	mNGS detected 86% of known infections. Only 33% of discordant samples confirmed initial PCR result.	Confirmation of respiratory viral diagnostic findings by mNGS	Clinical outcomes not provided.	Manuscript in preparation. Personal communication.
Ocular infections					
Intraocular fluid samples were obtained from 5 subjects with uveitis and 1 subject with bilateral chronic uveitis with unknown etiology.	Patient samples were positive with 423 reads of 16919211 (0.03%) reads from aqueous fluid aligning to HSV-1; 8469 of 4551967 (0.1% reads) reads from vitreous fluid aligned to *Cryptococcus neoformans*; 1853 of 10759511 (0.02%) reads from vitreous fluid aligning to *Toxoplasma gondii*; 585 of 1648220 (0.41%) reads of aqueous humor and 10 of 12111540 (0.01%) reads from vitreous fluid aligning to rubella virus. Two samples were determined to be negative for any pathogens.	For HSV-1, *C. neoformans* and *T. gondii*, the diagnosis was previously established by standard methods. The rubella virus uveitis was confirmed by RT-PCR of the aqueous fluid.	A new diagnosis of chronic rubella virus uveitis	Clinical outcomes not provided.	[47]

Table 1. Continued

Case(s)	mNGS Results and Specimen Types	Confirmatory Testing	Diagnosis	Outcome	Reference
A retrospective series of 16 cases of infectious keratitis. All specimens were FFPE tissues.	Sequencing generated 20 million to 46 million reads per sample. On average, 96% of the reads were classified as human, 1.7% represented microbial sequences. In total, 4 bacterial and mycobacterial cases, 5 of 6 fungal cases, 3 of 3 *Acanthamoeba* cases, and 1 case of 3 herpetic keratitis successfully identified by mNGS.	Retrospective analysis confirming previous culture and PCR results.	mNGS confirmed the diagnosis in 13 of 16 infectious keratitis cases.	Clinical outcomes not provided.	
Bloodstream infections					
Serum samples from 15 patients with known Zika virus infections in Brazil.	13 of 15 samples were positive for Zika virus by mNGS ranging from 2 to 281 099 reads per sample (0.0004%–4.1% of total reads). Five samples were also positive for Chikungunya virus.	All 15 samples were positive for Zika virus RT-PCR. Two of 5 samples were confirmed positive for Chikungunya virus by nested RT-PCR.	Confirmation of Zika virus infection and discovery of coinfection with Chikungunya virus in 2 patients.	Clinical outcomes not provided.	[48]
35-year-old man with ventricular septal defect admitted with fever of 3 months' duration.	Cardiac valve vegetation showed 97% of bacterial reads were genes of *Abiotrophia defectiva*.	Confirmation was not performed.	*Abiotrophia defectiva* endocarditis	Patient treated with antibiotics for 4 weeks and discharged.	[49]
Three patients in central Africa presenting with acute hemorrhagic fever.	Sequencing of the third patient's serum yielded 0.029% of reads with nucleotide or protein homology to a novel rhabdovirus.	Confirmatory PCR showed viral titers of 1.09×10^6 RNA copies/mL.	Novel rhabdovirus: Bas-Congo virus	First 2 patients died, third survived after treatment with fluids, blood transfusion and empiric antibiotics. No subsequent cases identified.	[50]
Four blood samples from patients with known viral infections.	Nanopore (MinION) sequencing detected viruses down to 1×10^5 copies/mL.	Chikungunya virus, Ebola virus, hepatitis C virus	Confirmation of known viral infections in 6-hour turnaround time.	Clinical outcomes not provided.	[41]
Gastrointestinal system infections					
A series of 6 patients who underwent cholecystectomy for acute cholecystitis. Four of 6 patients has mNGS analysis that demonstrated possible mono- or polymicrobial infections.	3 of 4 patients had a predominance of *Escherichia coli* and 1 sample had both *E. coli* and *Klebsiella pneumoniae* predominating among bacterial species. ESBL genes carried by a few *E. coli* strains were also detected.	mNGS results were confirmed by culture and AST.	Identification of Enterobacteriaceae pathogens	Detection of ESBL genes correlated with antimicrobial susceptibility test profiles. Clinical outcomes not provided.	[42]
Stool samples positive (n = 22) or negative (n = 5) for *Clostridioides* (previously *Clostridium*) *difficile* were tested by mNGS.	*C. difficile* was found in 86% of known positive samples by mNGS. Additional pathogens were detected in up to 27% of samples.	Co-detected pathogens were confirmed by PCR.	*C. difficile* infection	Clinical outcomes not provided.	[51]

Abbreviations: AST, antimicrobial susceptibility testing BAL, bronchoalveolar lavage; CDC, Centers for Disease Control and Prevention; CFU, colony-forming units; CMV, cytomegalovirus; CSF, cerebrospinal fluid; CVV, Cache Valley virus; EBV, Epstein-Barr virus; ESBL, extended-spectrum β-lactamase; FFPE, formalin-fixed, paraffin-embedded; HHV, human herpesvirus; HSV, herpes simplex virus; IgG, immunoglobulin G; IgM, immunoglobulin M; IVIG, intravenous immunoglobulin; JEV, Japanese encephalitis virus; LRTI, lower respiratory tract infection; mNGS, metagenomic next-generation sequencing; MRI, magnetic resonance imaging; NGS, next-generation sequencing; PCR, polymerase chain reaction; qPCR, quantitative polymerase chain reaction; RSV, respiratory syncytial virus; RT-PCR, reverse-transcription polymerase chain reaction.

approaches are being developed for mNGS. A recent example is the use of "spike-in"–based calibration to determine total microbial load as a proxy to convert percentage reads in relation to quantitating bacterial burden in colony-forming units (CFU) per milliliter [26]. Another approach to distinguish colonization from infection is to monitor and assess the host immune response. Langelier et al found that expression of a multigene immune response composite metric was significantly increased in hematopoietic cellular transplant patients with confirmed respiratory pathogens relative to those without a defined pathogen. These results suggest that, even in patients who have significant immunosuppression, the immune response may be used as a biomarker of active infection when utilizing host reads from mNGS analysis of respiratory specimens [19].

OTHER EXTRANEOUS SOURCES OF NUCLEIC ACID: PROVIDING FURTHER COMPLEXITY TO RESULT INTERPRETATION

The capability of unbiased sequencing directly from patient specimens is what makes mNGS an ideal, hypothesis-free diagnostic tool for infectious diseases. However, it also brings challenges as nucleic acid contamination occurs at several steps in the process from specimen acquisition to specimen processing and the environment, making interpretation of results challenging. Samples should be collected with much caution; for example, lumbar punctures to obtain CSF specimens can be contaminated by normal skin microbiota during the procedure. Similarly, surgical specimens may be contaminated during handling of the biopsy [17]. In addition, specimen collection containers are ideally verified to be DNA/RNA free as they too may serve as a source of contaminating nucleic acid. The development and strict adherence of specimen collection guidelines are crucial for reducing risk of contamination introduced at the time of specimen collection [9].

Furthermore, many researchers have shown that most reagents utilized for mNGS also introduce extraneous, unwanted sources of DNA during the sequencing process; this phenomenon is referred to as the "kit-ome" [9]. A study by Salter et al demonstrated that extraneous DNA is ubiquitous in commonly used DNA extraction reagents and other laboratory reagents used for NGS. The contamination can critically impact results obtained from samples, especially those containing low microbial biomass [27, 28]. This becomes particularly relevant when there is a lack of an obvious pathogen (eg, *Mycobacterium tuberculosis* or JC virus where they are not found as contaminants in reagents or part of normal microbiota) especially for mNGS analysis of CSF specimens where common reagent contaminants can include members of the Enterobacteriaceae or *Cutibacterium acnes* (previously known as *Propionibacterium acnes*), making clinical relevance difficult to discern. Thus, it is imperative that a no-template control (a sample [ie, water or extraction buffer] that is run through the entire mNGS procedure) be included in mNGS analyses to determine the nucleic acid background of the reagents used for sequencing. The no-template control can be used when interpreting results to help filter out contaminating background reads. In fact, the no template control has been utilized as part of a cutoff to determine the relevance of mNGS findings from patient specimens. A recent report used a cutoff of ≥10 ratio of the reads per million sample divided by the reads per million of the no-template control from any given taxon (species, genus, or family) [21]. Another approach is to align the reads to the genome of the pathogen to see if the reads span different areas of the genome. If the reads are localized to a restricted area of the genome, the reads are more likely to represent a contaminant, whereas if the reads span the genome, the organism is more likely to represent a true organism detection [29].

mNGS METHODS AND VALIDATION IN THE CLINICAL MICROBIOLOGY LABORATORY

Currently, there is no standardized protocol for mNGS. Thus, it is very important for the physician interpreting results to have a good understanding of the scope of the methods being applied as these impact the ability of the assay to detect certain pathogen types and affects interpretation of results. Methods of mNGS can include a DNA- and/or RNA-based approach. If a DNA-based approach is solely applied, pathogen types, except RNA viruses, will be detected. An RNA-based approach is required for detection of RNA viruses and further provides transcriptome-based analysis of other pathogen types and even the host immune response. A DNA approach will indicate what organism(s) are present, but an RNA approach can further reveal what organism(s) are transcriptionally active.

Another major factor that will impact results is the type of extraction method utilized in the laboratory. Nucleic acid recovery may not be equal for all pathogen types [30]. For example, mycobacteria require significant cell wall disruption to efficiently lyse the organisms for nucleic acid release. Different extraction methods may also be required for different specimen types [4]. Efficient extraction methods are a critical step to achieve truly unbiased sequencing of a sample.

In addition, the limit of detection is important to determine, as a negative mNGS result may simply reflect the high leukocyte count of a sample or low sequencing depth of a specimen rather than the absence of a pathogen. The reported limit of detection of mNGS from synthetic CSF matrix varies by organism type from a low of 0.01 CFU/mL for *Cryptococcus neoformans*, approximately 9 CFU/mL for *Klebsiella pneumoniae* and *Streptococcus agalactiae*, 9.4 copies/mL for cytomegalovirus, 55 organisms/mL for *Toxoplasma gondii*, and 100 copies/mL of human immunodeficiency virus (HIV) type 1, to a high of 130 CFU/mL of *Aspergillus niger* for detection [4]. This compares to 10^{4-5} CFU/mL for detection of pathogens by special stains directly from a specimen (ie, Gram stain), 10^{2-3} CFU/mL for

growth in culture, and 10–100 CFU/mL for nucleic acid amplification–based methods [31]. It also varies by specimen type, as the limit of detection of mNGS for hepatitis C virus and HIV from plasma has been reported as high as 1×10^4 copies/mL, as determined by quantitative PCR [32]. Use of an internal control is extremely important to identify analytic failures and specimens with unusual cellularity that may result in reduced analytical sensitivity. While a negative mNGS result does not completely rule out an infectious process, adequate recovery of the internal control can indicate the level of sensitivity achieved by the assay for that particular sample [4]. Results will still need to be interpreted within a clinical context.

Last, validation of mNGS by individual laboratories at the moment can be time consuming and extremely costly, as significant optimization is required for development, and de novo establishment of performance characteristics is required as there are no currently available US Food and Drug Administration (FDA)–cleared methods, instruments, and/or databases [4]. For example, validation of a single specimen type such as CSF will easily require an investment of more than US$100 000 simply for supply costs, not including technologist time and instrumentation fees. Thus, many hospitals may choose to send testing out to larger reference or commercial laboratories that have the resources to optimize and validate mNGS. Regardless of where testing is being performed, physicians need to have a strong understanding of the methods to appropriately interpret the results.

DATA STORAGE, ANALYSIS, AND DATABASES

A major bottleneck of mNGS for infectious disease diagnostics is the ability to decipher the data into clinically relevant information to positively impact patient care. The large sums of data generated by mNGS are burdensome in terms of storage and analysis. mNGS data include host reads, which introduces privacy concerns and requires HIPAA (Health Insurance Portability and Accountability Act)–compliant storage tools [33]. Host reads may be removed during data analysis steps, but can also be helpful to confirm a patient's identity to assist in ruling out mislabeling or contamination issues [34].

Rapid bioinformatics tools are required to allow for data analysis to obtain clinically actionable results in a meaningful timeframe. A diversity of data analysis tools exist, most of which first filter out host reads and then taxonomically classify the microbial reads to the most accurate taxonomic level (ie, species/subspecies level if possible, otherwise a genus-, family-, order-, class-, phylum-, or kingdom-level identification is provided) [4]. Two bioinformatics pipelines have been applied successfully in most clinical studies, the "sequence-based ultrarapid pathogen identification" (SURPI or SURPI+) pipeline and the Kraken pipeline [4, 17, 19–23, 29, 35]. Both of these pipelines rapidly align the sequencing reads to the National Center for Biotechnology Information (NCBI) nucleotide reference database and use taxonomic classification for more accurate read assignments [36, 37]. Bioinformatics programs also need to be straightforward with user-friendly interfaces for incorporation into clinical microbiology laboratories. Currently, most of these analytic tools require a significant degree of bioinformatics expertise that is typically not available in clinical laboratories [38]. Some suggest that many different algorithms using a variety of approaches, such as the use of k-mer, marker, and alignments, are required to analyze the data to ensure the highest sensitivity and specificity for taxonomic classification [9, 39].

Last, curated databases are required as several draft genomes or partial sequences available on NCBI contain erroneous information and can result in false-positive results. False-positive results using noncurated databases can result from low-complexity sequences matching low-quality reads from the sample (ie, computational noise), misannotated species, or contaminants from database entries that also contain reads to human DNA, sequencing adaptors, or vectors. False negatives may occur due to incomplete or lack of taxonomic representation in the databases. Due to these issues, there have been significant efforts put in place to create accurate, regulatory-grade databases, such as the FDA ARGOS database, that can be applied clinically [40].

TIME TO RESULTS

Currently, the turnaround time for mNGS has been reported to be anywhere from approximately 6 hours to 7 days (average of 48 hours) from specimen receipt depending on the sequencing technology, methods, and bioinformatics programs exploited [9, 22, 25, 41, 42]. In general, mNGS methodologies are labor intensive and require several steps from nucleic acid extraction, library preparation, and sequencing to data analysis. Depending on the sequencing chemistry (read lengths, paired-end vs single-end sequencing, depth of sequencing, or sequencing platforms), it can take up to 48 hours (average of ~ 24 hours) to generate the sequence data alone. Once the sequencing data are generated, rapid bioinformatic pipelines may yield results in as little as 1 hour with significant computational power to support the analysis [17]. mNGS reports can contain >2000 lines in length summarizing the reads from the sample (see Supplementary Data). Figure 3 demonstrates an example of a simplified mNGS Kraken report along with interpretation using a Web-based application tool for visualization of results incorporating heat maps to highlight predominant pathogens and alignment tools useful for result interpretation [29].

PRECISION MEDICINE TEAM FOR INTERPRETATION OF RESULTS

Perhaps the greatest challenge for the practicing infectious disease clinician is interpretation of the results generated by the sequencing laboratory. Due to the complexity of results

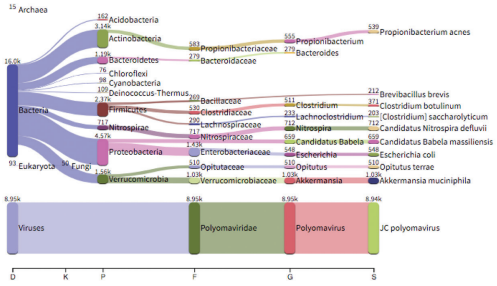
A

% of reads	Number of reads	Read Identity
96.92	26,089,388	Homo Sapiens
1.10	296,140	Unclassified
0.06	16,516	Bacterial reads
0.002	562	*Escherichia coli*
0.0003	78	*Pseudomonas putida group*
0.002	539	*Propionibacterium acnes*
0.04	9,441	Viruses
0.03	9,208	Polyomaviridae
0.03	9,208	Polyomavirus
0.03	9,208	JC Polyomavirus

C

Name	Max	PT1	PT2	PT3	PT4	PT5	PT6	PT7	PT8-S1	PT8-S2	PT9	PT10
Mycobacterium	71.43	0.4785%		0.4591%		0.02148%	0.1488%	0.5054%		71.43%	1.029%	
Polyomavirus	48.06					48.06%	1.116%					
Propionibacterium	37.5	0.9569%	37.5%	1.238%	22.45%	2.981%	5.134%	20.11%	20%	14.29%	15.23%	15.33%
Pseudomonas	16.91	3.828%	7.5%	3.892%	2.806%	0.3598%	2.679%	16.91%	10%	4.762%	3.909%	4.667%
Lymphocryptovirus	12										0.06859%	12%
Filobasidiella	10.05	10.05%				0.04297%	0.0744%	0.02732%				

B



D

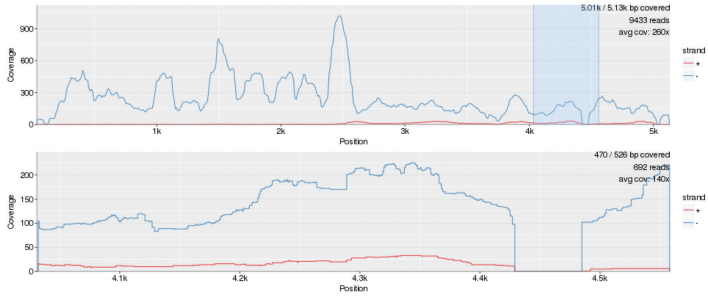


Figure 3. Example of results output and bioinformatics analysis tools for metagenomic next-generation sequencing data. *A*, A simplified Kraken report showing the number and percentage of sequence reads and their alignment identification using Kraken for a cerebrospinal fluid (CSF) specimen from a patient diagnosed with JC virus encephalitis [17]. The overall Kraken report summarizing the data from the CSF specimen is >2000 line listings long (see Supplementary Data). Of note, *Escherichia coli*, *Pseudomonas putida* group, and *P*ropionibacterium *acnes* (now *Cutibacterium acnes*) were considered reagent contaminants in this case as they were observed in the no-template control. *B–D*, Analysis modes of the Web-based Pavian program, a straightforward interface to analyze and compare complex metagenomics datasets. *B*, The number of sequence reads matching each taxa of interest are shown for the sample. Of note, almost all the virus reads align to JC polyomavirus. *C*, A heat map approach showing the percentage of microbially matched reads across multiple samples allowing for sample comparison. *D*, An interactive alignment tool showing the fold coverage of the reads over the whole JC virus genome [29].

generated from mNGS, some institutions have implemented precision medicine teams. These teams consist of representatives from medical microbiology, computational biology, infectious diseases, and other clinician groups who can discuss the results and provide interpretation of the mNGS results prior to reporting. This approach ensures that the most clinically relevant data are reported. Additionally, in the authors' experience, development of such a team was useful in determining direction of assay development and lobbying the institution for resources. Another approach, which is strongly advised, is to have an experienced laboratory director manually review the results prior to release [21].

SUCCESSFUL APPLICATIONS OF mNGS

mNGS directly from clinical specimens came to the forefront when the Chiu laboratory from the University of California, San Francisco (UCSF) applied the methodology to achieve diagnoses in a case of neuroleptospirosis and a case of neuroinvasive astrovirus [22, 35]. In both cases, standard microbial diagnostic techniques were applied, but no pathogens were identified. Unbiased mNGS was applied and successfully identified the infectious agents in these cases. These cases are examples of the power of NGS as a diagnostic tool. Initial applications of mNGS for diagnostics focused on central nervous system infections, mostly chronic infections, and have successfully diagnosed rare [22], novel [43], and atypical infectious etiologies [23] of encephalitis. One particular case of chronic Cache Valley virus encephalitis in an Australian patient highlights the ability of mNGS to identify pathogens not previously associated with a clinical phenotype [23]. A single report of the accuracy of mNGS compared to standard methods of 84 previously positive CSF specimens and 21 negative specimens observed a diagnostic sensitivity of 84.3% and specificity of 93.7% following discordant analysis [4]. Some recent articles describe the utility of the assay for detection of respiratory pathogens directly from bronchoalveolar lavage specimens from human stem cell transplant and lung transplant recipients [19, 25].

Table 1 summarizes the successful applications of mNGS in the literature to date (July 2017). All these studies have been performed using short read data (second-generation Illumina sequencing), with 2 exceptions. Recently, publications by Pendleton et al and Greninger et al demonstrated the utility of mNGS using the affordable, portable MinION sequencer (third-generation sequencing) that generates long reads with sequencing analysis that can occur in real time [25, 41]. The mNGS analysis of specimens by Pendleton et al resulted in a

more rapid turnaround time of the pathogens causing lower respiratory tract infections than standard culture methods [25]. For the most part, the successful application of mNGS has occurred for diagnoses of severe, insidious infections or has been performed in retrospective studies compared to standard-of-care results. Ideally, this technology would be applied in real time as an adjunct to current SoC testing.

USE OF mNGS FOR CLINICAL CARE: WHERE ARE WE NOW?

mNGS is currently being offered as a billable laboratory-developed test by both clinical and commercial laboratories. The question is, when should this rather expensive test be considered for clinical care? Based on the available literature, mNGS could be considered when SoC testing is unrevealing and can be used as a last resort effort to try to discern an infectious process. Alternatively, it may be considered for critically ill or severely immunocompromised patients where achieving a timely diagnosis is imperative for improved outcomes. Importantly, mNGS does not replace current SoC methods, but should rather be used as an adjunct to these methods. At this point in time, this question is still actively under investigation and further evidence is required to establish the use of mNGS in routine clinical care.

CONCLUSIONS

mNGS has emerged as a promising single, universal pathogen methodology for infectious disease diagnostics. In addition, mNGS will further be developed in years to come to evaluate antimicrobial resistance genes, strain typing, pathogen evolution, immune response to offending pathogens, and detection of virulence genes. As mNGS methods are still being developed and NGS technologies are rapidly evolving, there are still challenges ahead in terms of data interpretation and patient privacy. Infectious disease clinicians and other treating physicians can actively participate in the quality management of NGS diagnostic applications by appreciating the complexities and nuances of the methodologies being applied and ensuring that steps, as discussed in this and other reviews, are implemented to generate meaningful results beneficial to patient care.

Supplementary Data

Supplementary materials are available at *Clinical Infectious Diseases* online. Consisting of data provided by the authors to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the authors, so questions or comments should be addressed to the corresponding author.

Note

Potential conflicts of interest. K. C. C. reports grants from Accelerate Diagnostics, Curetis, BD Diagnostics, and GenePOC; personal fees from Roche Molecular, McGraw-Hill, American Society for Microbiology, and Elsevier. P. J. S. reports grants from Accelerate Diagnostics, BD Diagnostics, bioMérieux, Cellex, Check-Points Diagnostics, BV, and Hardy Diagnostics; and personal fees from the American Society for Microbiology and Accelerate Diagnostics. S. M. has a patent for the SURPI pipeline issued to UCSF, a patent for microbial standard reference materials for pathogen detection pending to UCSF, and a patent for historic contaminant database for mNGS applications pending to UCSF. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

1. Lozano R, Naghavi M, Foreman K, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet **2012**; 380:2095–128.
2. Tan KE, Ellis BC, Lee R, Stamper PD, Zhang SX, Carroll KC. Prospective evaluation of a matrix-assisted laser desorption ionization-time of flight mass spectrometry system in a hospital clinical microbiology laboratory for identification of bacteria and yeasts: a bench-by-bench study for assessing the impact on time to identification and cost-effectiveness. J Clin Microbiol **2012**; 50:3301–8.
3. Glaser CA, Honarmand S, Anderson LJ, et al. Beyond viruses: clinical profiles and etiologies associated with encephalitis. Clin Infect Dis **2006**; 43:1565–77.
4. Schlaberg R, Chiu CY, Miller S, Procop GW, Weinstock G; Professional Practice Committee and Committee on Laboratory Practices of the American Society for Microbiology; Microbiology Resource Committee of the College of American Pathologists. Validation of metagenomic next-generation sequencing tests for universal pathogen detection. Arch Pathol Lab Med **2017**; 141:776–86.
5. Rasko DA, Webster DR, Sahl JW, et al. Origins of the *E. coli* strain causing an outbreak of hemolytic-uremic syndrome in Germany. N Engl J Med **2011**; 365:709–17.
6. Salipante SJ, Roach DJ, Kitzman JO, et al. Large-scale genomic sequencing of extraintestinal pathogenic *Escherichia coli* strains. Genome Res **2015**; 25:119–28.
7. Salipante SJ, Sengupta DJ, Cummings LA, et al. Whole genome sequencing indicates *Corynebacterium jeikeium* comprises 4 separate genomospecies and identifies a dominant genomospecies among clinical isolates. Int J Med Microbiol **2014**; 304:1001–10.
8. Salipante SJ, Hoogestraat DR, Abbott AN, et al. Coinfection of *Fusobacterium nucleatum* and *Actinomyces israelii* in mastoiditis diagnosed by next-generation DNA sequencing. J Clin Microbiol **2014**; 52:1789–92.
9. Afshinnekoo E, Chou C, Alexander N, Ahsanuddin S, Schuetz AN, Mason CE. Precision metagenomics: rapid metagenomic analyses for infectious disease diagnostics and public health surveillance. J Biomol Tech **2017**; 28:40–5.
10. Goldberg B, Sichtig H, Geyer C, Ledeboer N, Weinstock GM. Making the leap from research laboratory to clinic: challenges and opportunities for next-generation sequencing in infectious disease diagnostics. MBio **2015**; 6:e01888–15.
11. Lavezzo E, Barzon L, Toppo S, Palù G. Third generation sequencing technologies applied to diagnostic microbiology: benefits and challenges in applications and data analysis. Expert Rev Mol Diagn **2016**; 16:1011–23.
12. Mardis ER. A decade's perspective on DNA sequencing technology. Nature **2011**; 470:198–203.
13. Metzker ML. Sequencing technologies—the next generation. Nat Rev Genet **2010**; 11:31–46.
14. Lindgreen S, Adair KL, Gardner PP. An evaluation of the accuracy and speed of metagenome analysis tools. Sci Rep **2016**; 6:19233.
15. Deurenberg RH, Bathoorn E, Chlebowicz MA, et al. Application of next generation sequencing in clinical microbiology and infection prevention. J Biotechnol **2017**; 243:16–24.
16. Ellington MJ, Ekelund O, Aarestrup FM, et al. The role of whole genome sequencing in antimicrobial susceptibility testing of bacteria: report from the EUCAST Subcommittee. Clin Microbiol Infect **2017**; 23:2–22. PMID: 27890457.
17. Salzberg SL, Breitwieser FP, Kumar A, et al. Next-generation sequencing in neuropathologic diagnosis of infections of the nervous system. Neurol Neuroimmunol Neuroinflamm **2016**; 3:e251.
18. Hasan MR, Rawat A, Tang P, et al. Depletion of human DNA in spiked clinical specimens for improvement of sensitivity of pathogen detection by next-generation sequencing. J Clin Microbiol **2016**; 54:919–27.
19. Langelier C, Zinter MS, Kalatar K, et al. Metagenomic next-generation seqeuceng detects pulmonary pathogens in hematopoietic cellular transplant patients with acute respiratory illnesses. bioRxiv **2017**. doi:https://doi.org/10.1101/102798.
20. Mai NTH, Phu NH, Nhu LNT, et al. Central nervous system infection diagnosis by next-generation sequencing: a glimpse into the future? Open Forum Infect Dis **2017**; 4:ofx046.
21. Mongkolrattanothai K, Naccache SN, Bender JM, et al. Neurobrucellosis: unexpected answer from metagenomic next-generation sequencing. J Pediatric Infect Dis Soc **2017**. doi:10.1093/jpids/piw066.
22. Wilson MR, Naccache SN, Samayoa E, et al. Actionable diagnosis of neuroleptospirosis by next-generation sequencing. N Engl J Med **2014**; 370:2408–17.

23. Wilson MR, Suan D, Duggins A, et al. A novel cause of chronic viral meningoencephalitis: Cache Valley virus. Ann Neurol **2017**; 82:105–14.
24. Wylie KM, Weinstock GM, Storch GA. Virome genomics: a tool for defining the human virome. Curr Opin Microbiol **2013**; 16:479–84.
25. Pendleton KM, Erb-Downward JR, Bao Y, et al. Rapid pathogen identification in bacterial pneumonia using real-time metagenomics. Am J Respir Crit Care Med **2017**. doi:10.1164/rccm.201703-0537LE.
26. Stämmler F, Gläsner J, Hiergeist A, et al. Adjusting microbiome profiles for differences in microbial load by spike-in bacteria. Microbiome **2016**; 4:28.
27. Naccache SN, Greninger AL, Lee D, et al. The perils of pathogen discovery: origin of a novel parvovirus-like hybrid genome traced to nucleic acid extraction spin columns. J Virol **2013**; 87:11966–77.
28. Salter SJ, Cox MJ, Turek EM, et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biol **2014**; 12:87.
29. Breitwieser FP, Salzberg SL. Pavian: interactive analysis of metagenomics data for microbiomics and pathogen identification. bioRxiv **2016**. doi:https://doi.org/10.1101/084715.
30. Hart ML, Meyer A, Johnson PJ, Ericsson AC. Comparative evaluation of DNA extraction methods from feces of multiple host species for downstream next-generation sequencing. PLoS One **2015**; 10:e0143334.
31. Leber AL, Everhart K, Balada-Llasat JM, et al. Multicenter evaluation of BioFire FilmArray meningitis/encephalitis panel for detection of bacteria, viruses, and yeast in cerebrospinal fluid specimens. J Clin Microbiol **2016**; 54:2251–61.
32. Kandathil AJ, Breitwieser FP, Sachithanandham J, et al. Presence of human hepegivirus-1 in a cohort of people who inject drugs. Ann Intern Med **2017**; 167:1–7.
33. Erlich Y, Narayanan A. Routes for breaching and protecting genetic privacy. Nat Rev Genet **2014**; 15:409–21.
34. Callaway E. Microbiome privacy risk. Nature **2015**; 521:136.
35. Naccache SN, Peggs KS, Mattes FM, et al. Diagnosis of neuroinvasive astrovirus infection in an immunocompromised adult with encephalitis by unbiased next-generation sequencing. Clin Infect Dis **2015**; 60:919–23.
36. Naccache SN, Federman S, Veeraraghavan N, et al. A cloud-compatible bioinformatics pipeline for ultrarapid pathogen identification from next-generation sequencing of clinical samples. Genome Res **2014**; 24:1180–92.
37. Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol **2014**; 15:R46.
38. Westblade LF, van Belkum A, Grundhoff A, et al. Role of clinicogenomics in infectious disease diagnostics and public health microbiology. J Clin Microbiol **2016**; 54:1686–93.
39. Afshinnekoo E, Meydan C, Chowdhury S, et al. Geospatial resolution of human and bacterial diversity with city-scale metagenomics. Cell Syst **2015**; 1:97–97.e3.
40. US Food and Drug Administration. Database for Reference Grade Microbial Sequences (FDA-ARGOS). Available at: https://www.fda.gov/MedicalDevices/ScienceandResearch/DatabaseforReferenceGradeMicrobialSequences/default.htm. Accessed 17 July 2017.
41. Greninger AL, Naccache SN, Federman S, et al. Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis. Genome Med **2015**; 7:99.
42. Kujiraoka M, Kuroda M, Asai K, et al. Comprehensive diagnosis of bacterial infection associated with acute cholecystitis using metagenomic approach. Front Microbiol **2017**; 8:685.
43. Hoffmann B, Tappe D, Höper D, et al. A variegated squirrel bornavirus associated with fatal human encephalitis. N Engl J Med **2015**; 373:154–62.
44. Greninger AL, Messacar K, Dunnebacke T, et al. Clinical metagenomic identification of *Balamuthia mandrillaris* encephalitis and assembly of the draft genome: the continuing case for reference genome sequencing. Genome Med **2015**; 7:113.
45. Murkey JA, Chew KW, Carlson M, et al. Hepatitis E virus-associated meningoencephalitis in a lung transplant recipient diagnosed by clinical metagenomic sequencing. Open Fourm Infect Dis **2017**; 4:ofx121.
46. Chiu CY, Coffey LL,Murkey J, et al. Diagnosis of fatal human case of mosquito-borne St. Louis encephalitis virus infection diagnosed by metagenomic sequencing, California, 2016. Emerg Infect Dis **2017**; 23:1964–68.
47. Doan T, Wilson MR, Crawford ED, et al. Illuminating uveitis: metagenomic deep sequencing identifies common and rare pathogens. Genome Med **2016**; 8:90.
48. Sardi SI, Somasekar S, Naccache SN, et al. Coinfections of Zika and Chikungunya viruses in Bahia, Brazil, identified by metagenomic next-generation sequencing. J Clin Microbiol **2016**; 54:2348–53.
49. Fukui Y, Aoki K, Okuma S, Sato T, Ishii Y, Tateda K. Metagenomic analysis for detecting pathogens in culture-negative infective endocarditis. J Infect Chemother **2015**; 21:882–4.
50. Grard G, Fair JN, Lee D, et al. A novel rhabdovirus associated with acute hemorrhagic fever in central Africa. PLoS Pathog **2012**; 8:e1002924.
51. Zhou Y, Wylie KM, El Feghaly RE, et al. Metagenomic approach for identification of the pathogens associated with diarrhea in stool specimens. J Clin Microbiol **2016**; 54:368–75.