



Metagenomics for Clinical Infectious Disease Diagnostics Steps Closer to Reality

John P. Dekker^a

^aDepartment of Laboratory Medicine, Clinical Center, National Institutes of Health, Bethesda, Maryland, USA

ABSTRACT Metagenomics approaches based on shotgun next-generation sequencing hold promise for infectious disease diagnostics. Despite substantial challenges that remain, work done over the past few years justifies excitement about the potential for these approaches to transform how clinical pathogen identification and analysis are performed. In an article in this issue of the *Journal of Clinical Microbiology*, M. I. Ivy et al. (J Clin Microbiol 56:e00402-18, 2018, <https://doi.org/10.1128/JCM.00402-18>) have applied a shotgun metagenomics approach to the diagnosis of prosthetic joint infections directly from synovial fluid. The results from this work demonstrate both the potentials and challenges of this approach applied in the clinical microbiology laboratory.

Affordable, high-throughput next-generation sequencing (NGS) approaches have unquestionably transformed biological research and have led to fundamental advances in the understanding of microbial life. These methods have also contributed immeasurably to the understanding of human pathogens, outbreaks, and infectious disease processes. Many in the field of clinical microbiology have long anticipated the application of NGS approaches to routine infectious disease diagnostics, but the substantial complexity involved in sequence analysis and lack of FDA-cleared test systems have presented formidable challenges to implementation of NGS-based methods in most clinical labs (1). Despite these challenges, work done over the past few years justifies excitement about the potential for NGS-based techniques to transform how clinical pathogen identification and analysis are performed (2–10).

Approaches to NGS-based microbial identification in primary specimens are generally divided into those based on sequencing of PCR-amplified targets and those based on shotgun sequencing (1). Both approaches have been referred to as “metagenomics” methods in the literature though many authors prefer to reserve this term for shotgun approaches. These two classes of approach have different domains of application and provide different kinds of information, as discussed in greater depth below. Both approaches, however, contend with some of the same challenges when applied to primary specimens, where diagnostic microbial genomic content is often present at relatively low concentrations. The first challenge is ubiquitous microbial DNA contamination present in reagents, on instrument surfaces, and in the environment (11). The second is the usually overwhelming amount of human host DNA present in primary specimens, derived from neutrophils and other human cell types (3, 7, 12). As culture-based methods are blind to much of the DNA that dominates NGS-based approaches, clinical microbiologists have only recently had to grapple seriously with the difficult problems associated with sorting meaningful sequencing signals from the mix of bacterial and other DNA present in all sequencing reaction mixtures.

Targeted amplicon sequencing approaches have been used routinely in the microbiome field for more than a decade and have certain advantages over shotgun methods for microbial identification in mixtures where microbial DNA is a small fraction of total DNA. These approaches are ordinarily based on amplification of well-characterized sequences

Accepted manuscript posted online 5 July 2018

Citation Dekker JP. 2018. Metagenomics for clinical infectious disease diagnostics steps closer to reality. J Clin Microbiol 56:e00850-18. <https://doi.org/10.1128/JCM.00850-18>.

Editor Karen C. Carroll, Johns Hopkins University School of Medicine

Copyright © 2018 American Society for Microbiology. All Rights Reserved.

Address correspondence to john.dekker@nih.gov.

For the article discussed, see <https://doi.org/10.1128/JCM.00402-18>.

The views expressed in this Commentary do not necessarily reflect the views of the journal or of ASM.

that can be used for taxonomic classification, including the 16S rRNA gene for broad-range bacterial identification and the internal transcribed spacer (ITS) region of the ribosomal gene cluster for broad-range fungal identification (1, 10, 13). Following PCR, amplicons are usually converted into barcoded libraries and sequenced with an NGS platform. In most approaches, the resulting sequencing reads are computationally clustered by sequence similarity with generation of a consensus for each cluster, followed by alignment to a reference database. These consensus sequences and the target(s) to which they align in the database usually contain information adequate for the differentiation of medically relevant bacteria and fungi (1, 13). Since targeted approaches work by selective amplification and sequencing of high-information-content regions of microbial genomes (without amplification of background human DNA), they generally allow taxonomic identification at a significantly lower depth of sequencing than shotgun-based approaches, with typical amplicon sequencing methods working with >100-fold-lower total read counts in primary specimens (3, 5–7, 10, 14). This allows sequencing on lower-throughput NGS instruments, such as the Illumina MiSeq, and usually less complex computational analysis.

In contrast, shotgun metagenomics methods involve extraction of total DNA and/or RNA (usually followed by conversion to DNA) from primary specimens, fragmentation, library preparation, and depth sequencing. A primary challenge with shotgun approaches is the overwhelming concentration of human host DNA as noted above. A typical human cell contains 1×10^3 to 5×10^3 times as much DNA as a bacterial cell and can contain up to 10^6 times as much DNA as a small virus. Furthermore, in abscess fluid or tissue biopsy specimens, microbial pathogens may be present in a numerically small proportion relative to neutrophils or other host cell types, particularly following antibiotic treatment. The product of these two factors can result in a ratio of microbial-to-human DNA that is vanishingly small. Since shotgun metagenomics approaches sequence DNA fragments approximately in proportion to their relative abundance in solution, this can result in host DNA representation of greater than 99.99% of total sequencing reads in some cases (3, 7).

A variety of methods have been developed for enriching for microbial DNA, with popular approaches based on differential lysis of human cells and chemical removal of human DNA (12). Though these methods may reduce human DNA concentration, host DNA is still usually the dominant library component, and sequencing to depths of tens of millions of reads or more is often required to get to rare microbial sequences. Human sequences are subtracted from sequencing output using a variety of methods prior to analysis to reduce downstream computational complexity. In this context, it should be noted that target amplification methods help to some extent with this problem within a certain range of microbial/host DNA proportions. However, in the limits of large relative concentrations of host DNA, targeted methods are affected as well, as most PCRs will begin to fail when the amplification target concentration drops below a certain percentage of total DNA in the reaction mixture, for example, in the range of a couple of copies of target in a background of 10^7 -fold-greater concentration of nontarget DNA.

Analysis of shotgun sequencing data is in many ways computationally more complex than analysis of targeted amplicon data. The input to shotgun metagenomics pipelines may consist of raw reads from a FASTQ file (direct unassembled output from the sequencing instrument), assembled contigs (draft genome output from an assembly pipeline), or translated open reading frames constructed from draft assemblies. Most approaches use one input or a combination of these inputs for alignment to reference databases (2, 15–20). Some of the common pipelines are distinguished by the statistical details of how they parse sequences that map ambiguously to shared regions of closely related genomes. The quality of the pathogen database used for alignment is also a critical determinant of differences in output between methods. A number of popular computational pipelines for analysis have been developed in recent years and include a mix of command-line tools and those with graphical user interfaces, as well as an array of commercial products (2, 15–20).

Shotgun approaches have been used for pathogen discovery and particularly for identifying pathogens for which universal target sequences are not available, such as viruses (4, 5, 8, 9). In certain cases, shotgun sequencing also allows recovery and assembly of whole genomes for bacteria and sometimes for fungi and parasites, and this affords advantages over standard 16S rRNA- and ITS-based targeted approaches. In theory, shotgun genome coverage can permit prediction of antimicrobial resistance, identification of virulence genes, high-resolution strain-typing approaches, and the study of noncultivable organisms (21–23).

Given the remarkable developments in NGS approaches outlined above, there has been sustained excitement about applications of these methods in the clinical lab, but only a small number of studies have examined implementation in real-world microbiology laboratories. In a timely and important study published in this issue of the *Journal of Clinical Microbiology*, Ivy and colleagues apply a shotgun metagenomics approach to the diagnosis of prosthetic joint infections (PJIs) (24). PJI represents a problem that may be particularly suited to metagenomics analysis, given the relative ease with which synovial fluid can be collected from a joint and the sometimes problematically poor yield of traditional culture-based methods for these specimens (25). The approaches used in this study are adopted from work recently published by members of the same group, which also deals with application of metagenomics to PJI (26, 27).

To reduce introduced environmental background contamination, synovial fluid samples were collected in vials pretreated with gamma irradiation, and processing was performed in a laminar flow hood cleaned with bleach before each sample was manipulated. To enrich for microbial DNA and reduce human DNA background, 1 ml of synovial fluid underwent treatment with the MoYsis Basic5 kit (Molzym, Bremen, Germany). This approach works by differential lysis of human cells, followed by degradation of released DNA, and has been demonstrated to yield significant enrichment for bacterial DNA for metagenomics sequencing (12). The remaining cells were pelleted and washed, and DNA was extracted using an MoBio Bacteremia DNA isolation kit (Qiagen, Hilden, Germany). To amplify DNA recovered following extraction, whole-genome amplification was performed using a Qiagen REPLI-g Single Cell whole-genome amplification (WGA) kit (Qiagen, Hilden, Germany), a multiple-displacement amplification approach. It should be noted that nonlinearity in sequence amplification with this technique can result in uneven representation of different regions of amplified genomes and may introduce and amplify additional background contamination as previously demonstrated (28). The consequences can be spurious amplification of components of mixed background contamination into proportionately large signals, particularly when low-level contamination is the dominant DNA component (true negative samples). Following REPLI-g amplification, DNA was purified, and sequencing libraries were prepared with a NEBNext Ultra DNA Library Prep kit (New England BioLabs, Ipswich, MA). Sequencing was performed on a HiSeq 2500 instrument to an average of 30 million paired-end reads per sample.

The bioinformatics pipeline used to analyze resulting sequencing data was constructed from public-domain tools and is available for download (https://github.com/pjeraldo/methods_pji_metagenomics) (26). Individual components are given below along with URLs. Adaptor trimming was accomplished with Trimmomatic (<https://github.com/timflutre/trimmomatic>), and BioBloom Tools (<https://github.com/bcgsc/biobloom>) was used to filter human reads. Two different popular analysis tools were used in parallel to analyze the sequences: Livermore Metagenomics Analysis Toolkit, version 1.2.6 (LMAT) (<https://computation.llnl.gov/projects/livermore-metagenomics-analysis-toolkit>), and MetaPhlan2 (<https://bitbucket.org/biobakery/metaphlan2>).

In building their computational analysis method, the authors take a sophisticated approach to dealing with the problem of ubiquitous reagent and environmental contamination mentioned above. To generate a list of known contaminants, negative controls (Tris-EDTA [TE] buffer without specimen) were sequenced in parallel, and any genus with at least 1,000 attributable reads was listed as a known contaminant and subject to differential thresholding. The genera compiled in this list included *Acineto-*

bacter, *Alishewanella*, *Ralstonia*, *Anaerococcus*, *Haemophilus*, *Malassezia*, *Enhydrobacter*, *Sphingomonas*, *Paenibacillus*, *Delftia*, *Corynebacterium*, *Cutibacterium* (*Propionibacterium*), and *Streptococcus*. The authors added *Bradyrhizobium* to the list based on multiple literature reports indicating its common role as a contaminant. They also removed *Streptococcus*, due to the difficulty of recovering the genome of this pathogen with their methods, which could result in true positives not meeting differential thresholds applied to listed contaminants. A series of computational filters were developed to distinguish cases in which species of these known contaminant genera were present in specimens as potential pathogens at higher read counts.

Analysis with LMAT was performed using two different approaches in parallel, one that made identifications at the genus level and another that made identifications at the species level. Paired-end reads were merged, assigned to taxonomic groups, and then subjected to a number of criteria prior to acceptance. Rules were applied at the levels of absolute number of reads, proportion of all bacterial reads, and fraction of genome covered by identified reads. For this purpose, genome alignment was performed with the Burrows-Wheeler Aligner (BWA) (<https://sourceforge.net/projects/bio-bwa/files/>), and coverage was calculated with BMap (<https://sourceforge.net/projects/bbmap/>). A genus-level identification required that the absolute number or proportion of genus-specific reads met a threshold value, and these values were higher for species of contaminant genera identified above. Parallel species-level identification was similar but included additionally a threshold for reference genome coverage. In parallel with LMAT analysis, sequencing output was analyzed with MetaPhlan2, which works by a different marker-gene-based approach.

To test their method, the authors studied 168 previously frozen synovial fluid samples collected between 1998 and 2017 from patients with clinical PJIs (positive or negative by culture) or aseptic implant failures. Clinical classification for PJI was based on a combination of the guidelines of the Infectious Diseases Society of America (IDSA) and blinded review of available clinical data. In total, 107 samples were classified as PJI by either IDSA guidelines or blinded clinical and laboratory test review, and 61 were classified as aseptic failure. The comparator was synovial fluid culture, performed by standard methods in the authors' clinical lab.

Of the 107 synovial fluids from patients classified as having PJI, 82 were culture positive, and 25 were culture negative. The pathogen(s) identified by culture was also detected by the metagenomics approach in 68/82 (82.9%) of cases, and metagenomics reported identical findings with culture in 67/82 of these cases (81.7%). In two cases of culture-positive PJI, the pathogen identified by metagenomics did not match that identified in culture. In addition to the matching pathogens, metagenomics identified additional putative pathogens not recovered in culture in the culture-positive PJI cases. The organisms identified by metagenomics but not by culture in these cases included, importantly, *S. aureus*, as well as *Salpingoeca rosetta* (a marine eukaryote), *Afpia broomeae*, and *Bradyrhizobium japonicum*, organisms for which there is less clear evidence for involvement in clinical prosthetic joint infections. The organisms detected by culture that were not detected by the species-level metagenomics approach (14 cases) included *S. aureus*, *Staphylococcus lugdunensis*, *Serratia marcescens*, and *Candida albicans*, as well as nine cases of *Staphylococcus epidermidis*. Notably, in five of these cases, the pathogen identified by culture was represented by reads in the sequencing data set but were excluded due to failure to meet either the genome-coverage or proportion threshold.

Metagenomics analysis of the 25 culture-negative cases of PJI resulted in identifications meeting thresholds in 4/25 (16%) of cases. These included *Enterococcus faecalis* and *Finegoldia magna* that were confirmed by culture of paired joint material (other than synovial fluid), and, importantly, *S. aureus*. The shotgun metagenomics approach also identified *Salpingoeca rosetta* in two specimens and *Anaerococcus vaginalis*, representing a genus on the contaminating DNA list and subject to differential thresholding.

Metagenomics analysis of synovial fluids from the 61 cases ruled as aseptic failure resulted in identifications meeting thresholds in 4/61 (6.6%) of cases. These included *S.*

aureus and also included *Dolosigranulum pigrum*, *Acinetobacter junii*, and *Acinetobacter johnsonii*. The authors note that the non-*S. aureus* identifications in these cases may represent contamination even though empirical read thresholds were met. Notably, *Acinetobacter* was the genus represented with the greatest read count on the reagent/environmental contamination list. The authors additionally noted a large number of reads representing contaminating genera in the aseptic failure specimens that were filtered by the thresholding rules, possibly a consequence of the overall lower DNA content and the relatively greater proportion that contaminating DNA made up of total DNA.

The results of this study generally compare with those of other recent work. In the previously mentioned study by Thoendel and colleagues, sonicate fluid cultures (as opposed to synovial fluid cultures) from PJI and aseptic failure cases were analyzed (26). Shotgun metagenomics was able to identify cultured pathogens in 109/115 (94.8%) of culture-positive PJIs, with additional pathogens not detected in culture in 11/115 (9.6%) in these cases. Potential pathogens were identified in a remarkable 43/98 (43.9%) of culture-negative PJIs, in contrast with identification of potential pathogens in 7/195 (3.6%) cases of aseptic failure.

Tarabachi and colleagues used a targeted 16S rRNA amplicon approach to study material from primary and revision arthroplasties (29). Identifications were reported by amplicon sequencing in 25/28 (89.3%) cases judged to be infected clinically, whereas cultures were positive in only 17/25 (60.7%) of these cases. The amplicon sequencing identification was concordant with culture in 15/17 cases. Amplicon sequencing also reported an identification in 9/36 (25%) aseptic revision arthroplasty specimens judged aseptic clinically (29). In a third study, Street and colleagues compared a shotgun metagenomics sequencing approach to culture for PJI and reported a genus-level sensitivity of 64/69 (93%) and calculated a species-level specificity of 85/97 (88%) (30).

For a variety of reasons, discussed above, the sensitivity of moderate-depth metagenomics sequencing of primary specimens may be less than that of culture, at least for the detection of cultivatable organisms. Obtaining the average ~30 million paired-end read depth in this study required sequencing with an Illumina HiSeq. While sequencing technologies will continue to provide greater throughput with smaller and less expensive instruments, this depth of sequencing for a single specimen falls outside the range of what can be reliably obtained in a single Illumina MiSeq run, a relevant comparison in this context. Sequencing (multiplexed at six samples per lane in the Illumina HiSeq 2500) in this study was less expensive than that using single Illumina MiSeq runs but is still associated with a cost of several hundred dollars per sample. This represents a materials cost that is substantially greater than that of culture, without taking into account the cost of bioinformatics expertise.

A significant potential value of sequence-based approaches is that of detecting true pathogens that may not have been recovered in culture due to impaired viability or unusual growth requirements. In this study, potentially important pathogens, including *S. aureus* and *Enterococcus*, were detected by metagenomics but not by culture in the PJI and aseptic failure cases. However, the significance of a few of the organisms found in this category may be questioned. This included three cases in which metagenomics identified *Salpingoeca rosetta*, a eukaryotic marine choanoflagellate that would appear more likely to represent reagent DNA contamination than a PJI pathogen, and genera included the reagent/environmental contamination list (*Acinetobacter junii*, *Acinetobacter johnsonii*, and *Anaerococcus vaginalis*). This demonstrates the difficulty of interpreting some sequencing results that are not supported by culture. In other work, metagenomics has revealed unusual or unexpected causes of disease (3, 7, 8, 27), but this difficulty in interpretation of culture-negative results will be a feature of any metagenomics approach, and this detailed study demonstrates the importance of caution in interpreting such results.

There are a number of regulatory considerations that any clinical lab would need to consider before rushing to implement a shotgun metagenomics approach such as the one presented in this work. Shotgun sequencing of primary specimens necessarily involves incidental sequencing of human genome, and such testing may require IRB

approval and patient consent in some institutions. As there are no FDA-approved metagenomics NGS approaches for infectious disease diagnosis at the time of this writing, any such approach would be implemented as a laboratory-developed test or sent to one of the commercial vendors currently offering such testing. The requirements for validation and quality control of wet lab and computational components of NGS-based workflows are still under discussion by regulatory committees and will undoubtedly continue to undergo refinement as technologies evolve. Until more routine experience with NGS-based testing has been acquired and evaluated, clinical labs will need to exercise caution in their communication and interpretation of results from these methods. But it is clear that NGS-based approaches are here to stay, and it is likely only a matter of time before they become more fully integrated into the standard of care in infectious disease diagnostics, just as they have in other fields of pathology.

ACKNOWLEDGMENTS

This work was supported by the Intramural Research Program of the National Institutes of Health Clinical Center. The content is solely the responsibility of the author and does not necessarily represent the official views of the National Institutes of Health, Department of Health and Human Services, or the U.S. Government.

REFERENCES

1. Lefterova MI, Suarez CJ, Banaei N, Pinsky BA. 2015. Next-generation sequencing for infectious disease diagnosis and management: a report of the Association for Molecular Pathology. *J Mol Diagn* 17:623–634. <https://doi.org/10.1016/j.jmoldx.2015.07.004>.
2. Naccache SN, Federman S, Veeraraghavan N, Zaharia M, Lee D, Samayoa E, Bouquet J, Greninger AL, Luk KC, Enge B, Wadford DA, Messenger SL, Genrich GL, Pellegrino K, Grard G, Leroy E, Schneider BS, Fair JN, Martinez MA, Isa P, Crump JA, DeRisi JL, Sittler T, Hackett J, Jr, Miller S, Chiu CY. 2014. A cloud-compatible bioinformatics pipeline for ultrarapid pathogen identification from next-generation sequencing of clinical samples. *Genome Res* 24:1180–1192. <https://doi.org/10.1101/gr.171934.113>.
3. Wilson MR, Naccache SN, Samayoa E, Biagtan M, Bashir H, Yu G, Salamat SM, Somasekar S, Federman S, Miller S, Sokolic R, Garabedian E, Candotti F, Buckley RH, Reed KD, Meyer TL, Serogy CM, Galloway R, Henderson SL, Gern JE, DeRisi JL, Chiu CY. 2014. Actionable diagnosis of neuroleptospirosis by next-generation sequencing. *N Engl J Med* 370:2408–2417. <https://doi.org/10.1056/NEJMoa1401268>.
4. Briese T, Kapoor A, Mishra N, Jain K, Kumar A, Jabado OJ, Lipkin WI. 2015. Virome capture sequencing enables sensitive viral diagnosis and comprehensive virome analysis. *mBio* 6:e01491-15. <https://doi.org/10.1128/mBio.01491-15>.
5. Greninger AL, Naccache SN, Federman S, Yu G, Mbala P, Bres V, Stryke D, Bouquet J, Somasekar S, Linnen JM, Dodd R, Mulembakani P, Schneider BS, Muyembe-Tamfum JJ, Stramer SL, Chiu CY. 2015. Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis. *Genome Med* 7:99. <https://doi.org/10.1186/s13073-015-0220-9>.
6. Naccache SN, Greninger A, Samayoa E, Miller S, Chiu CY. 2015. Clinical utility of unbiased metagenomic next-generation sequencing in diagnosis of acute infectious diseases: a prospective case series. *Open Forum Infect Dis* 2:103. <https://doi.org/10.1093/ofid/ofv131.23>.
7. Naccache SN, Peggs KS, Mattes FM, Phadke R, Garson JA, Grant P, Samayoa E, Federman S, Miller S, Lunn MP, Gant V, Chiu CY. 2015. Diagnosis of neuroinvasive astrovirus infection in an immunocompromised adult with encephalitis by unbiased next-generation sequencing. *Clin Infect Dis* 60:919–923. <https://doi.org/10.1093/cid/ciu912>.
8. Kennedy PGE, Quan PL, Lipkin WI. 2017. Viral encephalitis of unknown cause: current perspective and recent advances. *Viruses* 9:138. <https://doi.org/10.3390/v9060138>.
9. Somasekar S, Lee D, Rule J, Naccache SN, Stone M, Busch MP, Sanders C, Lee WM, Chiu CY. 2017. Viral surveillance in serum samples from patients with acute liver failure by metagenomic next-generation sequencing. *Clin Infect Dis* 65:1477–1485. <https://doi.org/10.1093/cid/cix596>.
10. Cummings LA, Kurosawa K, Hoogstraal DR, SenGupta DJ, Candra F, Doyle M, Thielges S, Land TA, Rosenthal CA, Hoffman NG, Salipante SJ, Cookson BT. 2016. Clinical next generation sequencing outperforms standard microbiological culture for characterizing polymicrobial samples. *Clin Chem* 62:1465–1473. <https://doi.org/10.1373/clinchem.2016.258806>.
11. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, Turner P, Parkhill J, Loman NJ, Walker AW. 2014. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol* 12:87. <https://doi.org/10.1186/s12915-014-0087-z>.
12. Thoendel M, Jeraldo PR, Greenwood-Quaintance KE, Yao JZ, Chia N, Hanssen AD, Abdel MP, Patel R. 2016. Comparison of microbial DNA enrichment tools for metagenomic whole genome sequencing. *J Microbiol Methods* 127:141–145. <https://doi.org/10.1016/j.mimet.2016.05.022>.
13. Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA, Chen W, Fungal Barcoding Consortium. 2012. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for fungi. *Proc Natl Acad Sci U S A* 109:6241–6246. <https://doi.org/10.1073/pnas.1117018109>.
14. Salipante SJ, Kawashima T, Rosenthal C, Hoogstraal DR, Cummings LA, Sengupta DJ, Harkins TT, Cookson BT, Hoffman NG. 2014. Performance comparison of Illumina and ion torrent next-generation sequencing platforms for 16S rRNA-based bacterial community profiling. *Appl Environ Microbiol* 80:7583–7591. <https://doi.org/10.1128/AEM.02206-14>.
15. Kilianski A, Carcel P, Yao S, Roth P, Schulte J, Donarum GB, Fochler ET, Hill JM, Liem AT, Wiley MR, Ladner JT, Pfeffer BP, Elliot O, Petrosov A, Jima DD, Vallard TG, Melendrez MC, Skowronski E, Quan PL, Lipkin WI, Gibbons HS, Hirschberg DL, Palacios GF, Rosenzweig CN. 2015. Pathosphere.org: pathogen detection and characterization through a web-based, open source informatics platform. *BMC Bioinformatics* 16:416. <https://doi.org/10.1186/s12859-015-0840-5>.
16. Byrd AL, Perez-Rogers JF, Manimaran S, Castro-Nallar E, Toma I, McCaffrey T, Siegel M, Benson G, Crandall KA, Johnson WE. 2014. Clinical PathoScope: rapid alignment and filtration for accurate pathogen identification in clinical samples using unassembled sequencing data. *BMC Bioinformatics* 15:262. <https://doi.org/10.1186/1471-2105-15-262>.
17. Hong C, Manimaran S, Shen Y, Perez-Rogers JF, Byrd AL, Castro-Nallar E, Crandall KA, Johnson WE. 2014. PathoScope 2.0: a complete computational framework for strain identification in environmental or clinical sequencing samples. *Microbiome* 2:33. <https://doi.org/10.1186/2049-2618-2-33>.
18. Ames SK, Hysom DA, Gardner SN, Lloyd GS, Gokhale MB, Allen JE. 2013. Scalable metagenomic taxonomy classification using a reference genome database. *Bioinformatics* 29:2253–2260. <https://doi.org/10.1093/bioinformatics/btt389>.
19. Wood DE, Salzberg SL. 2014. Kraken: ultrafast metagenomic sequence

- classification using exact alignments. *Genome Biol* 15:R46. <https://doi.org/10.1186/gb-2014-15-3-r46>.
20. Segata N, Waldron L, Ballarini A, Narasimhan V, Jousson O, Huttenhower C. 2012. Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat Methods* 9:811–814. <https://doi.org/10.1038/nmeth.2066>.
 21. Greninger AL, Messacar K, Dunnebacke T, Naccache SN, Federman S, Bouquet J, Mirsky D, Nomura Y, Yagi S, Glaser C, Vollmer M, Press CA, Kleinschmidt-DeMasters BK, Dominguez SR, Chiu CY. 2015. Clinical metagenomic identification of *Balamuthia mandrillaris* encephalitis and assembly of the draft genome: the continuing case for reference genome sequencing. *Genome Med* 7:113. <https://doi.org/10.1186/s13073-015-0235-2>.
 22. Schmidt K, Mwaigwisya S, Crossman LC, Doumith M, Munroe D, Pires C, Khan AM, Woodford N, Saunders NJ, Wain J, O'Grady J, Livermore DM. 2017. Identification of bacterial pathogens and antimicrobial resistance directly from clinical urines by nanopore-based metagenomic sequencing. *J Antimicrob Chemother* 72:104–114. <https://doi.org/10.1093/jac/dkw397>.
 23. Truong DT, Tett A, Pasolli E, Huttenhower C, Segata N. 2017. Microbial strain-level population structure and genetic diversity from metagenomes. *Genome Res* 27:626–638. <https://doi.org/10.1101/gr.216242.116>.
 24. Ivy MI, Thoendel MJ, Jeraldo PR, Greenwood-Quaintance KE, Hanssen AD, Abdel MP, Chia N, Yao JZ, Tande AJ, Mandrekar JN, Patel R. 2018. Direct detection and identification of prosthetic joint pathogens in synovial fluid by metagenomic shotgun sequencing. *J Clin Microbiol* 56:e00402-18. <https://doi.org/10.1128/JCM.00402-18>.
 25. Tande AJ, Patel R. 2014. Prosthetic joint infection. *Clin Microbiol Rev* 27:302–345. <https://doi.org/10.1128/CMR.00111-13>.
 26. Thoendel M, Jeraldo P, Greenwood-Quaintance KE, Yao J, Chia N, Hanssen AD, Abdel MP, Patel R. 10 April 2018. Identification of prosthetic joint infection pathogens using a shotgun metagenomics approach. *Clin Infect Dis* <https://doi.org/10.1093/cid/ciy303>.
 27. Thoendel M, Jeraldo P, Greenwood-Quaintance KE, Chia N, Abdel MP, Steckelberg JM, Osmon DR, Patel R. 2017. A novel prosthetic joint infection pathogen, *Mycoplasma salivarium*, identified by metagenomic shotgun sequencing. *Clin Infect Dis* 65:332–335. <https://doi.org/10.1093/cid/cix296>.
 28. Thoendel M, Jeraldo P, Greenwood-Quaintance KE, Yao J, Chia N, Hanssen AD, Abdel MP, Patel R. 2017. Impact of contaminating DNA in whole-genome amplification kits used for metagenomic shotgun sequencing for infection diagnosis. *J Clin Microbiol* 55:1789–1801. <https://doi.org/10.1128/JCM.02402-16>.
 29. Tarabichi M, Shohat N, Goswami K, Alvand A, Silibovsky R, Belden K, Parvizi J. 2018. Diagnosis of periprosthetic joint infection: the potential of next-generation sequencing. *J Bone Joint Surg Am* 100:147–154. <https://doi.org/10.2106/JBJS.17.00434>.
 30. Street TL, Sanderson ND, Atkins BL, Brent AJ, Cole K, Foster D, McNally MA, Oakley S, Peto L, Taylor A, Peto TEA, Crook DW, Eyre DW. 2017. Molecular diagnosis of orthopedic-device-related infection directly from sonication fluid by metagenomic sequencing. *J Clin Microbiol* 55: 2334–2347. <https://doi.org/10.1128/JCM.00462-17>.