The Journal of Infectious Diseases

SUPPLEMENT ARTICLE

From the Pipeline to the Bedside: Advances and Challenges in Clinical Metagenomics

Augusto Dulanto Chiang and John P. Dekker

Bacterial Pathogenesis and Antimicrobial Resistance Unit, Laboratory of Clinical Immunology and Microbiology, National Institute of Allergy and Infectious Diseases, Bethesda, Maryland

Next-generation sequencing (NGS) technologies have revolutionized multiple areas in the field of infectious diseases, from pathogen discovery to characterization of genes mediating drug resistance. Consequently, there is much anticipation that NGS technologies may be harnessed in the realm of diagnostic methods to complement or replace current culture-based and molecular microbiologic techniques. In this context, much consideration has been given to hypothesis-free, culture-independent tests that can be performed directly on primary clinical samples. The closest realizations of such universal diagnostic methods achieved to date are based on targeted amplicon and unbiased metagenomic shotgun NGS approaches. Depending on the exact details of implementation and analysis, these approaches have the potential to detect viruses, bacteria, fungi, parasites, and archaea, including organisms that were previously undiscovered and those that are uncultivatable. Shotgun metagenomics approaches additionally can provide information on the presence of virulence and resistance genetic elements. While many limitations to the use of NGS in clinical microbiology laboratories are being overcome with decreasing technology costs, expanding curated pathogen sequence databases, and better data analysis tools, there remain many challenges to the routine use and implementation of these methods. This review summarizes recent advances in applications of targeted amplicon and shotgun-based metagenomics approaches to infectious disease diagnostic methods. Technical and conceptual challenges are considered, along with expectations for future applications of these techniques.

Keywords. Metagenomics; next generation sequencing; clinical microbiology.

NEXT-GENERATION SEQUENCING (NGS) IN CLINICAL MICROBIOLOGY: A VARIETY OF APPLICATIONS

NGS technologies have revolutionized multiple areas in the field of infectious diseases, from high-resolution molecular epidemiologic studies [1, 2] and pathogen discovery [3] to characterization of genes mediating drug resistance [4, 5]. Routine sequencing of DNA in microbiology began, however, long before the availability of NGS approaches, with techniques developed by Frederick Sanger and colleagues >4 decades ago. The Sanger family of methods are based on template-directed DNA synthesis performed by primer extension in the presence of dideoxynucleotide chain terminators. Dideoxynucleotide incorporation is directed by the nucleotide sequence of the template, thereby generating a set of synthesized fragments terminated at different lengths that reflect the nucleotide sequence of the template. Electrophoretic separation of the synthesized fragments by length allows direct readout of the template sequence [6]. Automated derivatives of this original Sanger approach are still commonly used in clinical microbiology laboratories for the taxonomic identification of organisms on the basis of single-gene sequencing. Nevertheless, Sanger methods have a number of limitations, including an upper limit achievable contiguous sequencing length of approximately 1 kilobase (1000 bp), a requirement for a priori knowledge of target sequence content for primer design, and an inability to resolve mixed-sequence populations in a sample definitively.

Sanger techniques were followed by the development of NGS methods at the turn of the century. NGS approaches allow for parallel, high-throughput sequencing of large numbers (10^5–10^9) of individual DNA molecules, in contrast to the Sanger methods, which produce a single population-averaged sequence. The single-molecule discrimination provided by NGS permits resolution of heterogeneous genetic populations and allows shotgun-based assembly of whole-organism genomes. However, the sheer volume of sequence data generated by NGS methods and the complexity of analysis require sophisticated bioinformatics tools and expertise relative to simpler Sanger approaches, as will be explored in this review.

The basic unit of NGS data is the "read," a short stretch of sequence that reflects the order of nucleotides in a given single molecule of DNA in the sequencing reaction. Depending on the exact NGS approach, a single read will commonly range in length between tens of nucleotides to greater than a hundred thousand nucleotides [7]. A typical sequencing run, in turn, generates 10^5–10^9 individual reads. Depending on the analytic approach, the reads may be individually aligned to a

Correspondence: J. P. Dekker, MD, PhD, (john.dekker@nih.gov).

The Journal of Infectious Diseases® 2020;221(S3):S331–40

DOI: 10.1093/infdis/jiz151

reference database for identification and analysis of their origin, or they may be computationally assembled into larger contiguous pieces, called "contigs," representing larger fragments of genomes present in the sequenced material.

Broadly speaking, there are 2 types of NGS technologies: short-read and long-read methods [7, 8]. Short-read approaches, represented by Illumina and Ion Torrent technologies, produce reads between 35 and 300 nucleotides in length. Long-read sequencing technologies fall into a couple different groups. Pacific Biosciences (PacBio) single-molecule real-time sequencing can produce reads up to 30 000 nucleotides in length and of relatively high accuracy [9]. At the extreme in long-read sequencing, Oxford Nanopore Technologies nanopore sequencing instruments routinely produce reads that are 10^3 to $>10^5$ nucleotides in length, with rare single reads $>10^6$ nucleotides long reported in certain studies [10, 11]. An additional important feature of nanopore sequencing is speed, with individual reads available for analysis within minutes as they are produced. This compares with a waiting time of 4–60 hours with other approaches, in which reads are available at the end of a sequencing run. A significant challenge with current nanopore technology is that single sequencing reads demonstrate relatively high error rates as compared to other approaches, and the nonrandom nature of the error distribution places limits on correction with simple consensus averaging. An advantage of long reads (both PacBio and Oxford Nanopore) is that they can span repetitive elements that cannot be resolved by shorter reads, allowing complete contiguous closed assembly of chromosomes and plasmids [12, 13]. Additionally, for pathogens that develop substantial population variability during the course of infection, such as human immunodeficiency virus, long-read technologies can permit the phasing of mutations distributed throughout individual genomes [14]. However, as compared with short-read technologies, the long-read approaches currently tend to have lower throughput and higher total costs, limiting their widespread implementation. For these reasons, short-read technologies are more frequently used, with the Illumina platforms being the most popular, followed by the group containing Ion Torrent (Thermo Fisher) and the now phased-out 454 platform (Roche). A number of excellent reviews describe the different available technologies in more detail [7, 15].

Given the aforementioned characteristics, NGS is rapidly finding a variety of applications in both clinical and research microbiology laboratories. Whole-genome sequencing of viral, bacterial, and fungal isolates is now used for high-resolution outbreak investigations [1, 2, 16, 17], pathogen typing or identification of new species [18], and characterization of resistance in slow-growing organisms, such as *Mycobacterium tuberculosis* [4, 19]. Another application of NGS in the clinical microbiology laboratory that is becoming more common is taxonomic identification of unusual cultured clinical isolates, as an alternative to biochemical, mass spectrometry–based, or Sanger sequencing–based identification. Other potential future applications of NGS in the diagnostics laboratory include the genomic characterization of the host immune response during infection [20, 21] and diagnosis and treatment of disease conditions on the basis of changes in the host microbiome [22–24].

The focus of this review will be the application of NGS methods to the analysis of uncultured primary clinical specimens. There are 2 general classes of NGS approaches that can be used for identification of microorganisms in the clinical microbiology laboratory: targeted amplicon sequencing and shotgun sequencing. When applied to primary samples, these approaches have been referred to as "metagenomics" or "mNGS" approaches, although many authors have preferred to reserve these terms for shotgun-based NGS techniques. These 2 approaches are explored individually below.

TARGETED AMPLICON NGS APPROACHES

Targeted amplicon NGS approaches have proven to be a versatile set of methods in both microbiome studies and clinical microbiology. These approaches begin with polymerase chain reaction (PCR) amplification of a region of interest, using appropriate flanking primers. To prepare the product of this reaction for batch sequencing, the initial amplicon is usually converted into to bar-coded libraries. Targeted amplicon NGS–based approaches have been adapted for universal identification and taxonomic classification of bacteria and fungi, using conserved but information-rich regions in the microbial genome. For bacteria, the 16S ribosomal RNA gene is used, and for fungi, the internal transcribed spacer region in the ribosomal gene cluster or the 18S ribosomal RNA gene is ordinarily used, although other targets have been shown to have utility, as well. PCR primers that hybridize to highly conserved sequences within these regions are used to amplify adjacent variable, information-rich segments of sequence that can be used for taxonomic identification. Following sequencing, reads are clustered into groups by sequence similarity, consensus sequences are generated, and taxonomic identification is performed by alignment of consensus sequences to an appropriate reference database [25, 26].

A key advantage of targeted amplicon NGS over shotgun metagenomics is that significantly lower sequencing depths are required, as sequencing is restricted only to the region of interest of a single gene. Furthermore, less complex computational analysis is required owing to the amplification of only one genomic region. On the other hand, targeted amplicon NGS does not provide other genetic information about the pathogen outside the amplified region, such as the presence of virulence or resistance genes, unless these regions are specifically included in the targeted sequencing reaction. In the setting of very low amounts of pathogen genetic material in the sample, ultradeep shotgun approaches may improve sensitivity over targeted approaches, as the targets may not be present, whereas other small fragments of genome are [27, 28]. Targeted

approaches may also be prone to PCR amplification bias, in which preferential amplification of one or more targets may skew the inferred quantitative proportions of taxa identified in a specimen [29]. Finally, targeted amplicon NGS requires a hypothesis about which organism group (bacterial or fungal) is suspected, to ensure that appropriate amplification targets are chosen. Additionally, viruses as a group lack universal amplification targets analogous to the 16S or internal transcribed spacer regions in bacteria and fungi, although some conserved regions exist within viral families.

SHOTGUN NGS METAGENOMICS APPROACHES

Shotgun metagenomics approaches have been discussed by many authors as a potential universal diagnostic test, capable of detecting the presence of a wide range of pathogens, including viruses, bacteria, fungi, archaea, and parasites from a patient sample, without requiring an a priori hypothesis about what is present and with the potential to yield clinically relevant genomic features and possibly high-resolution epidemiological information. While many advances have been achieved toward making this approach a reality, many intrinsic and unique challenges remain (Table 1). To define the hurdles involved in shotgun metagenomic diagnostic methods, a brief overview of the technology is warranted.

Conceptually, this approach involves the extraction and sequencing of total DNA (and/or RNA, followed by reverse transcription) from the primary specimen. Both host and pathogen DNA are sequenced in approximate proportion to their abundance in solution, as well as any other potential sources of DNA, such as normal microbiome, environmental DNA, and DNA present in "sterile" reagents, which can be substantial.

One of the major challenges of shotgun metagenomics is the overwhelming amount of host DNA present in primary clinical specimens. This can be understood from the following considerations of the origins and proportions of different classes of DNA in clinical material. Bacterial cells, with haploid genome sizes typically ranging from 1 million to 5 million base pairs, contain 1/1000th or less the amount of DNA contained in a human cell, whose diploid genome contains on the order of 6 billion base pairs. Viruses contain orders of magnitude less DNA than bacterial cells. In addition to the fact that human cells contain much more DNA than pathogen cells, they are often present in greater proportions, particularly in tissue samples. In a tuberculous granuloma, for instance, there may be hundreds to thousands of histiocytes, lymphocytes, and neutrophils for each mycobacterial cell. Given that shotgun sequencing approaches produce reads in proportion to the relative concentrations of DNA in the sample as noted above, this can result in an extremely small ratio of microbial to human DNA, with a representation of human sequencing reads >99.99% [27, 28]. Several methods exist to enrich for microbial DNA in the sample; however, each method presents its own challenges [30, 31], and while these methods provide incremental enrichment of the sample for microbial DNA, host DNA often still represents >95% of the total. Furthermore, they can substantially reduce the total amount of remaining pathogen DNA before library preparation, making intermediate amplification steps necessary prior to sequencing, which can lead to biases and additional exogenous microbial DNA contamination [32].

One group of microbial DNA enrichment methods takes advantage of the lack of cell wall in human cells for differential cell lysis: an initial step of chemical lysis of human cells is followed by enzymatic removal of suspended DNA and, finally,

Table 1. Features and Challenges of Clinical Shotgun Metagenomics

Desired feature	Challenge(s)	Solutions (Reference[s])
Unbiased detection of all pathogens present in sample	Overwhelming amount of host nucleic acid, contaminant nucleic acid, reagent microbiome	Microbial enrichment methods [30, 31, 33], pathogen-specific enrichment [34–36], internal and external run controls [37, 38], query uniformity of coverage over organism genome [52]
Clinically relevant turnaround	Work flow of several days on most platforms	Use of nanopore technologies (but with limited read depth and low-accuracy reads) [13, 41, 56, 57], preferential use for detection of slow-growing, novel or culture-negative pathogens [27]
Clinical laboratory implementation	Test validation/standardization, expertise and computational resources needed, incidental sequencing of human genome	Standardized mixture of organisms for control [37,38], in silico simulation [54], user-friendly interface and streamlined work flows, dedicated informed consent [54], confidentiality and data protection safeguards
High accuracy for positive calls	Misidentification of reads to similar species, incorrect cross-assignment of ambiguous reads, novel organisms not present in database	Improved microbial genome database curation and completeness [40]
Virulence/resistance information	Assignment of genes located in mobile elements or plasmids to correct organism, insufficient read depth for point mutations, imperfect genotype-phenotype correlation	Proximity deconvolution [63–65], development of more extensive genotype-phenotype databases, application of machine learning techniques
Intuitive result interpretation	Complexity of result data	Multidisciplinary, precision medicine team [80], explicative result reports [80], end-user (clinician) training
Clinical correlation of positive results	Positive test result might not correlate with disease process	Clinical follow-up, prospective case series

microbial lysis and DNA extraction [30, 31]. Other methods use features present in human DNA to achieve selective depletion. For instance, CpG motif methylation can be used as a target for antibodies for differential removal of human DNA. A novel method referred to as depletion of abundant sequences by hybridization, which uses a modified Cas9 enzyme, has been used for removal of highly abundant human sequences [33]. Finally, highly multiplexed sequence capture approaches have recently been proven to provide substantial enrichment for viral or bacterial targets from a highly cellular background by selectively amplifying genetic material bound to pathogen-specific probe sets prior to sequencing [34–36].

CHALLENGES IN CLINICAL LABORATORY IMPLEMENTATION OF SHOTGUN METAGENOMICS

Widespread implementation of shotgun metagenomics in the clinical microbiology laboratory poses additional challenges in terms of validation and quality control [37, 38]. At present, there are no Food and Drug Administration–approved shotgun metagenomic diagnostic tests or standardized regulatory guidelines, and only a small number of centers offer limited metagenomic testing under Clinical and Laboratory Improvement Amendments certificates in a reference capacity. Successful shotgun metagenomic diagnostic approaches in the past have used a variety of techniques for DNA or RNA extraction, library preparation, and sequencing, as well as a number of different bioinformatics pipelines for human read subtraction and phylogenetic assignment, each with advantages and disadvantages [25, 27, 28, 39–53].

While it is tempting to develop a test that can be as versatile as possible in terms of organism detection, successful standardization depends on clearly defining the clinical uses for which the testing is intended. The range of clinical samples anticipated, the desired turnaround time, and the list of reportable pathogens are some factors that influence choice of nucleic acid extraction strategy (eg, DNA and/or RNA), method for host DNA depletion, sequencing depth, and analysis pipeline [37]. For instance, RNA viruses would not be detected with a strategy limited to DNA extraction. Sequencing to greater depths might provide more-detailed taxonomic resolution but at the expense of longer turnaround times, increased computational complexity, and greater cost. As sequencing technologies, bioinformatics pipelines, and both human and microbial reference sequence databases continue to evolve, established mechanisms for reference database updates and curation are also necessary. The complexities of shotgun metagenomics quality control and validation have been reviewed in detail elsewhere [37]. Some key aspects will be briefly discussed here.

Assay validation involves a determination of accuracy (agreement with findings verified by other reference methods), precision (reproducibility and repeatability), analytical sensitivity, and specificity [38]. The intrinsic characteristics of shotgun metagenomics make validation a challenge: How does one assess the characteristics of a test with such a vast potential range and diversity of results? It has been proposed that the accuracy evaluation be done with a combination of pathogen-positive and pathogen-negative specimens, whole-organism or purified nucleic acid, and in silico simulation [37]. As it is not feasible for a validation sample set to contain all classes of pathogens that are theoretically detectable by shotgun metagenomics, positive controls containing subsets of organisms belonging to major classes (eg, RNA and DNA viruses, gram-positive and gram-negative organisms, yeasts, and molds) have been used to provide standardized and reproducible input material for benchmarking given tests. Proficiency testing of bioinformatics pipelines by using sequence sets that have been "mutagenized" in silico to reflect naturally occurring error profiles and sequence variation might emerge as a complementary approach for validation of computational components of testing [54]. Assay sensitivity depends on multiple factors, including the relative representation of host and pathogen in the sample, the pathogen genome size and cell wall composition, and the test design. A practical approach to mitigate these multiple confounding factors is the use of a quantitative internal control that is added at a defined concentration into all specimens analyzed, which can give measures of absolute and relative sensitivity. A final consideration is the unavoidable incidental sequencing of human genetic information as part of the test. This raises additional ethical and logistic considerations, such as the need for additional confidentiality safeguards, information security infrastructure, review of testing plan by institutional review boards, and in many cases, patient informed consent.

SHOTGUN METAGENOMICS: MANAGING CLINICAL EXPECTATIONS

It is important to keep in mind that diagnostic shotgun metagenomics involves time- and resource-intensive, multistep testing procedures, as compared to traditional testing in the clinical microbiology laboratory. A misconception among some clinicians is that shotgun metagenomic testing is faster than currently available microbiologic diagnostic methods. The time to results for shotgun metagenomics is the sum of sample preparation time (DNA extraction plus library preparation), sequencing, and data analysis. The most common procedures have used Illumina technology, for which generating the sequencing data alone can take from 20 to 60 hours. The variety of bioinformatics pipelines available also vary in their computation times, which will depend, in turn, on the computational infrastructure that is available. For these reasons, many published diagnostic applications of shotgun metagenomics report turnaround times of 2–7 days from sample collection to results [55]. Nanopore sequencing technology offers real-time availability of sequence data and holds promise to revolutionize certain types of NGS diagnostic assays, particularly sequencing

of clinical isolates and resistance gene detection [13, 56]. However, nanopore sequencing approaches are still not ideally suited for shotgun metagenomics in primary specimens owing to their relatively low-depth sequencing at higher cost and to their high per-read error rates. That said, certain groups have successfully implemented shotgun metagenomics strategies on the nanopore sequencing platform, and reportable results have been obtained in as little as 6 hours [41, 57].

In comparison to shotgun metagenomics approaches, the work flow from sample collection to identification of common pathogens in the microbiology laboratory can take from a few hours (for nucleic acid amplification tests) to 1–3 days for bacteria that can be cultured on routine media. Antimicrobial susceptibility testing may add an additional 1–2 days. Thus, while shotgun metagenomics may indeed shorten the time to diagnosis in select cases, the greatest potential diagnostic advantages lie in the ability to detect unsuspected, uncharacterized, uncultivatable, or very slow-growing organisms, which produce negative results with standard assays. Rather than a replacement for current testing, shotgun metagenomics is most suited to be used in conjunction with traditional methods. This implies that implementing shotgun metagenomic diagnostic methods in the clinical microbiology laboratory may incur significant costs with little offset from discontinuation of other microbiologic tests. This stands in contrast to many other recent technologies, such as the replacement of automated biochemical approaches for isolate identification with matrix-assisted laser desorption/ionization time of flight mass spectrometry.

INTERPRETIVE CHALLENGES POSED BY THE UBIQUITY OF MICROBIAL DNA CONTAMINATION

With the increasing use of deep-sequencing approaches, it has become apparent that microbial DNA is ubiquitous, not only as cutaneous or other microbiota introduced into biological samples during diagnostic procedures, but also as present on plastics and in many laboratory reagents considered to be sterile and used for the preparation of the sequencing reaction [46, 58–60]. Different measures can be deployed to help with this challenge. Running parallel negative control samples with known DNA composition is often necessary to define the microbial nucleic acid background expected. An a priori knowledge of common reagent contaminants (both reported in the published literature and specific to each laboratory) can be used to generate a list of likely contaminants. This can be supplemented with sequences of actual contaminants identified in parallel negative samples, which can then be computationally subtracted from sequencing results or subjected to higher reporting thresholds, using a variety of methods. Finally, the distribution of sequencing reads over a reference genome can be used to identify likely contaminant DNA. If intact organisms are present, reads are more likely to cover the full genome, while if only genomic fragments of a contaminant are present (for instance, fragments that remain after sterilization procedures), reads may represent a small percentage of the genome. Measures of the standard distribution of read coverage have thus been used for this purpose [52]. However, contaminating DNA in reagents may be present as full genomes, and true pathogens may be present only in fragments, so this approach does not guarantee contaminant separation.

BIOINFORMATICS: METHODS AND RESOURCES

In contrast to other clinical microbiology tests more amenable to automation and less resource intensive, shotgun metagenomics requires bioinformatics expertise for results interpretation. The raw data generated can be processed through several different analysis tools or pipelines, which commonly subtract reads mapped to the human genome as a first step, followed by taxonomic assignment of the remaining reads to the appropriate phylogenetic group by comparing them to a genomic database. Different pipelines use different microbial databases, which in turn have different degrees of curation, accuracy, and completeness, resulting in varying sensitivity and specificity at this level. Use of an improperly curated or configured database can lead to inaccurate results due to assignment of reads to taxonomically misidentified microbial species, incorrect cross-assignment of reads that map ambiguously to shared regions of genomes of similar organisms (eg, *Staphylococcus aureus* vs other coagulase-negative staphylococci), or assignment of unfiltered human sequences or low-complexity sequences to microbial genomes. On the other hand, incomplete databases can result in false-negative results for organisms not included within them. As serious interpretive errors can occur with lack of full understanding of database composition, it is imperative that those implementing laboratory-developed shotgun NGS testing use standardized database resources or use expertise and caution in curation.

CHALLENGES IN CLINICAL INTERPRETATION OF RESULTS

Owing to their high cost and longer turnaround times, shotgun metagenomics tests are currently invoked as diagnostic approaches of last resort, used to solve puzzling cases with fastidious or obscure etiologies, when standard culture and/or molecular testing have failed to find an answer. In view of this, we must consider the unique difficulties of interpretation and independent validation of results from such testing, particularly when results of all other assays are negative [61]. In the case of shotgun metagenomics, where multiple otherwise undetected pathogens might be reported, it becomes critical to develop a clinical correlation for an isolated positive test result and the need for intervention. While these questions are ideally evaluated by randomized clinical trials, such trials might not be feasible for practical or ethical reasons. In such cases,

longitudinal follow-up and clinical correlation are required to define the clinical nature of diagnoses made solely on the basis of shotgun metagenomics findings [62].

For the above reasons, the complexity of results from shotgun metagenomics demands careful technical interpretation by the clinical microbiology laboratory directors before reaching clinicians and the medical record. It is likely that, in addition to information about the pathogens that are present, test reports will need to include other ancillary information, to aid interpretation by treating clinicians. In a similar fashion to interpretation of computed tomography results by end-user treating clinicians, infectious disease physicians with routine exposure to genomic testing might require some degree of specialized training to interpret the results.

MAKING USE OF SHOTGUN METAGENOMICS BEYOND PATHOGEN IDENTIFICATION: VIRULENCE AND RESISTANCE GENE DETECTION

As described above, a potential advantage of shotgun metagenomics over targeted (eg, 16S, 18S, or internal transcribed spacer) approaches lies in the ability to obtain sequencing information from other genomic regions, such as the presence or absence of antimicrobial resistance genes or virulence factors. To do this, a few challenges need to be overcome. First, in the case of bacterial pathogens, genes located on mobile plasmids are usually difficult to assign to a specific organism among of the possibly many detected in given sample, without the use of dedicated techniques. Approaches have been developed to help determine the organism of origin for a given plasmid-associated gene, but not without substantially increasing the complexity and cost of the overall process. One class of methods is based on proximity-ligation approaches [63, 64]. These rely on formaldehyde-induced cross-linking of DNA fragments (chromosome and plasmid) in close physical proximity in the intact cell, followed by ligation of cross-linked molecules, prior to sequencing [65].

A second challenge for resistance profile prediction lies in the often-inconsistent genotype-phenotype correlation that exists between the presence or absence of a given resistance element and (1) the measured in vitro minimum inhibitory concentration for the whole organism grown in the presence of a given antimicrobial or, more importantly, (2) the success or failure of clinical treatment. The presence of certain genes, such as those encoding carbapenemases, has been shown to correlate fairly well with in vitro resistance [66]. However, for other resistance genes, such as aminoglycoside-modifying enzymes, the degree of resistance to a given aminoglycoside often cannot easily be inferred solely from the presence of single genes. Last, overexpression of certain resistance genes as a consequence of intergenic promoter mutations can result in significantly different resistance phenotypes that may be difficult to infer from the genomic sequence alone. Shotgun methods based on RNA sequencing may be used to characterize gene expression [67].

On the other hand, while it would seem that susceptibility to a given antimicrobial might be easily inferred from the absence of any detected known resistance mechanism, matters are often not straightforward. Evidence points to the existence of a large number of resistance mechanisms that have not been genetically characterized, in particular for newer antimicrobials and those in uncommon pathogens. In addition, a certain minimum amount of sequencing coverage is required to declare the absence of a given gene. In cases in which there is low genomic coverage of a particular pathogen, it is possible that a resistance element might be present but simply not sequenced. This is especially true for point mutation–driven resistance, in which coverage must be sufficient for single-nucleotide variants to be determined with confidence.

ENTERING THE CLINICAL GROUNDS: SUCCESS STORIES AND ATTEMPTS AT IMPLEMENTATION

In recent years, a growing number of case reports and case series have emerged in the literature that demonstrate the power of shotgun metagenomic diagnostic methods. While by no means a comprehensive list, we describe below a representative set of reports. A large number of shotgun metagenomic diagnostic reports have studied central nervous system infection, likely in part because of the diagnostic challenges posed by cases of unexplained encephalitis, as well as because of the favorable characteristics of cerebrospinal fluid, particularly its low cellularity. Two of the earliest reports, by Quan et al [68] and Wilson et al [27], described the use of shotgun metagenomics in cerebrospinal fluid for the diagnosis of a novel astrovirus and a case of neuroleptospirosis, respectively. Since then, many other diagnoses have been made, including *Balamuthia mandrillaris* primary amoebic meningoencephalitis [40]; astrovirus progressive encephalitis [48]; infections caused by *Taenia solium, Aspergillus oryzae, Cryptococcus neoformans,* and *Candida dubliniensis* [53]; and neurobrucellosis [28]. Novel pathogens have been described as well, including variegated squirrel bornavirus causing fatal infection in a group of German squirrel fanciers [42] and Cache Valley virus [69], among others [49].

Successful approaches have also been reported from blood [50, 70–72], vitreal fluid [73], corneal tissue [74, 75], bile [55], and respiratory [45, 57] samples. A large collection of prosthetic joint fluids and explanted prosthetic joint sonicate material were analyzed by shotgun metagenomics in a series of detailed studies by Thoendel et al and Ivy et al [43, 51, 52], who reported that, compared with paired culture, shotgun metagenomics identified additional organisms in 8% of culture-positive cases and found organisms in 31% of culture-negative cases, including a novel joint infection agent, *Mycoplasma salivarium*.

SHOTGUN METAGENOMICS DIAGNOSTIC METHODS: PRESENT ROLE AND FUTURE OUTLOOK

With decreasing sequencing costs, the ever-growing number and diversity of pathogen sequences available, more-accurately curated databases, and more–user-friendly bioinformatics tools, the future looks bright in the terrain of shotgun metagenomics applications. Much work is being done to standardize shotgun metagenomics [37, 46, 76], and when achieved, it is likely that the barriers to a more extensive implementation in clinical laboratories will decrease.

What exactly the future clinical niche of shotgun metagenomics will be, however, remains to be seen. The current main competitors are commercial rapid multiplex PCR "syndromic" panels, in use in a large number of hospitals [77–79]. These panels have turnaround times of <2 hours, require little training and expertise to operate and interpret, and have significantly lower costs than current shotgun sequencing testing. While current multiplex PCR panels detect defined sets of 10–25 syndrome- or source-specific pathogens, they usually represent the most commonly encountered clinically relevant pathogens. These reasons make it unlikely for shotgun metagenomics to find a place in the arena of syndromic diagnostic methods. At the time of this writing, the roles that infectious diseases clinicians seek from shotgun metagenomics lie primarily in diagnosing cases in which an infectious etiology is strongly suspected but other available tests fail to identify a specific pathogen [47, 52]. The most-valuable results would identify treatable pathogens not easily covered empirically, resulting in patient management changes and improving outcomes.

As we have attempted to illustrate in this review, there is both excitement and challenge on the horizon for routine adoption of NGS metagenomics as a diagnostic approach in the infectious disease field. With the increasing presence of these methods in clinical practice, it will become progressively more important for physicians to be aware of the characteristics and potential uses of NGS metagenomics, as well as challenges in results interpretation, particularly as physicians begin to incorporate results of these tests into their approach to diagnostic workup.

Supplementary Data

Supplementary materials are available at *The Journal of 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.

Notes

Disclaimer. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health, the Department of Health and Human Services, or the US government.

Financial support. This work was supported by the Intramural Research Program of the National Institute of Allergy and Infectious Diseases.

Supplement sponsorship. This supplement is sponsored by WRAIR, LANL, USAMRIID, PUCP (Pontificia Universidad Catolica del Peru), USAFSAM, NIH.

Potential conflicts of interest. Both authors: No reported conflicts of interest. Both 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. Snitkin ES, Zelazny AM, Thomas PJ, et al.; NISC Comparative Sequencing Program Group. Tracking a hospital outbreak of carbapenem-resistant Klebsiella pneumoniae with whole-genome sequencing. Sci Transl Med **2012**; 4:148ra116.
2. Quick J, Loman NJ, Duraffour S, et al. Real-time, portable genome sequencing for Ebola surveillance. Nature **2016**; 530:228–32.
3. Paterson GK, Harrison EM, Murray GG, et al. Capturing the cloud of diversity reveals complexity and heterogeneity of MRSA carriage, infection and transmission. Nat Commun **2015**; 6:6560.
4. Gröschel MI, Walker TM, van der Werf TS, Lange C, Niemann S, Merker M. Pathogen-based precision medicine for drug-resistant tuberculosis. PLoS Pathog **2018**; 14:e1007297.
5. Schmidt K, Mwaigwisya S, Crossman LC, et al. Identification of bacterial pathogens and antimicrobial resistance directly from clinical urines by nanopore-based metagenomic sequencing. J Antimicrob Chemother **2017**; 72:104–14.
6. McGinn S, Gut IG. DNA sequencing–spanning the generations. N Biotechnol **2013**; 30:366–72.
7. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet **2016**; 17:333–51.
8. Loman NJ, Pallen MJ. Twenty years of bacterial genome sequencing. Nat Rev Microbiol **2015**; 13:787–94.
9. SMRT sequencing: read lengths. https://www.pacb.com/smrt-science/smrt-sequencing/read-lengths/. Accessed 18 March 2019.
10. Jain M, Koren S, Miga KH, et al. Nanopore sequencing and assembly of a human genome with ultra-long reads. Nat Biotechnol **2018**; 36:338–45.
11. Payne A, Holmes N, Rakyan V, Loose M. BulkVis: a graphical viewer for Oxford nanopore bulk FAST5 files. Bioinformatics **2018**; Nov 20:1–6. https://academic.oup.com/bioinformatics/advance-article/doi/10.1093/bioinformatics/bty841/5193712

12. George S, Pankhurst L, Hubbard A, et al. Resolving plasmid structures in *Enterobacteriaceae* using the MinION nanopore sequencer: assessment of MinION and MinION/Illumina hybrid data assembly approaches. Microb Genom **2017**; 3:e000118.
13. Lemon JK, Khil PP, Frank KM, Dekker JP. Rapid nanopore sequencing of plasmids and resistance gene detection in clinical isolates. J Clin Microbiol **2017**; 55:3530–43.
14. Dilernia DA, Chien JT, Monaco DC, et al. Multiplexed highly-accurate DNA sequencing of closely-related HIV-1 variants using continuous long reads from single molecule, real-time sequencing. Nucleic Acids Res **2015**; 43:e129.
15. 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.
16. Lockhart SR, Etienne KA, Vallabhaneni S, et al. Simultaneous emergence of multidrug-resistant Candida auris on 3 continents confirmed by whole-genome sequencing and epidemiological analyses. Clin Infect Dis **2017**; 64:134–40.
17. Rames E, Macdonald J. Evaluation of MinION nanopore sequencing for rapid enterovirus genotyping. Virus Res **2018**; 252:8–12.
18. Tortoli E. Microbiological features and clinical relevance of new species of the genus Mycobacterium. Clin Microbiol Rev **2014**; 27:727–52.
19. Shea J, Halse TA, Lapierre P, et al. Comprehensive whole-genome sequencing and reporting of drug resistance profiles on clinical cases of *Mycobacterium tuberculosis* in New York State. J Clin Microbiol **2017**; 55:1871–82.
20. Langelier C, Kalantar KL, Moazed F, et al. Integrating host response and unbiased microbe detection for lower respiratory tract infection diagnosis in critically ill adults. Proc Natl Acad Sci U S A **2018**; 115:E12353–62.
21. Bal A, Sarkozy C, Josset L, et al. Metagenomic next-generation sequencing reveals individual composition and dynamics of anelloviruses during autologous stem cell transplant recipient management. Viruses **2018**; 10.
22. Zhao J, Schloss PD, Kalikin LM, et al. Decade-long bacterial community dynamics in cystic fibrosis airways. Proc Natl Acad Sci U S A **2012**; 109:5809–14.
23. Morgan XC, Tickle TL, Sokol H, et al. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. Genome Biol **2012**; 13:R79.
24. van Nood E, Vrieze A, Nieuwdorp M, et al. Duodenal infusion of donor feces for recurrent *Clostridium difficile*. N Engl J Med **2013**; 368:407–15.
25. Cummings LA, Kurosawa K, Hoogestraat DR, et al. Clinical next generation sequencing outperforms standard microbiological culture for characterizing polymicrobial samples. Clin Chem **2016**; 62:1465–73.
26. Schoch CL, Seifert KA, Huhndorf S, et al.; Fungal Barcoding Consortium; Fungal Barcoding Consortium Author List. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. Proc Natl Acad Sci U S A **2012**; 109:6241–6.
27. Wilson MR, Naccache SN, Samayoa E, et al. Actionable diagnosis of neuroleptospirosis by next-generation sequencing. N Engl J Med **2014**; 370:2408–17.
28. Mongkolrattanothai K, Naccache SN, Bender JM, et al. Neurobrucellosis: unexpected answer from metagenomic next-generation sequencing. J Pediatric Infect Dis Soc **2017**; 6:393–8.
29. Kebschull JM, Zador AM. Sources of PCR-induced distortions in high-throughput sequencing data sets. Nucleic Acids Res **2015**; 43:e143.
30. Thoendel M, Jeraldo PR, Greenwood-Quaintance KE, et al. Comparison of microbial DNA enrichment tools for metagenomic whole genome sequencing. J Microbiol Methods **2016**; 127:141–5.
31. 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.
32. Probst AJ, Weinmaier T, DeSantis TZ, Santo Domingo JW, Ashbolt N. New perspectives on microbial community distortion after whole-genome amplification. PLoS One **2015**; 10:e0124158.
33. Gu W, Crawford ED, O'Donovan BD, et al. Depletion of Abundant Sequences by Hybridization (DASH): using Cas9 to remove unwanted high-abundance species in sequencing libraries and molecular counting applications. Genome Biol **2016**; 17:41.
34. Allicock OM, Guo C, Uhlemann AC, et al. BacCapSeq: a platform for diagnosis and characterization of bacterial infections. MBio **2018**; 9:e02007-18.
35. Wylie KM, Wylie TN, Buller R, Herter B, Cannella MT, Storch GA. Detection of viruses in clinical samples by use of metagenomic sequencing and targeted sequence capture. J Clin Microbiol **2018**; 56:e01123-18.
36. O'Flaherty BM, Li Y, Tao Y, et al. Comprehensive viral enrichment enables sensitive respiratory virus genomic identification and analysis by next generation sequencing. Genome Res **2018**; 28:869–77.
37. 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.
38. Lefterova MI, Suarez CJ, Banaei N, Pinsky BA. Next-generation sequencing for infectious disease diagnosis and

management: a report of the association for molecular pathology. J Mol Diagn **2015**; 17:623–34.
39. Briese T, Kapoor A, Mishra N, et al. Virome capture sequencing enables sensitive viral diagnosis and comprehensive virome analysis. MBio **2015**; 6:e01491–15.
40. 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.
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. 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.
43. Ivy MI, Thoendel MJ, Jeraldo PR, et al. Direct detection and identification of prosthetic joint infection pathogens in synovial fluid by metagenomic shotgun sequencing. J Clin Microbiol **2018**; 56:e00402-18.
44. Kennedy PGE, Quan PL, Lipkin WI. Viral encephalitis of unknown cause: current perspective and recent advances. Viruses **2017**; 9:138.
45. Langelier C, Zinter MS, Kalantar K, et al. Metagenomic sequencing detects respiratory pathogens in hematopoietic cellular transplant patients. Am J Respir Crit Care Med **2018**; 197:524–8.
46. Miller S, Naccache SN, Chiu C. Laboratory validation of a clinical metagenomic sequencing assay for pathogen detection in cerebrospinal fluid. **2018**. doi: 10.1101/gr.238170.118
47. Naccache SN, Greninger A, Samayoa E, Miller S, Chiu CY. Clinical utility of unbiased metagenomic next-generation sequencing in diagnosis of acute infectious diseases: a prospective case series. Open Forum Infect Dis **2015**; 2:103.
48. 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.
49. 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.
50. Somasekar S, Lee D, Rule J, et al. Viral surveillance in serum samples from patients with acute liver failure by metagenomic next-generation sequencing. Clin Infect Dis **2017**; 65:1477–85.
51. Thoendel M, Jeraldo P, Greenwood-Quaintance KE, et al. A novel prosthetic joint infection pathogen, mycoplasma salivarium, identified by metagenomic shotgun sequencing. Clin Infect Dis **2017**; 65:332–5.
52. Thoendel MJ, Jeraldo PR, Greenwood-Quaintance KE, et al. Identification of prosthetic joint infection pathogens using a shotgun metagenomics approach. Clin Infect Dis **2018**; 67:1333–8.
53. Wilson MR, O'Donovan BD, Gelfand JM, et al. Chronic meningitis investigated via metagenomic next-generation sequencing. JAMA Neurol **2018**; 75:947–55.
54. Duncavage EJ, Abel HJ, Merker JD, et al. A model study of in silico proficiency testing for clinical next-generation sequencing. Arch Pathol Lab Med **2016**; 140:1085–91.
55. 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.
56. Cao MD, Ganesamoorthy D, Elliott AG, Zhang H, Cooper MA, Coin LJ. Streaming algorithms for identification of pathogens and antibiotic resistance potential from real-time MinION™ sequencing. Gigascience **2016**; 5:32.
57. 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**; 196:1610–2.
58. de Goffau MC, Lager S, Salter SJ, et al. Recognizing the reagent microbiome. Nat Microbiol **2018**; 3:851–3.
59. Salter SJ, Cox MJ, Turek EM, et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biol **2014**; 12:87.
60. Thoendel M, Jeraldo P, Greenwood-Quaintance KE, et al. Impact of contaminating DNA in whole-genome amplification kits used for metagenomic shotgun sequencing for infection diagnosis. J Clin Microbiol **2017**; 55:1789–801.
61. Glasziou P, Irwig L, Deeks JJ. When should a new test become the current reference standard? Ann Intern Med **2008**; 149:816–22.
62. Lord SJ, Irwig L, Simes RJ. When is measuring sensitivity and specificity sufficient to evaluate a diagnostic test, and when do we need randomized trials? Ann Intern Med **2006**; 144:850–5.
63. Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. Science **2002**; 295:1306–11.
64. Burton JN, Liachko I, Dunham MJ, Shendure J. Species-level deconvolution of metagenome assemblies with Hi-C-based contact probability maps. G3 (Bethesda) **2014**; 4:1339–46.
65. Beitel CW, Froenicke L, Lang JM, et al. Strain- and plasmid-level deconvolution of a synthetic metagenome by sequencing proximity ligation products. PeerJ **2014**; 2:e415.
66. McMullen AR, Yarbrough ML, Wallace MA, Shupe A, Burnham CD. Evaluation of genotypic and phenotypic methods to detect carbapenemase production in gram-negative bacilli. Clin Chem **2017**; 63:723–30.
67. Versluis D, D'Andrea MM, Ramiro Garcia J, et al. Mining microbial metatranscriptomes for expression of antibiotic

resistance genes under natural conditions. Sci Rep **2015**; 5:11981.
68. Quan PL, Wagner TA, Briese T, et al. Astrovirus encephalitis in boy with X-linked agammaglobulinemia. Emerg Infect Dis **2010**; 16:918–25.
69. Wilson MR, Suan D, Duggins A, et al. A novel cause of chronic viral meningoencephalitis: Cache Valley virus. Ann Neurol **2017**; 82:105–14.
70. 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.
71. 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.
72. Grard G, Fair JN, Lee D, et al. A novel rhabdovirus associated with acute hemorrhagic fever in central Africa. PLoS Pathog **2012**; 8:e1002924.
73. Doan T, Wilson MR, Crawford ED, et al. Illuminating uveitis: metagenomic deep sequencing identifies common and rare pathogens. Genome Med **2016**; 8:90.
74. Shigeyasu C, Yamada M, Aoki K, et al. Metagenomic analysis for detecting Fusarium solani in a case of fungal keratitis. J Infect Chemother **2018**; 24:664–8.
75. Seitzman GD, Thulasi P, Hinterwirth A, Chen C, Shantha J, Doan T. Capnocytophaga keratitis: clinical presentation and use of metagenomic deep sequencing for diagnosis. Cornea **2019**; 38:246–8.
76. Simner PJ, Miller HB, Breitwieser FP, et al. Development and optimization of metagenomic next-generation sequencing methods for cerebrospinal fluid diagnostics. J Clin Microbiol **2018**; 56.
77. Babady NE. The FilmArray® respiratory panel: an automated, broadly multiplexed molecular test for the rapid and accurate detection of respiratory pathogens. Expert Rev Mol Diagn **2013**; 13:779–88.
78. Buss SN, Leber A, Chapin K, et al. Multicenter evaluation of the BioFire FilmArray gastrointestinal panel for etiologic diagnosis of infectious gastroenteritis. J Clin Microbiol **2015**; 53:915–25.
79. 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.
80. Simner PJ, Miller S, Carroll KC. Understanding the promises and hurdles of metagenomic next-generation sequencing as a diagnostic tool for infectious diseases. Clin Infect Dis **2018**; 66:778–88.