

Oxford nanopore sequencing in clinical microbiology and infection diagnostics

Dropen Sheka[†], Nikolay Alabi[†] and Paul M. K. Gordon 

Corresponding author: Paul M. K. Gordon, Department of Biochemistry & Molecular Biology, Faculty of Medicine, University of Calgary, Calgary, Alberta, T2N 1N4, Canada. Tel.: +14032106909; Fax: +1402109685; E-mail: gordonp@ucalgary.ca

[†]These authors contributed equally to this work.

Abstract

Extended turnaround times and large economic costs hinder the usage of currently applied screening methods for bacterial pathogen identification (ID) and antimicrobial susceptibility testing. This review provides an overview of current detection methods and their usage in a clinical setting. Issues of timeliness and cost could soon be circumvented, however, with the emergence of detection methods involving single molecule sequencing technology. In the context of bringing diagnostics closer to the point of care, we examine the current state of Oxford Nanopore Technologies (ONT) products and their interaction with third-party software/databases to assess their capabilities for ID and antimicrobial resistance (AMR) prediction. We outline and discuss a potential diagnostic workflow, enumerating (1) rapid sample prep kits, (2) ONT hardware/software and (3) third-party software and databases to improve the cost, accuracy and turnaround times for ID and AMR. Multiple studies across a range of infection types support that the speed and accuracy of ONT sequencing is now such that established ID and AMR prediction tools can be used on its outputs, and so it can be harnessed for near real time, close to the point-of-care diagnostics in common clinical circumstances.

Key words: bacterial pathogen identification; single molecule sequencing; antimicrobial resistance prediction; point-of-care diagnostics

Introduction

Current state of practice

A number of reasons portend the need for improvement in the accuracy, sensitivity and speed of diagnosing microbial infections in order to improve clinical outcomes. Primarily, a speedy diagnosis is needed when treating bacterial infections, where time is of the essence for positive patient outcomes. For instance, every hour delay in appropriate bloodstream infection-related sepsis treatment decreases survival by 7.6% [1]. Similarly, the long wait times and inaccuracies of tests for infections such as

tuberculosis and lyme disease often leave patients to endure the symptoms of their infections for extended periods of time [2, 3]. Prompt recognition is critical to the hastening of treatment plan initiation. A consequence of misdiagnosis is the prescription of antibiotics that are ultimately wasted and cause adverse side effects for patients. Economic impacts of misdiagnosis of bacterial infections have been studied in the context of cellulitis [4]. They found that for a single institution, misdiagnosis of cellulitis costs \$195 million to \$515 million annually in avoidable health-care spending, exclusive of the costs of the antibiotics themselves and the complications resulting from inappropriate treatment [4]. Another study evaluated the accuracy of diagnosis

Dropen Sheka is a data-driven and computational oncology researcher in the Department of Biochemistry and Molecular Biology in the Faculty of Medicine at the University of Calgary.

Nikolay Alabi is a data-driven and computational oncology researcher in the Department of Biochemistry and Molecular Biology in the Faculty of Medicine at the University of Calgary.

Dr Paul M. K. Gordon is the Bioinformatics Manager at the Cumming School of Medicine's Centre for Health Genomics and Informatics, a core facility of the University of Calgary. He has been doing nanopore sequence analysis since 2014 and microbial bioinformatics generally since 1996.

Submitted: 18 August 2020; Received (in revised form): 26 October 2020

© The Author(s) 2021. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oup.com

Table 1. Review of current microbial infection diagnostic techniques

Technique	Summary	ID capability	Time of fastest technique available for AST (h)	Associated cost	Automatic (A) or manual (M)
Culture-based techniques	Bacteria inoculated on agar plates with different concentrations of antibiotics	Limited to metabolic activity of pathogen on substrate	4.5–18 (MicroScan WalkAway)	Very expensive, exact figures are NA [9]	M
Molecular-based methods	Utilizes conventional PCR or quantitative real-time PCR to amplify specific sequences of nucleic acids, allowing simultaneous pathogen ID and AST	Limited to recognition of specific predetermined sequences of AMR genes	1.0 (BioFire)	>\$49,000 for BioFire Assay >\$100 per test [107]	A
Spectrometry MALDI-TOF MS	MALDI-TOF MS is based on the rapid ionization of the bacteria/yeast ribosomal proteins using a laser pulse. The calculated mass of the ions is the specific sample fingerprint of the bacterial/yeast species	Yes, but protein profile of each species must be predetermined; poor for similar species	< 5.0—Bruker Daltonics	\$0.50/sample and \$150 000 for machinery [108]	A
Spectrometry approaches combined with molecular tools	The PCR/electrospray-mass spectrometry (PCR/ESI-MS) is a fairly recent technology that couples a molecular method to a spectrometry approach	Yes, protein profile of each species must be predetermined	<6.0—Ibis Biosciences	\$50–100 per sample and \$450 000 for machinery [109]	A

and the appropriateness of prescribed antimicrobial courses at the Minneapolis Medical Center [5]. They determined that 42% of patients were incorrectly diagnosed of which 95% were prescribed inappropriate antibiotics. Inaccurate diagnosis can lead to potentially harmful antibiotic use for patients. A recent case study detailing the misdiagnosis of melioidosis reported two different patients having to be treated with multiple different courses of antibiotics as a result of inaccurate identification (ID) of the infecting bacteria [6]. Subsequently, one of the patients developed drug-induced hepatitis due to the prescription of an antitubercular medication from misdiagnosis, adding to their existing symptoms. The appropriateness of treatment a patient receives has a profound impact on their quality of life, and thus an accurate diagnosis is necessary.

Review of current techniques

Table 1 provides a summarized review of the most common bacterial infection diagnostic techniques. Current standard diagnostic techniques are typically based on serology, including the extensive multistep processes to detect microbial growth, followed by the isolation and ID of samples, and then antimicrobial susceptibility testing (AST), making blood cultures laborious and time consuming [7, 8]. Additionally, the accuracy and sensitivity of the microbial profiles are poor [9]. However, their automation and high degree of standardization across the field means these techniques remain valuable [10].

The main advantage of most molecular-based methods over culture-based methods in diagnostics is the ability to perform tests without culture—saving significant time and labor—directly on the patient samples. Most of the contemporary technologies require a deoxyribonucleic acid (DNA) extraction step in which low volumes of clinical sample lead to lower

volumes of DNA. This can lead to false-negative results, masking an infection and minimizing the clinical relevance of such technologies [11].

Molecular techniques can detect resistance traits in the pathogens, providing valuable information on the type of antimicrobial treatment to prescribe as well as on the identity of the pathogen [9]. A polymerase chain reaction (PCR) step is required to identify the resistance genes. However, this step requires previous knowledge of the sequences to amplify, and the presence of a resistance gene does not always correspond with the phenotypic resistance [12].

Relatively recent spectrometry methods offer new alternative technologies for AST. Such spectrometry methods include the matrix-assisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF MS). The MALDI-TOF MS holds utility as it does not require a unique test yet can be used for gram-positive and gram-negative bacteria and yeast, unlike other methods [13, 14]. Its ability to distinguish between species with similar ribosomal protein sequences is difficult, although a study by Fernández-Álvarez et al. [15] differentiated closely related *Flavobacterium* species (*Flavobacterium flevense*, *Flavobacterium succinicans*, *Flavobacterium columnare*, *Flavobacterium branchiophilum* and *Flavobacterium johnsoniae*) from *Flavobacterium psychrophilum* using ribosomal protein biomarkers and MALDI-TOF MS technology. In cases where ID is infeasible, supplementing with classical approaches is required. Other challenges for this technique include high amounts of host protein found in most clinical samples and time-consuming sample preparation that is still marginally less than cultures. This method also tends to under-evaluate infections with low bacterial counts, because of the low protein number, and can lead to false-negatives [9]. While the machinery for this technique is often very expensive, the cost of each test is inexpensive.

Spectrometry approaches combined with molecular tools, namely the IRIDICA BAC Bacterial Sepsis Infection (BSI) assay, overcomes weaknesses in the analysis of complex samples and performing of culture-independent analysis [16]. Similar to the MALDI-TOF MS approach, this method requires expensive machinery, however, the cost for each test is minimal. A point of significance is the rapid turnaround time, with results in under 6 h, plus the sample transport logistics time required for centralized testing.

Review of emerging techniques

In contrast to most current AST methods, technologies such as the oCelloScope—an image-based assay—can be used to perform AST in real time without the need to attach bacterial cells to an inert surface [9, 17]. Other technologies such as the Accelerate Pheno System can perform both ID and AST with high resolution, allowing the diagnosis of both mono- and poly-microbial infections. However, this technique still requires time-consuming blood cultures before its rapid assay results can be generated. Additionally, the limited support in species for bacterial ID and expensive machinery may prevent this technology from being applicable in a real-time clinical setting.

Many of the emerging biochemical methods share many common features, including their straightforward manipulation, low production costs and rapid turnaround time. Despite these favorable features, these methods mainly focus on AST with very limited capability for pathogen species ID. However, a notable biochemical method, the electronic nose (e-nose), does detect the chemical fingerprints of volatile organic compounds for pathogen ID [9]. The e-nose method can be done with minimal to no invasion and has a high specificity rate with a 10-min turnaround time. A drawback to the technology is its inability to identify and quantify chemical species in the usually complex volatile organic compounds (VOCs) mixtures. In terms of its clinical practicality, the e-nose has a large price and complexity of operation which requires expert operators—most likely limiting its implementation to hospital settings.

While enhancements in the accuracy of diagnostic tests for bacterial infections have improved therapies and avoided the unnecessary use of antibiotics in recent years, limitations still remain. The spectrum of pathogens and their varied clinical presentations and effects is wide; therefore, the current standards beg reexamination. Devising an effective method that can optimize timeliness and accuracy, while minimizing the cost and complexity, is still needed despite the recent advancements.

Table 2 shows that emerging trends such as imaging-based technologies and biochemical methods primarily focus on rapid AST, with limited capabilities of bacterial species/gene ID. Contrastingly, while e-nose technology is tailored for bacterial ID, its clinical roll out in hospital settings remains years away. An advantage of rapid bacterial antimicrobial resistance (AMR) gene ID is the accompanying information that informs a health-care practitioner of the appropriate antibiotic to prescribe to a patient, eliminating the need for AST. There are a growing number of databases that are continually updated, which offer drug options, such as those mentioned in Table 3, which depicts the integrated data and analysis tools required to support biomedical research on bacterial infectious diseases [18]. Thus, a rapid ID method for bacterial infections in combination with a growing AMR knowledge base could greatly improve the clinical treatment of bacterial infections.

DNA sequencing as a diagnostic tool

PCR-based testing has proven dominant among the molecular technologies for pathogen ID due to its speed, low complexity and cost. PCR though can be suboptimal in terms of resolving the strain or AMR [19], and it is sensitive to the a priori assumption of disease agents implied by the PCR panel selected [8]. Strain resolution can be improved by Sanger sequencing of the PCR products, but Sanger sequencing remains constrained by the pathogen hypotheses for which testing is performed.

Next-generation sequencing

An extant methodology for bacterial ID not yet mentioned is the next-generation sequencing (NGS) technology. While there are a variety of different types of NGS technologies, here, we use NGS as a shorthand to refer to second-generation sequencing technologies. NGS provides several possible applications and advantages to clinical microbiology. For instance, NGS can bypass many limitations of the current diagnostic approaches by enabling clinicians to assess for multiple pathogens during the initial diagnostic evaluation, thereby avoiding many rounds of testing to search for progressively less-common pathogens [20–22]. Furthermore, NGS has the potential to bypass the multitude of tests that are being conducted on clinical specimens [23, 24]. Using NGS to research the human microbiome in relation to health and disease has been critical for clinical applications such as fecal transplant therapy [25]. Another benefit of the NGS approach in clinical microbiology is the ability to develop new and enhanced diagnostic assays through the use of sequence information derived from NGS-generated data to improve specific DNA targets and primers used in multiplex assays [23, 26]. There are a variety of other ways that NGS can impact clinical microbiology that are detailed in the *Applications of Clinical Microbial Next-Generation Sequencing: Report on an American Academy of Microbiology Colloquium*, held in Washington, DC, in April 2015, but for the purposes of this review, we focus on AMR detection. There are a variety of high-throughput sequencing technologies available as commercial platforms, with Illumina's sequencing-by-synthesis as the dominant player. Its ability to rapidly sequence entire bacterial genomes and microbiomes with high resolution and analyze these sequences using improving bioinformatic tools lend credibility to the idea of NGS's clinical utility [27]. However, lack of practical NGS clinical adoption to date for clinical microbiology can be attributed to cost analysis in health economics, complex workflows, the need for quality controls and interfering contamination events [28, 29]. High throughput sequencing to guide clinical decision making could still be valuable if these downsides of NGS are addressed. Use of NGS in clinical microbiology is primarily limited to priority cases, such as outbreaks of food-borne illness [30–34] or highly pathogenic/drug-resistant diseases [35, 36] and in post hoc molecular epidemiology [29–32].

Nanopore sequencing

So-called third-generation sequencing (TGS) includes various technology platforms with the commonality that they sense the bases of individual DNA or ribonucleic acid (RNA) molecules, as opposed to NGS, where the signal is based on the amplification and synthesis of template DNA molecules. Raw TGS data are therefore more error prone, but methods to generate high-quality consensus from TGS signals approach the quality

Table 2. Review of emerging technologies and future trends in microbial infection diagnostic techniques

Technique	Summary	ID capability	AST capability
Imaging-based technologies	Usage of microscopy techniques and image-based assays that can provide pathogen ID and AST	Limited range of species and poor resolution of ID	High AST performance
Biochemical methods	Usage of biosensors to identify and detect bacterial growth through biochemical flags from cells	Limited to the interaction of bacteriophages with bacteria	High AST performance
Electric nose	Recognize single chemical fingerprint patterns through an array of semi-selective sensors for VOCs	Yes, but VOC fingerprint must be predetermined. High resolution	AST is not a feature of these devices

Table 3. Databases for AMR gene ID as well as drug options

Database	Summary	Suitable species for database [90]
PATRIC	Provides integrated data and analysis tools to support biomedical research on bacterial infectious diseases [18]	<i>Mycobacterium tuberculosis</i> , <i>S. pneumoniae</i> , <i>S. aureus</i> , <i>Acinetobacter baumannii</i> —non-serovar Typhi <i>S. enterica</i>
CARD	A bioinformatic database of resistance genes and proteins along with their corresponding phenotypes [110]	<i>Salmonella enterica</i> serovar Typhi, <i>Shigella sonnei</i> , non-serovar Typhi <i>S. enterica</i>
ARG-ANNOT	Detects existing and putative new AR genes in bacterial genomes. ARG-ANNOT uses a local alignment program in Bio-Edit software that allows the user to analyze sequences without a web interface [111]	NA
ResFinder	ResFinder identifies provided AMR genes and/or chromosomal mutations in total or partial sequenced isolates of bacteria [112]	<i>Escherichia coli</i> , <i>Enterococcus faecalis</i> , <i>E. faecium</i>
PlasmidFinder	The service identifies plasmids in total or partial sequenced isolates of bacteria [113]	NA
VFDB	VFDB is a repository to store, search, retrieve and update information about virulence factors from an assortment of bacterial pathogens [114]	NA
Mykrobe	Analyzes the entire genome of a bacterial sample and predicts to which drugs the bacteria of question is resistant to—all within a short turnaround period [115]	<i>M. tuberculosis</i> , <i>S. aureus</i>
PhyResSE	Determines strain lineage and AR of bacterial sample from NGS data [116]	<i>M. tuberculosis</i>

Note. NA, not available.

of individual NGS reads by using template molecule circularization and concatenation for repeated measurement [37, 38]. The leading platforms in the TGS space are the Pacific Biosciences and Oxford Nanopore Technologies (ONT). PacBio is similar to ONT in terms of long-read capability, but PacBio is more similar to core sequencing facility NGS platforms in terms of turnaround time for individual reads, capital cost and physical requirements (cooling, power, etc.). While we acknowledge there are other nanopore technologies that remain in the research and development stage (e.g. Stoddart et al. [39] and solid-state nanopore technology recently reviewed in Manrao et al. [40]), henceforth we will refer to the existing commercially available nanopore technology, ONT, whose use has been documented in a clinical setting [41]. The remainder of this review will therefore focus on the tools that are common to NGS and TGS infection analysis and on specific clinic-oriented aspects unique to the existing nanopore sequencing capabilities.

Recently, a variety of studies have demonstrated nanopore sequencing's capability for bacterial ID and AMR gene detection in a clinical setting. Table 4 shows studies that utilized nanopore sequencing technologies for bacterial ID. Notable innovations within each study include improving accuracy and shortened turnaround times. In Matsumoto and colleagues' [42] 2019 study, a hybrid assembly using sequencers from Illumina and ONT and the development of a novel multilocus sequence typing (MLST) database provided the most accurate detection of a

variety of nontuberculous mycobacteria (NTM) among all currently used clinical methods of bacterial ID. In terms of improving turnaround times from sampling to results, Leggett and colleagues' [43] 2019 study demonstrated a capacity of an optimized workflow to shorten turnaround times for fecal infant sample bacteria and AMR ID to less than 5 h.

The versatility of ONT sequencing has also been demonstrated to identify bacterial species via both whole genome approaches and PCR amplification of 16S ribosomal RNA (rRNA) sequences. While many of the studies used targeted rDNA or rRNA approaches to identify species in environmental samples [44–46], studies such as Kamathewatta et al. [47] sequenced clinical samples. A metagenomics approach to identify bacteria using ONT sequencing is becoming increasingly popular due to its rapid turnaround time and handheld convenience. Implementation in the hospital setting could markedly increase the utility of ONT in real-time point-of-care clinical bacterial pathogen diagnostics.

In terms of AMR detection, some studies in Table 4, such as Lim et al. (2019), showed the potential for AMR detection either through PCR amplification of the target gene or through whole genome sequencing and ID of the gene. ONT sequencing allows rapid ID of these genes in comparison to other NGS approaches.

ONT sequencing can overcome several challenges faced by NGS as a diagnostic tool. Generalizing and extending the workflows described in [42–47], we present and discuss the

Table 4. Studies involving bacterial species and AMR gene ID using nanopore (ONT) technologies

Date of publication	Study	Description of method	Bacterial species detected	Reported time	Reported nanopore accuracy
23 July 2016	Shin et al. 2016 [117]	16S rRNA amplification and sequencing of entire mouse gut microbiota	16 phylogenetically distinct species distributed in 13 genera	<3 h for library construction, real-time detection	79.6% coverage
26 September 2016	Schmidt et al. [118]	WGS of bacterial DNA enrichment from urine samples using MoYsis and MagNA Pure Compact Nucleic Acid Isolation Kit or NEBNext Microbiome DNA Enrichment Kit	<i>Klebsiella pneumoniae</i> <i>E. coli</i> <i>Escherichia cloacae</i>	4 h from sample to result	70–85% coverage. 100% correct bacterial ID
15 September 2017	Kerkhof et al. 2017 [44]	16S rRNA amplification and sequencing of soil samples mixed with bioreactor DNA	<i>Acidovorax wautersii</i> , <i>Comamonas</i> <i>nitrativorans</i> and <i>Stenotrophomonas</i> <i>rhizophila</i>	6 h MinION runs	79–100% coverage
31 October 2017	Xia et al. 2017 [119]	WGS of coliform bacteria with hybrid assembly of nanopore and Illumina sequences. DNA from bacterial isolates were extracted through Fast DNA® Spin Kit for soil usage	Coliform bacteria	30 h from sample to result	84.6% accuracy in alignment with Illumina assembly
1 November 2017	Moon et al. 2017 [120]	16S rRNA amplicon sequencing performed by MinION with PCR amplification of 16S rRNA genes. Full-length 16S rRNA genes were confirmed by Sanger sequencing	<i>Campylobacter fetus</i>	Unspecified—51 min to generate 43 044 reads	99.9% correct alignment to <i>Campylobacter fetus</i> , 89% and 60% correct alignment to correct species and subspecies, respectively
18 June 2018	Tanaka et al. 2018 [121]	WGS through usage of nanopore, Illumina or hybrid assembly. DNA extracted through Wizard® Genomic DNA Purification Kit with NucleoSpin® Tissue—protocols corresponding to gram-negative bacteria	<i>Vibrio</i> <i>aphrogenes</i> , <i>Vibrio</i> <i>alginovorans</i> , <i>Vibrio</i> <i>casei</i> , <i>Vibrio</i> <i>litoralis</i> and <i>Vibrio</i> <i>rumoiensis</i>	Unspecified	100% coverage from Nanopore–MinION presents 99% average nucleotide identity (ANI) between <i>Ponticus</i> clade strains and 95–97% ANI between <i>Splendius</i> clade strains
19 October 2018	Chalupowicz et al. 2018 [122]	NucleoSpin Plant II Midi Kit was used to extract DNA from leaf tissue, and the MasterPure Complete DNA Purification Kit was used to isolate total DNA from seeds, stems and fruits. The WIMP workflow was utilized for bacterial ID	<i>Clauibacter michiganensis</i> , <i>Pseudomonas corrugata</i> , <i>Acidovorax citrulli</i> , <i>Pantoea agglomerans</i> and <i>Candidatus</i> <i>Phytoblasma aurantifolia</i>	Pathogen ID within 1–2 h of nanopore sequencing	Mean coverage for bacteria was typically <1—with mean coverage = aligned bases divided by genome size
October 2018	Leggett et al. 2020 [123]	WGS of genomic DNA from a microbial mock community used in the Human Microbiome Project	20 bacterial strain sequences were correctly identified	4–5 h for pathogen and AMR detection from sample to result	82–89% for both bacterial identity and sequence coverage
4 December 2018	Golparian et al. 2018 [52]	Gonococcal and clinical isolates were cultured, and the Wizard Genomic DNA Purification Kit was used for DNA isolation prior to WGS	<i>Neisseria gonorrhoeae</i> and the <i>de novo</i> assembly of gonococcal genomes	<24 h for sequencing	95.2–98.3% coverage

(Continued)

Table 4. Continue

Date of publication	Study	Description of method	Bacterial species detected	Reported time	Reported nanopore accuracy
12 January 2019	Kai et al. 2019 [124]	16S rRNA gene amplicon nanopore sequencing—16S rRNA genes were amplified by PCR and bacterial DNA was purified by DNeasy Blood and Tissue Kit	<i>E. coli</i> and <i>S. aureus</i>	<2 h of analysis for bacterial ID	>90% correct read alignment
18 March 2019	McManus et al. 2019 [125]	Hybrid assemblies based on Illumina MiSeq short-read and Oxford Nanopore MinION long-read whole genome sequences of oral rinse samples. DNA was extracted from bacterial isolates using <i>S. aureus</i> Genotyping Kit 2 DNA microarray kit and the DNeasy Blood and Tissue kit	<i>S. aureus</i>	Unspecified	–
27 March 2019	Hamner et al. 2019 [126]	WGS of water samples after DNA isolation using PowerWater DNA isolation kit	<i>Acidovorax</i> and <i>Aeromonas salmonicida</i> among several others, including eukaryotic protists and fungi	Unspecified	93–96% correct read alignment to reference genomes
5 March 2019	Lim et al. 2019 [127]	Genomic DNA from bacterial isolates was extracted using the OMEGA Bio-tek EZNA [®] bacterial DNA Kit. WGS by nanopore sequencing was applied to characterize AMR through functional annotative databases	<i>Mannheimia haemolytica</i>	<48 h for the genotyping of 12 microbial genomes	91% correct assignment to species, 93% to correct genus and >99% to the correct family
30 May 2019	Kamathewatta et al. 2019 [47]	WGS of DNA from samples isolated using the Wizard Genomic DNA Purification Kit. Followed by 16S rRNA amplification	Genus level taxonomic classification to 83% and 94% of reads	<16 h for total sequence output, including human DNA	80–90% sequence coverage
24 June 2019	Sakai et al. 2019 [128]	Bacterial DNA was extracted through bead beating and chemical lysis protocols. WGS was performed with the EPI2ME workflow—gram-positive bacterial ID indicated poor accuracy	ESBL-producing gram-negative bacteria	Approximately 30 min	100% ID for gram-negative bacteria, but 60% accuracy for gram-positive strains
9 July 2019	Matsumoto et al. 2019 [42]	MLST-based ID in conjunction with hybrid assemblies from Illumina and nanopore sequencers. Genomic DNA from bacterial colonies were extracted through the DNeasy PowerSoil Kit	<i>Mycobacterium avium</i> complex, <i>Mycobacterium kansasii</i> and the <i>Mycobacteroides abscessus</i> complex	Detection of NTM within 10 min of nanopore sequencing	16S rRNA gene sequence homology only correctly identified 11 of 175 NTM species with 99% identity—this does not include MLST scoring
24 July 2019	Bialasiewicz et al. 2019 [129]	WGS of blood culture sample	<i>Capnocytophaga canimorsus</i>	<18 h for total sequence output, including human DNA	Unspecified

Note. ESBL: extended spectrum beta-lactamase.

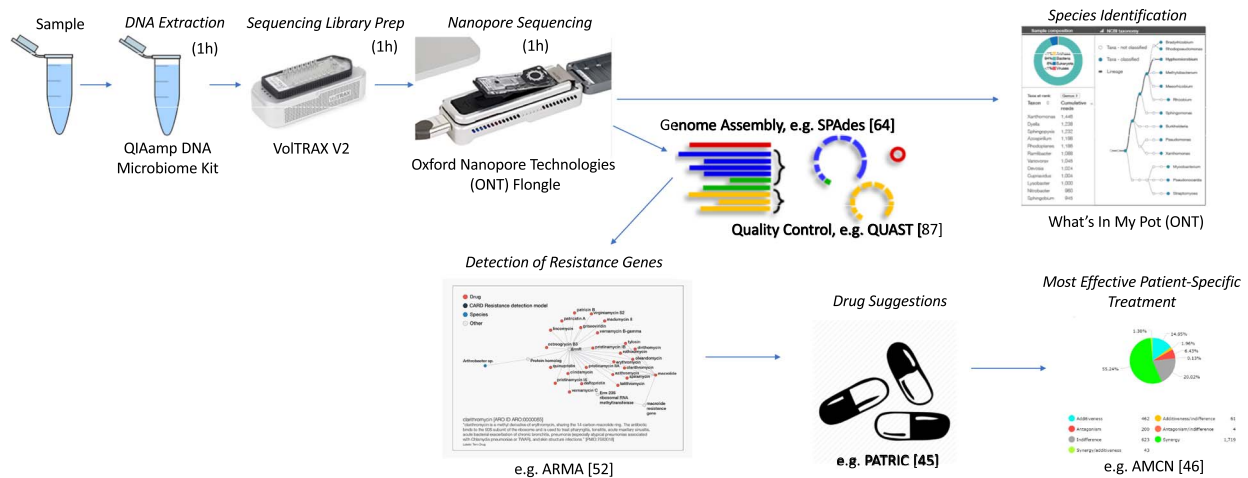


Figure 1. End-to-end steps for rapid, largely automated clinical infection diagnostics, with steps in *italics* and example hardware/tools/databases in regular font.

practicality of a pipeline that has the potential to streamline nanopore sequencing for clinical microbiological use. The ONT Flongle is a novel TGS sequencing instrument that utilizes cost-effective and real-time long-read sequencing technology to routinely generate individual read lengths of tens and thousands of nucleotides. Uniquely, ONT sequencers provide read-by-read data availability [48, 49], meaning there is no fixed minimum run time before the biological sample interpretation can commence. The clinical laboratory suitability of Flongle is also supported by its track record of sequencing genomes of bacterial and viral species to an accuracy needed for identifying and diagnosing AMR genes in bacteria [48, 50]. The Flongle is also an attractive option for the clinic based on its short turnaround times, low-cost and low space occupancy. Additionally, the recent production of the VolTRAX V2 is a piece of technology that can make TGS for microbiological applications viable. The VolTRAX is a small device designed to perform all of the molecular biological manipulations required to convert a raw biological sample to a form ready for analysis on a nanopore sensing device (i.e. MinION) without the need for human intervention.

Figure 1 illustrates how the aforementioned ONT sequencing devices can help automate and inform treatment options in a clinical setting, using ONT's EPI2ME software (Figure 1) and existing NGS bioinformatic databases/tools for microbiology (PATRIC [18] and antimicrobial combinations network (AMCN) [51]) as representative steps in a pipeline.

Patient sample collection and preparation

While the VolTRAX can be utilized for the PCR amplification of the 16S–23S rRNA region for bacterial ID, the entire genome sequence of a potential pathogen is far more valuable for a number of reasons. Firstly, the 16S–23S rRNA region can only be used to identify the species, while in contrast, having the entire genome of a pathogen can help to both identify it and to characterize antibiotic resistance (AR) genes [17, 42]. Furthermore, having the genomic sequence of a pathogen is valuable for further needs such as molecular epidemiology [52, 53], or for tracking the evolution of the pathogen within the host if it gains AMR. For hard-to-culture bacteria such as tuberculosis-causing mycoplasma, genome-based AMR testing can be especially valuable [54].

The primary aim of this step is isolating reliable bacterial DNA from the host (human) sample for nanopore genomic sequencing, while maintaining low cost, labor and time consumption. In order to isolate the microbiome and deplete any human DNA contaminants in a reasonable time span—that also minimizes complexity—the usage of VolTRAX V2 with an affordable isolation kit and workflow may be warranted (Table 5).

The VolTRAX is capable of DNA isolation and PCR amplification on bacterial cultures, with equivalent performance in terms of read length distribution to manually perform the prep protocol with pipettes. Additionally, since there are several reagent ports on the device, the same sample can go through different protocols using different reagents simultaneously to select the best output for DNA quantity. If the DNA outputs are still very low, whole genome amplification can be applied to the resulting DNA to increase the concentration of DNA.

A review of current bacterial DNA isolation kits shows that most kits follow a general workflow that can be adapted to a variety of sample types, including blood, stool, urine, etc. The procedure adapted to the VolTRAX is shown in Figure 1, where a variety of reagents, buffers and lysis solutions are going to be present at each step. The VolTRAX applies a charge to the sample and reagents and moves them to a path programmed by the software. The device can also complete library preparation in tandem with the separation process essentially automating the preparation process—reducing complexity and associated contamination risks. A plausible clinical upside of using the VolTRAX is the automation of the entire preparation process. Potential users could insert the results of their DNA isolation kits into the VolTRAX with a specified protocol, and the rest of the preparation process would be completed by the machine. Depending on the selected protocol, preparation processes could be as short as approximately 1 h.

Some infections present particularly low abundances of microbial populations, and when DNA quantities are barely detectable, small variations in sample quantity or quality may cause misleading results [55]. Thus, choosing kits that produce enriched microbiome DNA from samples is essential. This consideration is critical, and thus choosing kits that optimize the DNA yield, time and cost are important decisions. Table 5 presents some of the commercial choices available for

Table 5. Commercial isolation kits and associated costs

Isolation kit	Description	Cost (USD)	Time (h)	Preps	Cost per sample (USD)
PSP Spin Stool DNA Plus Kit (Strattec Biomedical)	Requiring approximately 400 mg of human or animal stool samples, the protocol has optimized bacterial DNA yield and purity found in human gut microbiota. Produces a considerably higher yield of bacterial DNA when isolated from stool than QIAamp DNA Stool Mini Kit [55]	345	1	50	6.90
PureLink Microbiome DNA Purification Kit (ThermoFisher Scientific)	Kit utilizes PureLink Column technology with optimal recovery of highly purified DNA. Also utilizes beads in conjunction with heat and mechanical disruption to lyse a variety of microorganisms. DNA is compatible with NGS for WGS	311	1.0	50	6.22
Optimized workflow described by Ackerman et al. [55]	Through an optimized pipeline described in Ackerman et al. [55], isolated microbial community DNA yields were maximized greater than any individual isolation kit yield. DNA yields were maximized by combining modified DNA extraction methods such as lyticase/lysozyme digestion, bead beating, boil/freeze cycles, proteinase K and carrier DNA use. The downside to this methodology is its considerably longer preparation periods in comparison to commercially viable kits	+1200	2.5–3	NA	NA
QIAamp DNA Microbiome Kit	Novel protocol that aims to enzymatically deteriorate host/human DNA contaminants prior to bacterial cell lysis and attenuates biases attributed to variable susceptibility to lysis procedures in different bacterial species. Results in optimal representation of microbiome	538	1.0	50	10.76
Microbiome DNA Isolation Kit (Norgen Biotek)	The process provides a simple and convenient DNA isolation protocol for a variety of microbiome samples preserved using Norgen's Swab Collection and DNA Preservation Kit	233	0.5	50	4.66

a variety of DNA isolation kits and protocols, with varied costs, turnaround times and preps.

For example, comparing the costs of commercially available isolation kits in accordance with purified DNA yields, the PSP spin tool DNA Plus Kit or the PureLink Microbiome DNA Purification kits are both economical and clinically applicable isolation kits (Table 5), and they have defined protocols for a wide variety of tissues. Although the optimized workflow described in Ackerman *et al.* [55] is promising, the complexity and economic burden of the preparation and longer wait time may pose an obstacle in widespread clinical adoption. In conjunction with VolTRAX, the kits could prepare clinical samples of patients for Oxford nanopore sequencing with minimal human intervention and in a time-efficient manner. A recent study also optimized the isolation process to eliminate 99.99% of the host nucleic acid contaminants before downstream nanopore sequencing of bacteria conferring lower respiratory infection [56]. The optimized methodology of the study provides turnaround times as short as 6 h—from sample collection to result—but could be simplified and hastened if partially automated with VolTRAX [56]. VolTRAX upholds its potential utility in this workflow as a result of its capabilities in incubation, PCR and library preparation automation, attenuating the complexity and further reducing the length of time of the novel pipeline lending credibility to its future clinical application.

Regarding library preparation, the VolTRAX Sequencing Kit generates sequencing libraries from extracted gDNA in 5–10 min hands-on time plus 45 min of automation. The kit is optimized for simplicity and speed rather than for obtaining maximum throughput. However, some limitations to this step are the cost at \$50/reaction. Ultimately, less than 2 h are required from the isolated DNA to sequencing initiation. Additionally, flow cells are required for the sequencing reaction to occur on the Flongle. Each flow cell is listed at \$90 USD.

Assuming the cost for each sample extraction is approximately \$10 (QIAamp DNA Microbiome Kit), an additional \$50 is required for the library preparation and \$90 is required for the flow cell. Ultimately, \$150 per sample is costly compared to the aforementioned techniques. However, as the VolTRAX technology gains momentum and advances are made, the cost of the library preparation kits and flow cells will likely decrease, and additional capabilities are required to reduce costs implemented, such as sample multiplexing when the volumes are higher.

Oxford nanopore sequencing

Through the combination of speed and thorough coverage of all microorganisms present, metagenomic sequencing-based approaches have the ability to overcome the limitations of both culture and PCR methods. NGS technologies such as Ion Torrent and Illumina are widely used for metagenomics sequencing, but they generally face the limiting requirement of the run to be complete prior to analysis. However, LiveKraken, a recent method, enables analysis of raw Illumina data prior to the completion of the sequencing [57]. The highly parallel throughput of Illumina devices makes it more economically suited to a centralized, batch processing setup. ONT sequencing brings the advantages of built-in real-time data analysis of small or no-batch samples and low capital cost, which are more suited to the constraints of the clinical setting.

Nanopore MinION sequencing technology is approaching comparable performance to NGS at the consensus level, while showing advantages in terms of the size of the device,

simple library preparation workflow, real-time sequencing data generation and analysis, and most importantly, long-read length to provide higher accuracy in terms of complex repeat and/or rearrangement structures (including plasmids) related to AMR. Its main limitation has been per-sample cost. However, Flongle is designed to be the quickest, most accessible and cost-efficient sequencing system for smaller tests and experiments, making it ideal for point-of-care clinical use. It utilizes the same core nanopore technology as MinION, GridION and PromethION, offering direct DNA or RNA analysis, simple preparation, real-time data and analysis. The Flongle is ideal for small samples (such as a microbiome as opposed to the human genome), single runs (removes the issues of sample cross-contamination due to multiplexing and barcoding), performing rapid quality check and species ID.

This hypothesis-free approach is a form of metagenomics, and while metagenomics projects with high diversity of bacteria of bacterial populations may require larger sequencing platforms like the MinION that are less cost-effective, we used the working assumption that, in an acute infection, the etiology of the disease is a bacterium in relatively high abundance. With this assumption, the lower throughput of Flongle will suffice for accurate detection. We realize that, depending on the type of infection, this requires a careful selection by the clinician of the sampling site [29].

Species ID

Currently, NGS as a tool for microbial ID utilizes the sequencing of the 16S–23S rRNA region as it is unique and is found in all bacteria [58]. This method has been reported to be more accurate than other commonly used ID methods that may identify bacteria on the basis of phenotype rather than sequence [19]. NGS allows for the simultaneous ID of several pathogens; however, while this method has very high resolution, some bacterial species can be too genetically similar to differentiate [58]. This issue was recently overcome in a method devised by Sabat *et al.* [58]. While the NGS of 16S–23S rRNA region is culture-independent and more discriminative at the species and genus level than popular approaches such as the MALDI-TOF MS, its main limitations come from cost and speed. With costs per sample—not including labor—amounting to approximately \$75 USD. More importantly, the practicality of such a method in clinical microbiology is infeasible as turnaround times can reach 4 days. Another considerable limitation is the lack of easily implementable and utilizable reference databases for 16S–23S rRNA sequences and complementary software for automated, easy and reliable species ID. While reference 16S–23S rRNA databases do exist, such as Greengene, RDP and SILVA, they require software such as Kraken, QIIME and MOTHUR for use [23, 59–64]. These analyses are computationally intensive and primarily intended for card-carrying bioinformaticians comfortable with the command line as opposed to regular clinicians [23, 59].

A relatively novel pipeline that could be utilized in real-time ID of bacterial species utilizing nanopore technology involves the WIMP workflow. Centrifuge technology allows rapid and accurate ID of potentially pathogenic bacteria through quantification and labeling of reads [56]. The tool utilizes k-mer-based ID indexing methods centered around Burrow-Wheelers Transform (BWT) and the Ferragina-Manzini (FM) index to classify reads into species by referencing RefSeq in real time [56, 65]. Ultimately, the WIMP workflow classifies and identifies species in real time: as soon as a strand of DNA passes through the pore, it can be base-called and analyzed. WIMP is a part of an EPI2ME

AMR pipeline (ONT) that can be used to identify bacterial, viral and human reads.

This pipeline was used within Charalampous et al.'s [56] 2019 study of the ID of the lower respiratory infection-conferring bacteria. A drawback to the usage of WIMP is its relatively high false-positive detection rates as evidenced in the study. High misID rates can be attributed to the inefficacy of k-mer-based read ID methods in differentiating highly similar genomes of closely related organisms (at the species level). False-positives attributable to poor k-mer classification could be mitigated by lightweight scripting that highlights those reads that are assigned to the same species by multiple classification programs. Since each tool has slight algorithmic variations and underlying database indices, commonly classified reads are more likely to be true positives, with some chance of losing true positives that are called by only one tool. This combine-existing-predictions approach is widely used in bioinformatics when false-positive predictions are troublesome, for example, *de novo* transcriptome assembly, gene splicing and miRNA targeting [66–68].

Despite WIMP's limitations, its potential for point-of-care clinical utility to identify species in conjunction with Oxford Nanopore Sequencing in real time is valuable. As the program improves through updates, so will the number of bacterial reference genomes it possesses. The output from the program can be pipelined to a bacterial database, where only species that are harmful to the patient are selected.

The repertoire of extant species and AMR genes that are relevant to defining the etiology of disease are both incomplete in the databases and are ever expanding (i.e. emerging pathogens). This highlights both the importance of maintaining up-to-date versions of AMR gene databases (discussed below) and the opportunity for using clinical sequencing to supplement existing pathogen sentinel programs (the complexity of which is beyond the scope of this work, see for clinical and veterinary reviews [69, 70]).

AMR gene detection

The implementation of wide genome sequencing (WGS) in AMR detection in clinical microbiology has been contentious until recently. Like many other sequencing techniques to determine AMR in bacterial strains (PCR, microarray, etc.), WGS has the ability to detect genetic factors that confer AMR such as genes and mutations [71]. Genomics-based prediction of antimicrobial susceptibility has been shown to be up to 98% concordant with minimum inhibitory concentration (MIC) testing by dilution [72], which gives equivalent results as standard AST disk tests [73]. However, WGS offers the advantage over PCR by being more efficient in bacterial AMR detection, where resistance is predominantly induced by mutational factors [11]. Furthermore, WGS displays strong suits in its ability to detect multiple different targets concurrently to rapidly record novel target sequences into databases and curate gene variants of target sequences [71]. Setbacks in the forms of economic feasibility and clinical inefficiency hinder WGS application to clinical microbiology, but newer technologies from Oxford Nanopore Sequencing mitigate these [71, 73].

In recent years, WGS has been carried out by complex and relatively inexpensive per-base cost sequencing technology, primarily machines manufactured by Illumina and to a lesser extent by Thermo Fisher (Ion Torrent). These 'NGS' technology platforms were a large step up from the Sanger sequencing sequence in terms of parallel data generation capacity and

hence shortened turnaround times for genome-scale projects [71]. NGS machines produce relatively short but highly accurate sequence reads that are comparable to conventional Sanger sequencing methods. These highly accurate but very short read sequences may not unambiguously resolve an AMR gene, even after assembly, especially if repeated sequences are present [71, 74] and as AMR genes often reside in genomic islands containing nearly identical genetic mobility genes and insertion sequences [75]. As a consequence of the short-read lengths, assemblies of bacterial genomes that are enriched for these repeats are often profoundly fragmented and usually incomplete, hindering the potential for clinical usage [48]. However, the NGS approach for AMR detection in clinical microbiology has been shown by a study in Texas where WGS via the Illumina technology accurately predicted resistance to four β -lactams in major gram-negative bacterial pathogens commonly used in the treatment of neutropenic fever [76]. The study demonstrated the merit of WGS usage specifically in regard to detecting insertions, deletions and single-nucleotide polymorphisms (SNPs) that widely used PCR assays fail to detect in certain AMR bacterial species. Although the study utilized manual sequence inspection which was time consuming and not ideal for clinical microbiology, further advancements in the field could circumvent these obstacles [76]. Another noteworthy prerequisite that must be considered when clinically applying NGS is the risk of the full genome analysis of off-target human DNA in samples being sequenced due to the techniques' sensitivity. The aforementioned approaches to bacterial enrichment will hopefully mitigate this problem [71].

To optimize the clinical usage of WGS in AMR detection, some have proposed the potential usage of Oxford Nanopore Sequencing in conjunction with the NGS technology [74, 77] to reduce the error susceptibility and improve the repeats resolution of the latter. These so-called hybrid assemblies are typically accomplished with SPAdes [78] or MaSURCA [79], and the results are often polished with Pilon [80]. Wrappers combining these tools into an assembly pipeline are available, and the foremost among them being UniCycler [81, 82]. npScarf can do the hybrid assembly as the data are being produced off the MinION/Flongle [83], allowing parsimonious reuse of flow cells. Another strategy that has been formerly deployed is to polish the long-reads with the short read information before assembly with PBCr [84], Nanocorr [84] or NaS [85], though the need for such correction has been greatly diminished by improvements in the ONT-provided base-caller Guppy and nanopore chemistry since 2017.

For use at point of care, long-read-only assembly is the most practical strategy [86], with a number of software packages now being able to effectively deal with modern nanopore raw read error rates of ~5%. Popular long-read-only assemblers include canu [87], shasta [88], Flye [89], wtdbg [90], raven [91], miniasm [92], reviewed in [93], and ABruijn [94]. With 100 \times coverage (easily achievable with a Flongle for most microbes), most tools provide single contig assembly for the genome and good elucidation of plasmids within minutes. These assemblies can be polished with racon [95], medaka (ONT) or nanopolish [96]. The practicalities of some of these tool combinations are explored in Senol et. al's 2019 study [97]. In practice, the vast majority of errors that remain after long-read assembly and polishing of recent nanopore-only data are in homopolymer stretches of greater than 6 bases with the commonly used R9.4.1 pore, though even this is mitigated with newer R10.3 pores that have a larger contact interface, allowing accurate homopolymer sequencing up to 10 bases. These errors can also be mitigated in R9.4.1

by training the deep neural network base-callers in a species-specific manner [98], though this is counter to the wide-range microbial detection that is sought in point-of-care applications. If homopolymers are miscalled relative to reference microbial gene, it is reasonable to assume the reference homopolymer length is correct in most cases, where with reference-biased correction, we would err on the side of overpredicting rather than underpredicting AMR. Popular software for assessing quality of assemblies enumerate potential frameshift and the presence of essential one-copy genes, and these software include BUSCO [99], CheckM [100] or Quast [101].

Once the patient sample is sequenced with this method, the sequence could be compared to reference sequences of known bacterial strains and AMR genes for real-time diagnosis [48]. Reference sequences for AMR gene detection are growing in variety, given the expanding library of AMR gene databases such as the Comprehensive Antibiotic Resistance Database (CARD), RAST and ResFinder. The detection of AMR genes in a similar fashion were demonstrated in Lemon *et al.*'s [48] study involving MinION ID of plasmid AMR genes with polishing from MiSeq to provide accurate reads. It is noteworthy that the study focused on sequencing plasmids. However, with technological improvements in ONT and informatics, the entire bacterial genome including plasmids can be sequenced for AMR conferring genes or mutations in real time. Among others, common bioinformatic tools that could be used for this workflow are listed in Table 6.

Many of the tools mentioned in Table 6, such as ABRicate, SRST2, SSTAR, Resistance Gene Identifier (RGI), MEGARes and ARIBA, possess the ability to screen a customizable database that can be made of multiple resistance gene databases as they are based on Basic Local Alignment Search Tool (BLAST). Thus, the screening against all available resistance gene databases can be automated into one workflow. The other tools take different approaches. For instance, AMRfinder uses gene prediction to find novel AMR genes, which is advantageous for new discoveries and to address poor results from aforementioned programs. KmerResistance also provides an opportunity for poor quality reads and samples with limited reads to have clinical value, as it does not exclusively depend on sequence identity when comparing against a database.

The aforementioned EPI2ME pipeline holds large merit in its optimized bacterial ID capabilities through WIMP; however, the EPI2ME pipeline can also identify AMR genes in identified sample bacteria through the ARMA—AMR tool. ARMA is a workflow integrated with CARD analyzing FASTQ data for AMR-conferring genes found within the bacterial genomes of species identified with WIMP [56]. With growing databases in both RefSeq and CARD, the EPI2ME pipeline for ONT technologies such as the MinION could be ideal for diagnosing a variety of bacterial infections and for optimizing treatment options depending on the AMR genes present. We propose that the usage of the EPI2ME pipeline in addition to Flongle and VolTRAX technologies could largely automate patient sample preparation processes, considerably reduce turnaround times and attenuate the complexity of convoluted workflows—upholding its potential point-of-care clinical applicability and utility.

Adjusting patient-specific antibiotic therapy

Conventionally applied tools for AMR screening are time consuming and resource heavy [102]. These conventional methods of AMR detection in bacterial isolates have lengthy turnaround times of weeks and are limited by the panel of antibiotics that are used to determine MIC in accordance with sensitivity or

resistance to antibiotics [102, 103]. The usage of the Flongle in AMR determination through the EPI2ME workflow would avoid the lengthy turnaround time and identify the resistance to certain antibiotics as well as associated resistance mechanisms in AMR bacterial samples. Consequently, antibiotic susceptibility could be deduced from the sequencing results of AMR genes and could be used to treat patients accordingly. PATRIC is one of a number of growing databases that can help inform users of antibiotic therapies based on the pathogen species and the AMR genes present [18]. Another growing database that can be used to specify treatment and avoid the use of broad-spectrum antibiotics is the National Database of Antibiotic Resistant Organisms (NDARO). The NDARO aims to increase standardization and make AMR-related data more widely available by collecting the genetic and antibiotic susceptibility data. Another database that can be used to evaluate potential treatments is the Sing group's AMCN. The AMCN is a novel database system allowing users to see information on the efficacies and concentrations of certain drug combinations against a particular bacteria or AMR gene [51].

Through bioinformatics, pipelines can be created to streamline the output of EPI2ME to pathogen and antimicrobial databases that can help inform clinical point-of-care use. This automated pipeline can help inform clinicians of possible antibiotic treatments and can be used in conjunction with other approaches to help improve approaches. As an example of this use, we can take a look at a common case of a patient with cellulitis. If WIMP identifies *Staphylococcus aureus* as a pathogen present, PATRIC suggests 17 possible antibiotics that are used to treat the infection caused by *S. aureus*. If ARMA detects that *mecA*, a common AMR gene of *S. aureus* is present, the automated pipeline can denote methicillin as a poor treatment option. Additionally, the user can utilize AMCN to deduce the most efficacious combination of potential drug treatment to optimize the recovery time, drug usage and the patient quality of life. In this instance, we take two drug candidates, Vancomycin and Linezolid, which were suggested from PATRIC. AMCN shows us the percent combinations of these drugs in combination with other drugs that can provide the most synergistic effect against *S. aureus*.

This information is very valuable in providing the most effective treatment option and for avoiding the use of broad-spectrum antibiotics.

Comparison to current approaches

Ultimately, the approach is designed to take less than 3 h from sample collection to AMR detection, bacterial ID and antibiotic treatment suggestion. Depending on the kit applied, microbiome DNA isolation will take approximately 1 h, with an additional hour for library preparation. The sequencing, bacterial ID, AMR detection and antibiotic information are all provided in real time. For the entire microbiome sequence, we estimate that it will likely take less than 1 h [104]. In terms of cost, the Flongle is listed at \$1800 USD, and the VolTRAX V2 is priced at \$8150 USD. Thus, total machinery costs are under \$10 000 USD. The costs per sample vary depending on the kit used and are less than or equal to \$150 USD.

In terms of resolution, this approach offers the highest resolution and theoretically the greatest accuracy so long as the microbiome reads are accurate and reliable. Regarding time and cost optimization, this approach has a faster turnaround time than all of the current fastest commercial techniques available for AST (identified in Table 1), with the exception of Biofire.

Table 6. Bioinformatic tools that could be utilized within the proposed workflow

Tool	Summary	Matching technique	Advantages and disadvantages
ABRicate	Tool used for mass screening of contigs for AMR or virulence genes with eight pre-downloaded databases [130]	Uses BLAST [131] to screen whole genome against multiple databases	Can be bundled with multiple databases: NCBI, CARD, ARG-ANNOT, Resfinder, MEGARES, EcoH, PlasmidFinder, Ecoli_VF and VFDB.
RGI	Software to predict resistomes, based on protein or nucleotide data, as well as metagenomics data, based on homology and SNP models [110]	(i) Predict protein sequences using Prodigal [132]	Only supports contigs and not FASTQ reads—only detects acquired resistance genes and not point mutations Specifically designed for assembly contigs or whole genomes—effective based on CARD's canonical reference sequences. Has a false-negative rate due to a lack of sequence diversity found in CARD. Also contains false-positives for a specific epidemic strain of <i>P. aeruginosa</i> due to gaps in CARD curation [134]
SRST2	Short-read sequence typing for bacterial pathogens [135], which includes various database versions from Table 4 and others	(ii) Homolog detection using DIAMOND [133] for protein models. (iii) BLASTP for variant protein models (iv) BLASTN for rRNA mutation models (v) BLASTP for protein overexpression models Performs of genes and alleles direct from short sequencing reads using any sequence database(s) and calculates combinatorial sequence types defined in MLST-style databases	Limitations for SRST2 include: paired-end reads have to be in the FR orientation, mate-pair samples may not be supported, poor sample alignment with reference gene databases can produce errors in result outputs and all samples must contain UNIQUE sample names
AMRFinder	Tool that identifies AMR genes using either protein annotations or nucleotide sequence [136]	Uses a database of protein-based HMMs (gene prediction files) to find novel AMR genes	As of right now, does not include the tools necessary to analyze adaptive resistance mutations like point mutations in rRNA genes or promoter-related mutations
KmerResistance	Correlates mapped genes with the predicted species of WGS sample [137, 138]	Examines the co-occurrence of k-mers between the WGS data and a database of resistance genes	Performs better than other known methods when data are contaminated or contain small amounts of sequence reads [137]. Cannot identify or interpret SNP variants that confer resistance—must identify resistance from specific genes or predefined alleles
SSTAR	AMR gene predictor [139]	Combines a locally executed BLASTN search against a customizable database with an intuitive graphical user interface for identifying AMR (AR) genes from genomic data	Can seamlessly apply any AMR database of interest for reference comparisons and can manually specify/add genes that may impact resistance [139]
MEGARes	Contains sequence data for approximately 8000 hand-curated AMR genes accompanied by an annotation structure that is optimized for use with high throughput sequencing [140]	Uses a translated BLAST search and USEARCH [140] to search annotations made through a unique hierarchical structure against a hand-curated MEGARES database for AMR genes	Does not provide detailed and comprehensive gene descriptions and multiple category annotations in order to provide simpler and hierarchical annotation schemes instead [140]
Ariba	AMR gene ID By assembly directly from short reads [141]	Uses BLAST to match reference genes against the assembled genome from short reads	Highly customizable and can be seamlessly compared with phenotypic resistance data. However, not recommended to be used with long-read technologies such as Oxford Nanopore. Samples can only have one gene per reference cluster—not suggested for metagenomic data [141]

Note. FR: Forward; HMM: hidden markov model.

However, the machinery cost of Biofire is considerably higher than a nanopore-based approach and is more suited to a centralized testing approach, with inherent longer turnaround times due to the logistics of sample movement. A key advantage to the proposed approach is the avoidance of centralized hospital machinery. Most hospitals and clinical laboratories usually have only system for bacterial ID and AST, for instance, hospitals and labs often share systems such as the MicroScan Walkaway and the MALDI-TOF MS. A downside to this type of design is that there is often a long waiting time on a patient-to-patient basis for the result. With a variety of cases depending on the machine, this type of approach increases the amount of time the infection has to occur. With the use of the handheld Flongle, this can be avoided with a Flongle for each case, and the subsequent analysis being performed on modern laptops. This point-of-care diagnosis approach is critical in terms of lessening turnaround times.

This approach can be recreated using other sequencing technologies and different bioinformatic downstream analysis programs. For instance, one could use the Illumina sequencing data and the previously mentioned LiveKraken analysis, a real-time read classification that uses streams of raw data from Illumina sequencers to classify reads taxonomically [57]. However, the ONT data can report greater taxonomic resolution than Illumina reads classification, making it more advantageous in this approach [104]. Another similar design can use PacBio data and sppIDer [111], a species classification tool designed for working with PacBio data, however, this combination is not intended to function in real time.

Some of the limitations of this approach include the degree of information available on databases such as PATRIC and NDARO. However, these databases are constantly growing with new AMR genes, bacterial species and antibiotic susceptibility information. Additionally, the Oxford Nanopore Sequencing technologies are relatively recent and will improve in terms of speed, accuracy and price as their popularity increases. Additionally, the performance of lysis methods may be subject to several factors including sensitivity of the target microorganism to the lysis [106]. Further research will be needed to optimize lysis reagents to improve the workflow. Furthermore, this approach is somewhat dependent on the level of background reads from the human host in comparison microbial reads [106]. This can have an impact on the sensitivity and time length of the method. Additionally, this type of approach to AMR detection is prone to false-negative results [106], however, this approach would be considered a frontline ID method that could deal with a significant number of cases. In the case of negative frontline sequence-based testing, existing, established practices should be used for ID. The deployment of sequencing in the clinical context requires it to be considered as part of larger systems, including validation and regulatory standards [106].

Conclusions

Avoiding the use of broad-spectrum antibiotics is critical in improving the therapies and for preventing the production of AMR bacteria. Generalizing and extending the reviewed examples, we propose a pipeline framework using ONT (Flongle, VoITRAX V2 and EPI2ME), third-party software and databases (assembly and AMR ID) and DNA extraction processes, which should aid in the exigent diagnosis and treatment of bacterial infections. This approach could improve on a variety of current diagnostic techniques in terms of turnaround time, cost and accuracy. The main improvements over NGS in microbial

applications are the: (1) automation and clinical point-of-care potential use that this pipeline presents and (2) ability to analyze sequence data as they are produced (i.e. within seconds rather than hours/days).

While all the bioinformatics and hardware pieces of the clinical microbiology sequencing puzzle exist and have been enumerated here, we have also highlighted additional bioinformatics components that would aid in maximizing the utility of the approach: (1) systematically combining the results of multiple read-to-species classification tools to reduce false-positives and (2) automated and frequent consolidated updates to species and AMR gene ID databases.

Key Points

- Commonly used methods for clinical infection diagnostics rely on high capital cost equipment and require centralized processing, sample batching and/or skilled labor.
- The emerging availability of automated sample preparation and low capital cost DNA sequencing equipment (nanopore-based) will enable diagnostic testing closer to point of care by combining OEM tools with third-party software and databases.
- Sequencing-based clinical diagnostics have been demonstrated repeatedly at small scale by researchers.
- Sequencing-based diagnosis leverages large (and ever-growing) public catalogs of both existing clinical isolates and AMR genes.
- Quality control mechanisms established for whole microbial genome sequencing can provide indication of diagnostic quality to point-of-care sequencing users.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Conflict of interest

Drophen Sheka, Nikolay Alabi, and Paul M.K. Gordon declare that they have no conflict of interest.

References

1. Kumar A, Roberts D, Wood KE, et al. Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. *Crit Care Med* 2006;**34**(6):1589–96.
2. Eldin C, Parola P, Raoult D. Limitations of diagnostic tests for bacterial infections. *Med Et Malad Infect* 2019;**49**(2):98–101.
3. Nahid P, Pai M, Hopewell P. Advances in the diagnosis and treatment of tuberculosis. *Proc Am Thorac Soc* 2006;**3**(1):103–10.
4. Weng Q, Raff A, Cohen J, et al. Costs and consequences associated with misdiagnosed lower extremity cellulitis. *JAMA Dermatol* 2017;**153**(2):141–6.
5. Filice G, Drekonja D, Thurm J, et al. Diagnostic errors that lead to inappropriate antimicrobial use. *Infect Cont Hospital Epidemiol* 2015;**36**(8):949–56.

6. Shrestha N, Adhikari M, Pant V, et al. Melioidosis: misdiagnosed in Nepal. *BMC Infect Dis* 2019;**19**(1):176.
7. Loonen AJ, Wolffs PF, Bruggeman CA, et al. Developments for improved diagnosis of bacterial bloodstream infections. *Eur J Clin Microbiol Infect Dis* 2014;**33**(10):1687–702.
8. Zhang Y, Hu A, Andini N, et al. A 'Culture' shift: application of molecular techniques for diagnosing polymicrobial infections. *Biotechnol Adv* 2019;**37**(3):476–90.
9. Maugeri G, Lychko I, Sobral R, et al. Identification and antibiotic-susceptibility profiling of infectious bacterial agents: a review of current and future trends. *Biotechnol J* 2019;**14**(1):e1700750.
10. Leclercq R, Canton R, Brown DF, et al. EUCAST expert rules in antimicrobial susceptibility testing. *Clin Microbiol Infect* 2013;**19**(2):141–60.
11. Loonen AJ, Bos MP, van Meerbergen B, et al. Comparison of pathogen DNA isolation methods from large volumes of whole blood to improve molecular diagnosis of bloodstream infections. *PLoS One* 2013;**8**(8):e72349.
12. Pulido MR, García-Quintanilla M, Martín-Peña R, et al. Progress on the development of rapid methods for antimicrobial susceptibility testing. *J Antimicrob Chemother* 2013;**68**(12):2710–7.
13. McElvania Tekippe E, Shuey S, Winkler DW, et al. Optimizing identification of clinically relevant gram-positive organisms by use of the Bruker Biotyper matrix-assisted laser desorption ionization-time of flight mass spectrometry system. *J Clin Microbiol* 2013;**51**(5):1421–7.
14. Barnini S, Ghelardi E, Bruculeri V, et al. Rapid and reliable identification of gram-negative bacteria and gram-positive cocci by deposition of bacteria harvested from blood cultures onto the MALDI-TOF plate. *BMC Microbiol* 2015;**15**:124.
15. Fernández-Álvarez C, Torres-Corral Y, Santos Y. Use of ribosomal proteins as biomarkers for identification of *Flavobacterium psychrophilum* by MALDI-TOF mass spectrometry. *J Proteomics* 2017;**170**:59–69.
16. Stevenson M, Pandor A, Martyn-St James M, et al. Sepsis: the LightCycler SeptiFast test MGRADE®, SepsisTest™ and IRIDICA BAC BSI assay for rapidly identifying bloodstream bacteria and fungi—a systematic review and economic evaluation. *Health Technol Assess* 2016;**20**(46):1–246.
17. Syal K, Mo M, Yu H, et al. Current and emerging techniques for antibiotic susceptibility tests. *Theranostics* 2017;**7**(7):1795–805.
18. Wattam AR, Davis JJ, Assaf R, et al. Improvements to PATRIC, the all-bacterial bioinformatics database and analysis resource center. *Nucleic Acids Res* 2017;**45**(D1):D535–42.
19. Clarridge JE, III. Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. *Clin Microbiol Rev* 2004;**17**(4):840–62.
20. Sherry NL, Porter JL, Seemann T, et al. Outbreak investigation using high-throughput genome sequencing within a diagnostic microbiology laboratory. *J Clin Microbiol* 2013;**51**:1396–401.
21. Long SW, Williams D, Valson C, et al. A genomic day in the life of a clinical microbiology laboratory. *J Clin Microbiol* 2013;**51**:1272–7.
22. Koser CU, Ellington MJ, Cartwright EJ, et al. Routine use of microbial whole genome sequencing in diagnostic and public health microbiology. *PLoS Pathog* 2012;**8**:e1002824.
23. Applications of Clinical Microbial Next-Generation Sequencing: Report on an American Academy of Microbiology Colloquium held in Washington, DC, in April 2015. American Society for Microbiology at Washington (DC), 2016. <https://www.ncbi.nlm.nih.gov/books/NBK513764/> doi: 10.1128/AAMCol.Apr.2015. (12 November 2019, date last accessed).
24. Didelot X, Bowden R, Wilson DJ, et al. Transforming clinical microbiology with bacterial genome sequencing. *Nat Rev Genet* 2012;**13**:601–12.
25. Guinane CM, Cotter PD. Role of the gut microbiota in health and chronic gastrointestinal disease: understanding a hidden metabolic organ. *Therap Adv Gastroenterol* 2013;**6**:295–308.
26. Olsen RJ, Long SW, Musser JM. Bacterial genomics in infectious disease and the clinical pathology laboratory. *Arch Pathol Lab Med* 2012;**136**:1414–22.
27. Orsini M, Cuccuru G, Uva P, et al. Bacterial genomic data analysis in the next-generation sequencing era. *Methods Mol Biol* 2016;**1415**:407–22.
28. Gu W, Miller S, Chiu CY. Clinical metagenomic next-generation sequencing for pathogen detection. *Annu Rev Pathol* 2019;**14**:319–38.
29. Frickmann H, Künne C, Hagen RM, et al. Next-generation sequencing for hypothesis-free genomic detection of invasive tropical infections in poly-microbially contaminated, formalin-fixed, paraffin-embedded tissue samples—a proof-of-principle assessment. *BMC Microbiol* 2019;**19**:75.
30. Dallman T, Byrne L, Launder N, et al. The utility and public health implications of PCR and whole genome sequencing for the detection and investigation of an outbreak of Shiga toxin-producing *Escherichia coli* serogroup O26:H11. *Epidemiol Infect* 2015;**143**(8):1672–80.
31. Lee SC, Billmyre RB, Li A, et al. Analysis of a food-borne fungal pathogen outbreak: virulence and genome of a *Mucor circinelloides* isolate from yogurt. *MBio* 2014;**5**(4):e01390.
32. Andrews-Polymenis HL, Santiviago CA, McClelland M. Novel genetic tools for studying food-borne *Salmonella*. *Curr Opin Biotechnol* 2009;**20**(2):149–57.
33. Drali R, Deriet A, Verhaegen B, et al. Whole-genome sequencing of *Listeria monocytogenes* serotype 4b isolated from ready-to-eat lentil salad in Algiers, Algeria. *New Microbes New Infect* 2019;**33**:100628.
34. Pietzka A, Allerberger F, Murer A, et al. Whole genome sequencing based surveillance of *L. monocytogenes* for early detection and investigations of listeriosis outbreaks. *Front Public Health* 2019;**7**:139.
35. Rutvisuttinunt W, Chinnawirotpisan P, Simasathien S, et al. Simultaneous and complete genome sequencing of influenza A and B with high coverage by Illumina MiSeq Platform. *J Virol Methods* 2013;**193**(2):394–404.
36. Quainoo S, Coolen J, van Hijum S, et al. Whole-genome sequencing of bacterial pathogens: the future of nosocomial outbreak analysis. *Clin Microbiol Rev* 2017;**30**(4):1015–63.
37. Wenger AM, Peluso P, Rowell WJ, et al. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nat Biotechnol* 2019;**37**(10):1155–62.
38. Volden R, Palmer T, Byrne A, et al. Improving nanopore read accuracy with the R2C2 method enables the sequencing of highly multiplexed full-length single-cell cDNA. *Proc Natl Acad Sci U S A* 2018;**115**(39):9726–31.
39. Stoddart D, Heron AJ, Mikhailova E, et al. Single-nucleotide discrimination in immobilized DNA oligonucleotides

- with a biological nanopore. *Proc Natl Acad Sci U S A* 2009;**106**(19):7702–7.
40. Manrao EA, Derrington IM, Pavlenok M, et al. Nucleotide discrimination with DNA immobilized in the MspA nanopore. *PLoS One* 2010;**6**(10):e25723.
 41. Okoniewski M, Ploski R, Wiewiorka M, et al. Clinical applications for next-generation sequencing. In: *Future Directions*. Cambridge, Massachusetts, United States: Academic Press 2016, 281–94.
 42. Matsumoto Y, Kinjo T, Motooka D, et al. Comprehensive subspecies identification of 175 nontuberculous mycobacteria species based on 7547 genomic profiles. *Emerg Microbes Infect* 2019;**8**(1):1043–53.
 43. Leggett RM, Alcon-Giner C, Heavens D, et al. Rapid MinION profiling of preterm microbiota and antimicrobial-resistant pathogens. *Nat Microbiol* 2019;**5**(3):430–42.
 44. Kerkhof LJ, Dillon KP, Häggblom MM, et al. Profiling bacterial communities by MinION sequencing of ribosomal operons. *Microbiome* 2017;**5**(1):116.
 45. Roguet A, Eren AM, Newton RJ, et al. Fecal source identification using random forest. *Microbiome* 2018;**6**(1):185.
 46. Schriefer AE, Cliften PF, Hibberd MC, et al. A multi-amplicon 16S rRNA sequencing and analysis method for improved taxonomic profiling of bacterial communities. *J Microbiol Methods* 2018;**154**:6–13.
 47. Kamathewatta KI, Bushell RN, Young ND, et al. Exploration of antibiotic resistance risks in a veterinary teaching hospital with Oxford Nanopore long read sequencing. *PLoS One* 2019;**14**(5):e0217600.
 48. Lemon JK, Khil PP, Frank KM, et al. Rapid nanopore sequencing of plasmids and resistance gene detection in clinical isolates. *J Clin Microbiol* 2017;**55**(12):3530–43.
 49. Jain M, Olsen H, Paten B, et al. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol* 2016;**17**(1):239.
 50. Ashton PM, Nair S, Dallman T, et al. MinION nanopore sequencing identifies the position and structure of a bacterial antibiotic resistance island. *Nat Biotechnol* 2014;**33**(3):296–300.
 51. Jorge P, Pérez-Pérez M, Rodríguez GP, et al. A network perspective on antimicrobial peptide combination therapies: the potential of colistin, polymyxin B and nisin. *Int J Antimicrob Agents* 2017;**49**(6):668–76.
 52. Golparian D, Donà V, Sánchez-Busó L, et al. Antimicrobial resistance prediction and phylogenetic analysis of *Neisseria gonorrhoeae* isolates using the Oxford Nanopore MinION sequencer. *Sci Rep* 2018;**8**(1):17596.
 53. Gardy J, Loman N. Towards a genomics-informed, real-time, global pathogen surveillance system. *Nat Rev Genet* 2018;**19**(1):9–20.
 54. Allix-Béguec C, Arandjelovic I, Bi L, et al. CRyPTIC consortium and the 100,000 Genomes Project. Prediction of susceptibility to first-line tuberculosis drugs by DNA sequencing. *N Engl J Med* 2018;**379**(15):1403–15.
 55. Ackerman AL, Anger JT, Khaliq MU, et al. Optimization of DNA extraction from human urinary samples for microbiome community profiling. *PLoS One* 2019;**14**(4):e0210306.
 56. Charalampous T, Kay GL, Richardson H, et al. Nanopore metagenomics enables rapid clinical diagnosis of bacterial lower respiratory infection. *Nat Biotechnol* 2019;**37**(7):783–92.
 57. Tausch SH, Strauch B, Andrusch A, et al. LiveKraken—real-time metagenomic classification of illumina data. *Bioinformatics* 2018;**34**(21):3750–2.
 58. Sabat AJ, van Zanten E, Akkerboom V, et al. Targeted next-generation sequencing of the 16S-23S rRNA region for culture-independent bacterial identification—increased discrimination of closely related species. *Sci Rep* 2017;**7**(1):3434.
 59. Lu J, Salzberg SL. Ultrafast and accurate 16S microbial community analysis using Kraken 2. *bioRxiv* 2020; 3:27.
 60. McDonald D, Price M, Goodrich J, et al. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J* 2012;**6**:610–8.
 61. Cole JR, Wang Q, Fish JA, et al. Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res* 2014;**42**(Database issue):D633–42.
 62. Quast C, Pruesse E, Yilmaz P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 2013;**41**(1):590–6.
 63. Bolyen E, Rideout JR, Dillon MR, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* 2019;**3**:852–7.
 64. Schloss PD. Reintroducing mothur: 10 years later. *Appl Environ Microbiol* 2020;**86**(2):e02343–19.
 65. Kim D, Song L, Breitwieser FP, et al. Centrifuge: rapid and sensitive classification of metagenomic sequences. *Genome Res* 2016;**26**(12):1721–9.
 66. Cerveau N, Jackson DJ. Combining independent de novo assemblies optimizes the coding transcriptome for non-conventional model eukaryotic organisms. *BMC Bioinformatics* 2016;**525**(17).
 67. Zickmann F, Renard BY. IPred—integrating ab initio and evidence based gene predictions to improve prediction accuracy. *BMC Genomics* 2015;**16**(1):134.
 68. Oliveira AC, Bovolenta LA, Nachtigall PG. Combining results from distinct MicroRNA target prediction tools enhances the performance of analyses. *Front Genet* 2017;**8**(59):1–10.
 69. Revez J, Espinosa L, Albiger B. Survey on the use of whole-genome sequencing for infectious diseases surveillance: rapid expansion of European National Capacities, 2015–2016. *Front Public Health* 2017;**5**(347).
 70. Neo JPS, Tan BH. The use of animals as a surveillance tool for monitoring environmental health hazards, human health hazards and bioterrorism. *Vet Microbiol* 2017;**203**:40–8.
 71. Anjum M, Zankari E, Hasman H. Molecular methods for detection of antimicrobial resistance. *Microbiol Spectr* 2017;**5**(6).
 72. European Society of Clinical Microbiology and Infectious Diseases (ESCMID). Determination of minimum inhibitory concentrations (MICs) of antibacterial agents by agar dilution. *Clin Microbiol Infect* 2000;**6**(9):509–15.
 73. 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**(1):2–22.
 74. Goldstein S, Beka L, Graf J, et al. Evaluation of strategies for the assembly of diverse bacterial genomes using MinION long-read sequencing. *BMC Genomics* 2018;**20**(1):23.
 75. Langille MG, Hsiao WW, Brinkman FS. Detecting genomic islands using bioinformatics approaches. *Nat Rev Microbiol* 2010;**8**(5):373–82.

76. Shelburne SA, Kim J, Munita JM, et al. Whole-genome sequencing accurately identifies resistance to extended-spectrum β -lactams for major gram-negative bacterial pathogens. *Clin Infect Dis* 2017;**65**(5):738–45.
77. Vasudevan K, Ragupathi NKD, Jacob JJ, et al. Highly accurate single chromosomal complete genomes using IonTorrent and MinION sequencing of clinical pathogens. *Genomics* 2020;**112**(1):545–51.
78. Bankevich A, Nurk S, Antipov D, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 2012;**19**(5):455–77.
79. Zimin AV, Marçais G, Puiu D, et al. The MaSuRCA genome assembler. *Bioinformatics* 2013;**29**(21):2669–77.
80. Walker BJ, Abeel T, Shea T, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 2014;**9**(11):e112963.
81. Wick RR, Judd LM, Gorrie CL, et al. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* 2017;**13**(6):e1005595.
82. Wick RR, Judd LM, Gorrie CL, et al. Completing bacterial genome assemblies with multiplex MinION sequencing. *Microb Genom* 2017;**3**(10):e000132.
83. Cao MD, Nguyen SH, Ganesamoorthy D, et al. Scaffolding and completing genome assemblies in real-time with nanopore sequencing. *Nat Commun* 2017;**8**:14515.
84. Goodwin S, Gurtowski J, Ethe-Sayers S, et al. Oxford Nanopore sequencing, hybrid error correction, and de novo assembly of a eukaryotic genome. *Genome Res* 2015;**25**(11):1750–6.
85. Madoui MA, Engelen S, Cruaud C, et al. Genome assembly using Nanopore-guided long and error-free DNA reads. *BMC Genomics* 2015;**16**(1):327.
86. Berbers B, Saltykova A, Garcia-Graells C, et al. Combining short and long read sequencing to characterize antimicrobial resistance genes on plasmids applied to an unauthorized genetically modified bacillus. *Sci Rep* 2020;**10**:4310.
87. Koren S, Walenz BP, Berlin K, et al. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res* 2017;**27**(5):722–36.
88. Shafin K, Pesout T, Lorig-Roach R, et al. Efficient de novo assembly of eleven human genomes using PromethION sequencing and a novel nanopore toolkit. *BioRxiv* 2019;**38**:715722.
89. Kolmogorov M, Yuan J, Lin Y, et al. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* 2019;**37**(5):540–6.
90. Ruan J, Li H. Fast and accurate long-read assembly with wtdbg2. *Nat Methods* 2019;**17**:155–8.
91. Vaser R, Šikić M. Yet another de novo genome assembler. *BioRxiv* 2019;141–6.
92. Li H. Minimap and minimap: fast mapping and de novo assembly for noisy long sequences. *Bioinformatics* 2016;**32**(14):2103–10.
93. Wick RR, Holt KE. Benchmarking of long-read assemblers for prokaryote whole genome sequencing. *F1000Research* 2019;**8**:2138.
94. Lin Y, Yuan J, Kolmogorov M, et al. Assembly of long error-prone reads using de Bruijn graphs. *Proc Natl Acad Sci U S A* 2016;**113**(52):e8396–405.
95. Vaser R, Sović I, Nagarajan N, et al. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res* 2017;**27**(5):737–46.
96. Loman N, Quick J, Simpson J. A complete bacterial genome assembled de novo using only nanopore sequencing data. *Nat Methods* 2015;**12**(8):733–5.
97. Senol Cali D, Kim JS, Ghose S, et al. Nanopore sequencing technology and tools for genome assembly: computational analysis of the current state, bottlenecks and future directions. *Brief Bioinform* 2019;**20**(4):1542–59.
98. Wick RR, Judd LM, Holt KE. Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biol* 2019;**20**(1):129.
99. Seppey M, Manni M, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness. *Methods Mol Biol* 2019;**1962**:227–45.
100. Parks DH, Imelfort M, Skennerton CT, et al. Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 2014;**25**(7):1043–55.
101. Gurevich A, Saveliev V, Vyahhi N, et al. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 2013;**29**(8):1072–5.
102. Anjum M. Screening methods for the detection of antimicrobial resistance genes present in bacterial isolates and the microbiota. *Future Microbiol* 2015;**10**(3):317–20.
103. Randall L, Lemma F, Rogers J, et al. Prevalence of extended-spectrum- β -lactamase-producing *Escherichia coli* from pigs at slaughter in the UK in 2013. *J Antimicrob Chemother* 2014;**69**(11):2947–50.
104. Cheng J, Hu H, Kang Y, et al. Identification of pathogens in culture-negative infective endocarditis cases by metagenomic analysis. *Ann Clin Microbiol Antimicrob* 2018;**17**(1):43.
105. Langdon QK, Peris D, Kyle B, et al. sppIDer: a species identification tool to investigate hybrid genomes with high-throughput sequencing. *Mol Biol Evol* 2018;**1**:2835–49.
106. Chiu CY, Miller SA. Clinical metagenomics. *Nat Rev Genet* 2019;**20**:341–55.
107. Beal S, Tremblay E, Toffel S, et al. A gastrointestinal PCR panel improves clinical management and lowers health care costs. *J Clin Microbiol* 2018;**56**(1):e01457–17.
108. Dhiman N, Hall L, Wohlfel SL, et al. Performance and cost analysis of matrix-assisted laser desorption/ionization-time of flight mass spectrometry for routine identification of yeast. *J Clin Microbiol* 2011;**49**(4):1614–6.
109. Kaleta E, Clark A, Cherkaoui A, et al. Comparative analysis of PCR-electrospray ionization/mass spectrometry (MS) and MALDI-TOF/MS for the identification of bacteria and yeast from positive blood culture bottles. *Clin Chem* 2011;**57**(7):1057–67.
110. Alcock BP, Raphenya AR, Lau TTY, et al. CARD 2020: antibiotic resistance surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res* 2020;**48**(D1):D517–25.
111. Gupta SK, Padmanabhan BR, Diene SM, et al. ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. *Antimicrob Agents Chemother* 2014;**58**(1):212–20.
112. Zankari E, Hasman H, Cosentino S, et al. Identification of acquired antimicrobial resistance genes. *J Antimicrob Chemother* 2012;**67**(11):2640–4.
113. Carattoli A, Zankari E, García-Fernández A, et al. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob Agents Chemother* 2014;**58**(7):3895–903.
114. Chen L, Yang J, Yu J, et al. VFDB: a reference database for bacterial virulence factors. *Nucleic Acids Res* 2004;**33**:D325–8.

115. Walker TM, Kohl TA, Omar SV, et al. Whole-genome sequencing for prediction of *Mycobacterium tuberculosis* drug susceptibility and resistance: a retrospective cohort study. *Lancet Infect Dis* 2015;**15**(10):1193–202.
116. Feuerriegel S, Schleusener V, Beckert P, et al. PhyResSE: a web tool delineating *Mycobacterium tuberculosis* antibiotic resistance and lineage from whole-genome sequencing data. *J Clin Microbiol* 2015;**53**(6):1908–14.
117. Shin J, Lee S, Go M, et al. Analysis of the mouse gut microbiome using full-length 16S rRNA amplicon sequencing. *Sci Rep* 2016;**6**(1):29681.
118. 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**(1):104–14.
119. Xia Y, Li AD, Deng Y, et al. MinION nanopore sequencing enables correlation between resistome phenotype and genotype of coliform bacteria in municipal sewage. *Front Microbiol* 2015;**2017**:8.
120. Moon J, Kim N, Lee HS, et al. *Campylobacter fetus* meningitis confirmed by a 16S rRNA gene analysis using the MinION nanopore sequencer, South Korea, 2016. *Emerg Microbes Infect* 2017;**6**(11):e94.
121. Tanaka M, Mino S, Ogura Y, et al. Availability of Nanopore sequences in the genome taxonomy for *Vibrionaceae* systematics: *Rumoiensis glade* species as a test case. *PeerJ* 2018;**6**(6):e5018.
122. Chalupowicz L, Dombrovsky A, Gaba V, et al. Diagnosis of plant diseases using the Nanopore sequencing platform. *Plant Pathol* 2019;**68**(2):229–38.
123. Leggett RM, Alcon-Giner C, Heavens D, et al. Rapid MinION profiling of preterm microbiota and antimicrobial-resistant pathogens. *Nat Microbiol* 2020;**5**(3):430–42.
124. Kai S, Matsuo Y, Nakagawa S, et al. Rapid bacterial identification by direct PCR amplification of 16S rRNA genes using the MinION™ nanopore sequencer. *FEBS Open Bio* 2019;**9**(3):548–57.
125. Mcmanus BA, O'Connor AM, Egan SA, et al. First description of arginine catabolic mobile element (ACME) type VI harboring the *kdp* operon only in *Staphylococcus epidermidis* using short and long read whole genome sequencing: further evidence of ACME diversity. *Infect Genet Evol* 2019;**71**:51–3.
126. Hamner S, Brown BL, Hasan NA, et al. Metagenomic profiling of microbial pathogens in the little Bighorn River. *Montana Int J Envir Res Public Health* 2019;**16**(7):1097.
127. Lim A, Naidenov B, Bates H, et al. Nanopore ultra-long read sequencing technology for antimicrobial resistance detection in *Mannheimia haemolytica*. *J Microbiol Methods* 2019;**159**:138–47.
128. Sakai J, Tarumoto N, Kodana M, et al. An identification protocol for ESBL-producing gram-negative bacteria bloodstream infections using a MinION nanopore sequencer. *J Med Microbiol* 2019;**68**(8):1219–26.
129. Bialasiewicz S, Duarte TPS, Nguyen SH, et al. Rapid diagnosis of *Capnocytophaga canimorsus* septic shock in an immunocompetent individual using real-time Nanopore sequencing: a case report. *BMC Infect Dis* 2019;**19**:660.
130. Seemann T. Abricate, Github. <https://github.com/tseemann/abricate>. (29 November 2019, date last accessed).
131. Camacho C, Coulouris G, Avagyan V, et al. BLAST+: architecture and applications. *BMC Bioinformatics* 2019;**10**:421.
132. Hyatt D, Chen GL, Locascio PF, et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC bioinformatics* 2010;**11**:119.
133. Buchfink B, Xie C, Fast HDH. Sensitive protein alignment using DIAMOND. *Nat Methods* 2015;**12**:59–60.
134. Freschi L, Jeukens J, Kukavica-Ibrulj I, et al. Clinical utilization of genomics data produced by the international *Pseudomonas aeruginosa* consortium. *Front Microbiol* 2015;**6**:1036.
135. Inouye M, Dashnow H, Raven L, et al. SRST2: rapid genomic surveillance for public health and hospital microbiology labs. *Genome Med* 2014;**6**:90.
136. Feldgarden M, Brover V, Haft DH, et al. Validating the AMRFinder tool and resistance gene database by using antimicrobial resistance genotype-phenotype correlations in a collection of isolates. *Antimicrob Agents Chemother* 2019;**63**(11):e00483–19.
137. Clausen PTLC, Zankari E, Aarestrup FM, et al. Benchmarking of methods for identification of antimicrobial resistance genes in bacterial whole genome data. *J Antimicrob Chemother* 2016;**71**(9):2484–8.
138. Clausen PTLC, Aarestrup FM, Lund O. Rapid and precise alignment of raw reads against redundant databases with KMA. *BMC Bioinformatics* 2018;**19**:307.
139. de Man TJB, Limbago BM. SSTAR, a stand-alone easy-to-use antimicrobial resistance gene predictor. *mSphere* 2016;**1**(1):e00050–15.
140. Doster E, Lakin SM, Dean CJ, et al. MEGARes 2.0: a database for classification of antimicrobial drug, biocide and metal resistance determinants in metagenomic sequence data. *Nucleic Acids Res* 2018;**48**(D1):D561–9.
141. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 2010;**26**:2460–1.