



Long-Reads-Based Metagenomics in Clinical Diagnosis With a Special Focus on Fungal Infections

Minh Thuy Vi Hoang^{1,2}, Laszlo Irinyi^{1,2,3}, Yiheng Hu⁴, Benjamin Schwessinger⁴ and Wieland Meyer^{1,2,3,5*}

¹Molecular Mycology Research Laboratory, Centre for Infectious Diseases and Microbiology, Faculty of Medicine and Health, Sydney Medical School, Westmead Clinical School, The University of Sydney, Sydney, NSW, Australia, ²Westmead Institute for Medical Research, Westmead, NSW, Australia, ³Sydney Infectious Disease Institute, The University of Sydney, Sydney, NSW, Australia, ⁴Research School of Biology, Australia National University, Canberra, ACT, Australia, ⁵Westmead Hospital (Research and Education Network), Westmead, NSW, Australia

OPEN ACCESS

Edited by:

Marcos Dias Pereira,
Federal University of Rio de Janeiro,
Brazil

Reviewed by:

Leho Tedersoo,
Tartu University Hospital, Estonia
Sandip Paul,
JIS Institute of Advanced Studies
and Research, India

*Correspondence:

Wieland Meyer
wieland.meyer@sydney.edu.au;
wieland.meyer@curtin.edu.au

Specialty section:

This article was submitted to
Microbial Physiology and Metabolism,
a section of the journal
Frontiers in Microbiology

Received: 12 May 2021

Accepted: 03 December 2021

Published: 06 January 2022

Citation:

Hoang MTV, Irinyi L, Hu Y,
Schwessinger B and Meyer W (2022)
Long-Reads-Based Metagenomics in
Clinical Diagnosis With a Special
Focus on Fungal Infections.
Front. Microbiol. 12:708550.
doi: 10.3389/fmicb.2021.708550

Identification of the causative infectious agent is essential in the management of infectious diseases, with the ideal diagnostic method being rapid, accurate, and informative, while remaining cost-effective. Traditional diagnostic techniques rely on culturing and cell propagation to isolate and identify the causative pathogen. These techniques are limited by the ability and the time required to grow or propagate an agent *in vitro* and the facts that identification based on morphological traits are non-specific, insensitive, and reliant on technical expertise. The evolution of next-generation sequencing has revolutionized genomic studies to generate more data at a cheaper cost. These are divided into short- and long-read sequencing technologies, depending on the length of reads generated during sequencing runs. Long-read sequencing also called third-generation sequencing emerged commercially through the instruments released by Pacific Biosciences and Oxford Nanopore Technologies, although relying on different sequencing chemistries, with the first one being more accurate both platforms can generate ultra-long sequence reads. Long-read sequencing is capable of entirely spanning previously established genomic identification regions or potentially small whole genomes, drastically improving the accuracy of the identification of pathogens directly from clinical samples. Long-read sequencing may also provide additional important clinical information, such as antimicrobial resistance profiles and epidemiological data from a single sequencing run. While initial applications of long-read sequencing in clinical diagnosis showed that it could be a promising diagnostic technique, it also has highlighted the need for further optimization. In this review, we show the potential long-read sequencing has in clinical diagnosis of fungal infections and discuss the pros and cons of its implementation.

Keywords: pathogenic fungi, long-read sequencing, metagenomics, identification, diagnosis, mycoses

INTRODUCTION

Rapid and accurate diagnosis of pathogens is essential in the management of infectious disease and is the ultimate goal for clinical microbiology laboratories. Successful diagnosis ensures effective treatment, with patient outcome improved the faster the pathogen is identified (Perfect, 2013). The ideal diagnostic test would be capable of reliably identifying any potential pathogen, provide additional information, such as the antimicrobial resistance profile, and offer the potential for further epidemiological analysis, while remaining cost-effective (Wickes and Wiederhold, 2018). The current available identification techniques for fungal pathogens are unable to meet all these criteria. The main drawback of the wide range of traditional identification techniques is that they rely in general on the *in vitro* growth of the causative pathogen before its diagnosis (Irinnyi et al., 2016).

Sequence-based metagenomics is a powerful culture independent tool for the identification of mixed microbial communities, regardless of the ability of member organisms to be grown *in vitro* (Miller and Chiu, 2018). Recently, developed long-read sequencing technologies generate sequencing reads far longer than previous sequencing technologies and are capable of addressing the shortcomings of conventional identification, inaccurate, time consuming, requires specific expertise, and short-read based metagenomics, which is unable to detect all pathogens when based on short DNA fragments (Tedersoo et al., 2021).

This review explores previous and potential use of long-read sequencing in the clinical diagnosis, focusing on fungal infections.

NON-SEQUENCING-BASED IDENTIFICATION TECHNIQUES

Many identification techniques have been developed and are currently used in clinical diagnosis, public health, animal welfare, plant protection, quarantine implications, and many other industrial fields (Figure 1). These conventional diagnostic tools vary in their success depending on the type of sample. However, despite advancements in molecular identification techniques of pathogens, these traditional methods still are an important component of routine diagnosis.

Morphology

Morphology-based methods rely on the identification of phenotypic features of the pathogens at the macroscopic or microscopic level. Physical characteristics can be observed within patient samples through tissue staining. However, growth on media is often required for diagnosis (Guarner and Brandt, 2011; Austin, 2017). Fungal pathogens may be identified based upon simple characteristics, such as the color, size, and smell of the isolates. Often, these do not allow for complete diagnosis, as these features may be shared among many species. More in-depth observations

are required, including the analysis of the complexity of morphology, the presence of sexual reproductive structures, the effect on surrounding media, and the effect of different staining techniques (Kurtzman et al., 2011; Schelenz et al., 2015).

Biochemical Methods

Commercial diagnostic tests based on characteristic biochemical profiles produced by fungi are available for diagnostic use. These tests provide a positive or negative result for metabolic, enzymatic, or fermentation products produced by specific fungal species and so are suitable for confirmatory testing after a suspected pathogen has been found through more broad range methods (Chen et al., 2014; Lackner et al., 2014; Schelenz et al., 2015; Wagner et al., 2017).

Serology

Serology-based tests identify the presence of antigens and antibodies within patient serum indicating current and previous infections (Im et al., 2019). Serological tests for fungal infections include testing for the presence of Galactomannan antigen for cases of invasive Aspergillosis (Pfeiffer et al., 2006; de Heer et al., 2019) and (1-3)- β -d-glucan (BDG) antigen testing for diagnosis of Candidemia and *Pneumocystis jirovecii* Pneumonia (Karageorgopoulos et al., 2011; White et al., 2017).

Targeted PCR Amplification

Targeted amplification of DNA regions using species-specific primers by PCR has been widely employed for the identification of fungal pathogens (Kidd et al., 2019). A wide array of primers targeting the ITS1/2 region of the rDNA gene cluster has been designed for specific fungal species, including the prominent pathogenic fungal species *Candida*, *Aspergillus*, and *Fusarium* (Carvalho et al., 2007; Springer and Loffler, 2017). Additional regions have also been targeted by species-specific primers, e.g., the D1/D2 region of the large ribosomal subunit (LSU) of the rDNA gene cluster is used for the identification of *Candida* species (Mannarelli and Kurtzman, 1998) and for *P. jirovecii* (Tia et al., 2012). The latest targeted amplification-based identification method is the T2Candida Panel (T2 Biosystems), which is a new diagnostic tool for rapid, sensitive, and specific detection of invasive candidiasis from whole blood without culturing or nucleic acid extraction (Neely et al., 2013). It relies on PCR amplification and T2 Magnetic Resonance which enables nanoparticle-mediated rapid detection of candidemia with a sensitivity of one colony forming unit (CFU)/ml of whole blood (Mylonakis et al., 2015; Clancy et al., 2018). The use of species-specific PCR is limited in fungal diagnostics, as it is reliant on prior suspicion of the fungal pathogen to select the correct PCR target, primers, and amplification conditions (Wickes and Wiederhold, 2018).

Matrix-Assisted Laser Desorption/Ionization Time of Flight Mass Spectrometry

Matrix-assisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF MS) is a high-throughput, reliable technology used to identify and analyze large biomolecules

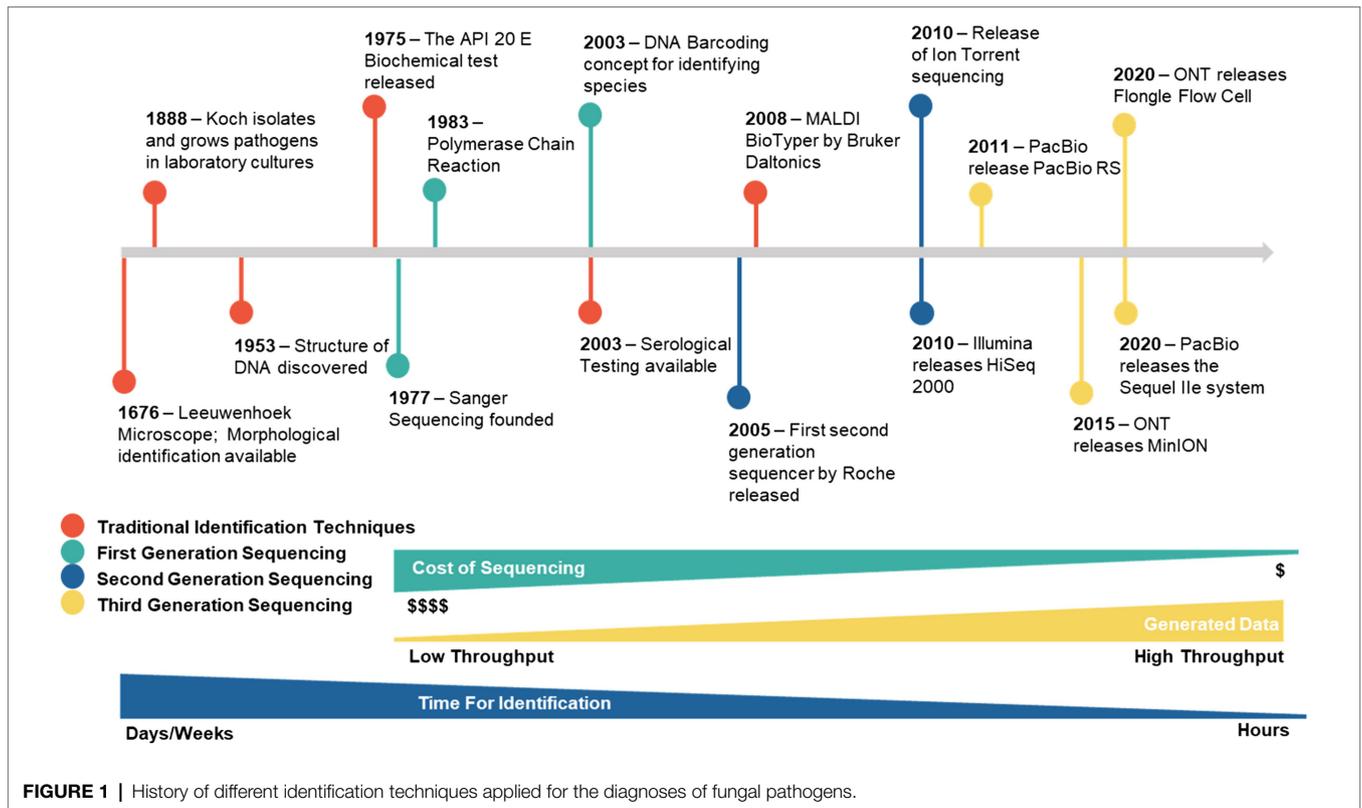


FIGURE 1 | History of different identification techniques applied for the diagnoses of fungal pathogens.

(Fenselau and Demirev, 2001; Clark et al., 2013). Starting material for MALDI-TOF can be found within sterile patient samples with high microbial load; however, this is not commonly possible and so prior pathogen growth is required. The processed material is then suspended in a crystalline matrix, vaporized, and ionized by a laser. Charged particles are then separated in a high voltage field, and the time of flight of the particles is measured and recorded as mass spectrum. Organism identification is obtained *via* the comparison of the obtained mass spectra with a reference database, which is currently the limiting factor (Angeletti, 2017).

Traditional identification techniques largely rely on prior growth of pathogens before testing may commence. This puts a strain on diagnosis as some fungal species may require up to 4 weeks before discriminatory features can be observed (Irinyi et al., 2016). Closely related species and species complexes, which may have different responses to antifungals, are unable to be reliably identified as characteristic phenotypic traits are not displayed and can only be genetically distinguished, such as in the genera *Fusarium*, *Aspergillus*, and *Scedosporium* (Goldstein et al., 2015; De Hoog et al., 2019). Traditional identification methods have been shown to be more insensitive than more advanced techniques, as results are highly subjective and error prone, leading to potential misidentifications (Table 1; Irinyi et al., 2016; Gustafson et al., 2019). Additionally, a number of fungal species, such as *P. jirovecii*, cannot be cultured under laboratory conditions and so culture-based methods are inappropriate for the identification (White et al., 2017). MALDI-TOF most closely meets the criteria for an ideal

diagnostic test; however, it remains reliant on culturing and still lacks an exhaustive curated database. Although MALDI-TOF has been demonstrated to be more cost-efficient than traditional culturing techniques, the initial instrument cost limits widespread use in smaller diagnostic laboratories (Patel et al., 2017).

Overall, non-sequencing-based identification techniques are inadequate as the gold standard diagnostic technique for fungal infections.

SEQUENCING-BASED IDENTIFICATION TECHNIQUES

The advancement in the understanding of the genetic information of microorganisms has been fundamental for the development of molecular diagnosis techniques of pathogenic fungi. These techniques include PCR - restriction fragment length polymorphism analysis (Dendis et al., 2003), random amplified polymorphic DNA analysis (Brandt et al., 1998; Mobasherizadeh et al., 2016), hybridization with genus/species specific DNA or RNA probes (Sandhu et al., 1995; Lindsley et al., 2001), PCR-fingerprinting (Lieckfeldt et al., 1993), species-specific PCR assays (Martin et al., 2000; Kulik et al., 2004), real-time PCR (Klingspor and Jalal, 2006), and increasingly DNA sequencing (Pryce et al., 2003; Romanelli et al., 2010). DNA sequencing introduced by Sanger et al. (1977) dominated the sequencing technology for 20 years, being recognized as first-generation sequencing. When applied to the identification of

TABLE 1 | Features of diagnostic tests available for fungal infections.

Diagnostic test	Available since	Culture dependency	Need for single culture	Test turnaround time without culture time	Accuracy	Curated database needed	Expertise	Cost	Portability	Resolution power	Potential for virulence and drug resistance detection	Generated clinical information
Non-sequence-based	Morphology	Yes	Yes	Long	Low	No	High	Low	No	Medium	No	Low
	Biochemical	Yes	Yes	Long	Low	Yes	Low	Medium	No	Medium	No	Low
	Serological	Yes	No	Short	Low	No	Low	Low	No	None	No	Low
	Targeted PCR	No	No	Short	High	No	Low	Low	No	High	No	Low
Sequence-based	amplification (T2Candida panel)											
	MALDI-TOF	Yes	Yes	Short	High	Yes	Low	Low	No	High	No	Low
	PCR based assays	Yes	No	Medium	High	No	Medium	Low	No	High	No	Medium
	Sanger sequencing/DNA barcoding	Yes	Yes	Medium	High	Yes	Medium	Medium	No	High	No	Medium
	Whole genome sequencing	Yes	Yes	Long	High	Yes	High	High	No	High	Yes	Very high
Long-read metagenomics/ Metabarcoding	2015	No	No	Short	High	Yes	Medium	Low	Yes	High	Yes	Very high

microorganisms, it relies on the analysis of pathogen-specific genome regions for comparison to reference sequence databases.

DNA barcoding is one of the most promising and efficient sequencing-based methods enabling rapid identification of species and recognition of cryptic species. The concept was first proposed by Hebert et al. (2003). DNA barcodes are standardized, easily amplified, universal, short DNA sequences (500–800 bp), which show a high divergence at the species level. These allow for the rapid identification of an organism through comparison of the generated genetic barcode to a reference collection of DNA barcodes from well-identified species. Two barcoding regions have been established for use in the identification of pathogenic fungi. The primary fungal DNA barcoding region is the internal transcribed spacer region (ITS1/2; Schoch et al., 2012) and the secondary fungal DNA barcoding region is the *translational elongation factor 1α* (Stielow et al., 2015). DNA barcoding using Sanger sequencing has major inherent limitations, including: (1) the requirement of a single fungal organism in the sample, (2) high target amplicon yield to avoid biases and errors (Polz and Cavanaugh, 1998), and (3) intra-individual variability (heteroplasm). Especially, the high intra-genomic diversity of the ITS1/2 region in fungi cannot be simultaneously detected from one sample using Sanger sequencing alone, as samples with multiple amplicons will appear as mixed peaks in the sequencing chromatograms (O'Donnell and Cigelnik, 1997). The high intra-genomic diversity was only able to be detected after cloning (Jumpponen, 2003; Simon and Weiss, 2008) and the application of Next-Generation Sequencing (NGS) technologies (Ganley and Kobayashi, 2007; Colabella et al., 2021).

Sequencing technologies have recently rapidly advanced. The early NGS technologies were classified as second-generation sequencing technologies and achieved increased sequencing throughput after the amplification of thousands of DNA templates and the simultaneous sequencing of the resulting DNA fragments, saving time, and decreasing costs (Heather and Chain, 2016; Ravi et al., 2018). The most prominent short-read sequencing technology is those produced by Illumina. While Sanger sequencing suffered from low throughput, it did produce long-reads (up to 1,000 bp) with high accuracy. Short-read NGS technologies focused on achieving high throughput while sacrificing read length (100–600 bp; van Dijk et al., 2014; Illumina, 2021). The high throughput of NGS allowed for in-depth sequencing, recovering more data and thus more species from complex samples (Shokralla et al., 2014).

The introduction of sequence-based identification methods improved the speed and accuracy of fungal diagnostics and the rapid advancements of sequencing technologies enables the identification of fungi in complex samples.

METABARCODING AND METAGENOMICS

Metabarcoding/metagenomic sequencing is a method in which nucleic acid (DNA or RNA) of all organisms in a sample is extracted and sequenced using NGS techniques. The generated

sequences are then used to identify organisms present in the sample. It was first mentioned in 1998 (Handelsman et al., 1998), when it was applied for the culture independent analysis of a complex and diverse (“meta”) community of microorganisms.

There are two main approaches which can be used to characterize the microbiome of a specific sample: (i) **targeted amplicon sequencing (metabarcoding)** and (ii) **shotgun metagenomics**. While the amplicon sequencing is based on the amplification of well-characterized genetic regions, shotgun metagenomics attempts to sequence the entire genetic content in a sample. The detailed differences between the two approaches have been previously discussed elsewhere (Forbes et al., 2017).

Targeted amplicon sequencing (metabarcoding) enables the direct identification of multiple species simultaneously from any environmental or clinical sample without culturing. It has been extensively used in microbiome studies and its biggest advantage over shotgun sequencing is the low amount of microbial DNA needed for identification and targeted enrichment of the sequence of interest. It is based on the amplification of specific taxonomic genetic regions (barcodes), such as the 16S ribosomal RNA (rRNA) gene cluster for bacteria (Riesenfeld et al., 2004; Janda and Abbott, 2007) or the internal transcribed spacer region (ITS1/2; Nilsson et al., 2009; Schoch et al., 2012) for fungi, by universal degenerated primers in a PCR reaction. These barcodes are ubiquitous in varying copy numbers, can be easily amplified, and have two fundamental characteristics: high taxonomic coverage and high resolution (Schoch et al., 2012). This approach combines the concept of DNA barcoding (Hebert et al., 2003) and the application of high-throughput sequencing technology (metabarcoding). The generated barcodes are then computationally clustered by sequence similarity into operational taxonomic units (OTUs) and queried against a reference database, such as UNITE¹ (Koljalg et al., 2005), BOLD² (Ratnasingham and Hebert, 2007), RefSeq³ (Schoch et al., 2014), and the ISHAM Barcoding Database⁴ (Irinnyi et al., 2015; Meyer et al., 2019). The targeted approach only sequences the amplified barcodes and not the whole genome. Due to the smaller size of the generated data, the computational analyses are less complex and faster. However, amplicon-based sequencing has PCR biases due to its inability to amplify all microorganisms across multiple taxa and it does not provide any additional characterization beyond taxonomic information.

Short-read sequencing has been widely used in metabarcoding studies; however, it only targets short barcodes, such as the ITS1 or ITS2 region for fungi, as second-generation sequencer read lengths (500–600bp for paired reads) are unable to span the full length of the established barcodes, the entire ITS1/2 region (Edouard et al., 2018; Hamad et al., 2018; Tsang et al., 2021). In addition, short-reads are unable to resolve repeated sequences that are longer than the reads generated by short-read NGS and as such these may lead to misassemblies and gaps (Salzberg and Yorke, 2005; Treangen and Salzberg, 2011;

Teng et al., 2017). Furthermore, even the short-reads lend themselves to identifying single-nucleotide variants and short indels accurately, longer structural variations are more challenging to resolve, as the variation is different within the ITS1 or ITS2 regions (**Figure 2**), leading to different discriminatory powers (van Dijk et al., 2018). As a result, in some species, the ITS1 or the ITS2 alone cannot resolve all species, which will only be possible using the combined ITS1/2 region (Irinnyi et al., 2016). Longer sequencing reads would eliminate these issues and so drove the development of long-read NGS, see below.

In contrast, **shotgun metagenomic** sequencing allows for a higher resolution, as it sequences most parts of the genomes of every organism present in the sample. This feature enables the technology not only to identify the organism, but also to characterize their extended profiles, such as antimicrobial resistance, genetic subtypes, and virulence. It involves extraction of nucleic acid (DNA or RNA) from primary specimens (e.g., soil samples, water samples, blood, feces, bronchoalveolar lavage (BAL), and sputum, tissue), fragmentation of the obtained DNA or RNA extracts, library preparation, in-depth sequencing, and subsequent data analysis. Shotgun metagenomics is significantly more expensive (depending on the sequencing depth) and computationally more extensive than amplicon sequencing. Another challenge is the overwhelming amount of host or background DNA compared to the microbial DNA. Many methods have been developed for enriching microbial DNA to combat this for both environmental and clinical samples, see below. Additionally, in cases of non-sterile patient samples or co-infections, genomic material from multiple microbial species may prove difficult to assign correctly to the particular species present (Bharti and Grimm, 2021).

The application of shotgun metagenomics to diagnostics eliminates the need for culturing as identification would be made directly from patient samples (e.g., tissue, blood, feces, BAL, and sputum). As previously discussed, culturing greatly limits the detection of causative pathogens, as not all pathogens can be grown in a laboratory setting and it also greatly delays the timing of accurate identification (Garrido-Cardenas and Manzano-Aguigliaro, 2017). In addition, metagenomics eliminates any sequencing bias that may occur as a result of PCR amplification (Wang et al., 2015). Comparing to amplicon sequencing, shotgun metagenomics also allows for better estimation of the relative abundance of microbes in samples based on the number of reads generated (Laudadio et al., 2018). Most importantly, shotgun metagenomics data contain additional genomic information from the patient samples, potentially informing not only on pathogen identities, but also underlying host factors. Previously established antimicrobial markers may also be identified from metagenomic data and as such can be used to guide targeted treatment, leading to the reduction of inappropriate use of antimicrobials and a subsequent reduction in the development of antimicrobial resistance. Additionally, further genomic information can be generated, to be used for epidemiological studies and outbreak tracing *via* the utilization of multi-locus sequence typing data generated as part of the metagenomics dataset (Couto et al., 2018).

¹<https://unite.ut.ee/>

²<https://www.boldsystems.org/>

³<https://www.ncbi.nlm.nih.gov/refseq/>

⁴<https://its.mycologylab.org/>

The quantity and quality of data generated from metagenomic studies is reliant upon the NGS technology used and as such the advantages and disadvantages must be considered. The characteristics of the main technologies used for the identification of fungal pathogens are summarized in Table 1.

LONG-READ SEQUENCING TECHNOLOGIES

Third-generation sequencing technologies are characterized by the generation of ultra-long-reads and offer a number of advantages over short-read sequencing. Long-read sequencing routinely generates reads over 10 kb (Pollard et al., 2018; Mantere et al., 2019). Currently, commercial long-read sequencing is supported by two companies, Pacific Biosciences and Oxford Nanopore Technologies.

Pacific Biosciences

Pacific Biosciences (PacBio) released the first commercial long-read sequencer in 2011 and continue to maintain and release improved sequencing instruments (Ardui et al., 2018; Tedersoo et al., 2021). The basis of Pacific Biosciences sequencers is known as single molecule real-time sequencing (SMRT), which takes place in single use SMRT Cells. These contain multiple Zero-mode waveguiders (ZMWs) that hold an immobilized polymerase. Double-stranded DNA (dsDNA) is prepared for sequencing by forming SMRT bells by adding hairpin adaptors to both ends of the dsDNA template (Eid et al., 2009). Each SMRT bell is diffused into one ZMW and nucleotides with identifying fluorescent labels are added to the ZMW alongside reagents required for PCR. From there, the immobilized polymerase binds to the hairpin adaptor of the SMRT bells and begins replication. As a nucleotide is added to the complementary strand, fluorescence at a characteristic wavelength is emitted and the sequence of fluorescence pulses is recorded into a movie. The wavelengths emitted are then converted into a nucleotide sequence called a continuous long-read. After the polymerase completes replication of one DNA strand, it continues to sequence the opposite adapter and second strand. As a result, it is possible to generate multiple passes of the same template depending on the lifetime of the polymerase. Upon sequencing completion and conversion to continuous long-reads, subreads are formed by identifying and cleaving the adaptor region sequences. Multiple computationally merged subreads of the same template DNA strand are called circular consensus sequences or HiFi reads, and the more subreads, the higher the accuracy, currently reaching up to Q30–50 for consensus HiFi (Rhoads and Au, 2015; Tedersoo et al., 2021).

PacBio currently have three sequencers on the market, the Sequel System, the Sequel II system, and the Sequel IIe system, released in October 2020 (Pacific Biosciences, 2020). The newer iterations Sequel II and Sequel IIe boast a new SMRT Cell, dubbed the SMRT Cell 8M, containing 8 million ZMWs capable of running for 30h and generating up to 4,000,000 reads with over 99% (Q20+) accuracy (Wenger et al., 2019; Pacific Biosciences, 2021a).

The Sequel IIe improves upon the Sequel II by upgrading the software and data analysis function through data processing capability on the instrument and cloud facilitation leading to potentially 70% reduction in secondary analysis time and 90% reduction in file transfer and data storage (Pacific Biosciences, 2020).

SMRT sequencers initially reported low read accuracy of 85–87%, when at the same time, short-read sequencers from Illumina had read accuracy >99% (Chin et al., 2013; Ardui et al., 2018). Errors in sequencing through insertions and deletions were more common in SMRT sequencing. However, improvements in sequencing chemistry and polishing methods increased the read accuracy of SMRT sequencing to 99.8% (Ardui et al., 2018; Tedersoo et al., 2018; Wenger et al., 2019). This allowed its use in direct sequencing studies rather than only be used to scaffold short-read Illumina reads (Parker et al., 2017; Teng et al., 2017; De Maio et al., 2019). The increased throughput of SMRT sequencers has made adequate coverage with multiplexed samples achievable (Armanhi et al., 2016; Pacific Biosciences, 2021b). The current SMRT Cell 8M supports the simultaneous sequencing of 348 samples when using the largest barcoding kit PacBio currently offers, and the sequencing depth can achieve reference quality for *de novo* assemblies for genomes up to 2GB of a single sample and full length 16S region sequences for 192 samples for strain level identification (Pacific Biosciences, 2021b).

Disadvantages of PacBio instruments are, that they require a high initial investment cost, limiting widespread use. The throughput of Illumina sequencers remains higher than PacBio, as PacBio sequencers are limited by the number of ZMWs available in the SMRT Cells. Not all ZMWs are guaranteed to successfully sequence, as there maybe more than one DNA fragment being present in a ZMW, and there maybe issues with the polymerase binding. The nature of SMRT bells results in a trade-off between read length and read quality as longer DNA inserts are sequenced less often and hence the generated consensus sequence is of lower quality (Tedersoo et al., 2018). Further, the library preparation for the Sequel II Instrument is estimated to be 1 day, far longer than needed for the Oxford Nanopore Technology library preparation kits (see below), which may not be suitable for time sensitive diagnostic applications.

Overall, Pacific Biosciences has successfully created sequencing instruments capable of generating high accuracy ultra-long-reads.

Oxford Nanopore Technologies

The idea of nanopore-based technologies originated from the Coulter counter and ion channels (Cornell et al., 1997). The single molecule sequencing using biological nanopores was first proposed in 1995 and published in 1998 (Church et al., 1998). Kasianowicz et al. (1996) then described the detection of ssDNA passing through an α -hemolysin nanopore (2.6 nm in diameter; Kasianowicz et al., 1996). Oxford Nanopore Technologies (ONT) officially launched the concept of MinION™ in 2012 as a new generation nucleic acid sequencing technology based on nanopores, free from fluorescence labels and amplification requirements. It became commercially available in 2015 (Eisenstein, 2012; Deamer

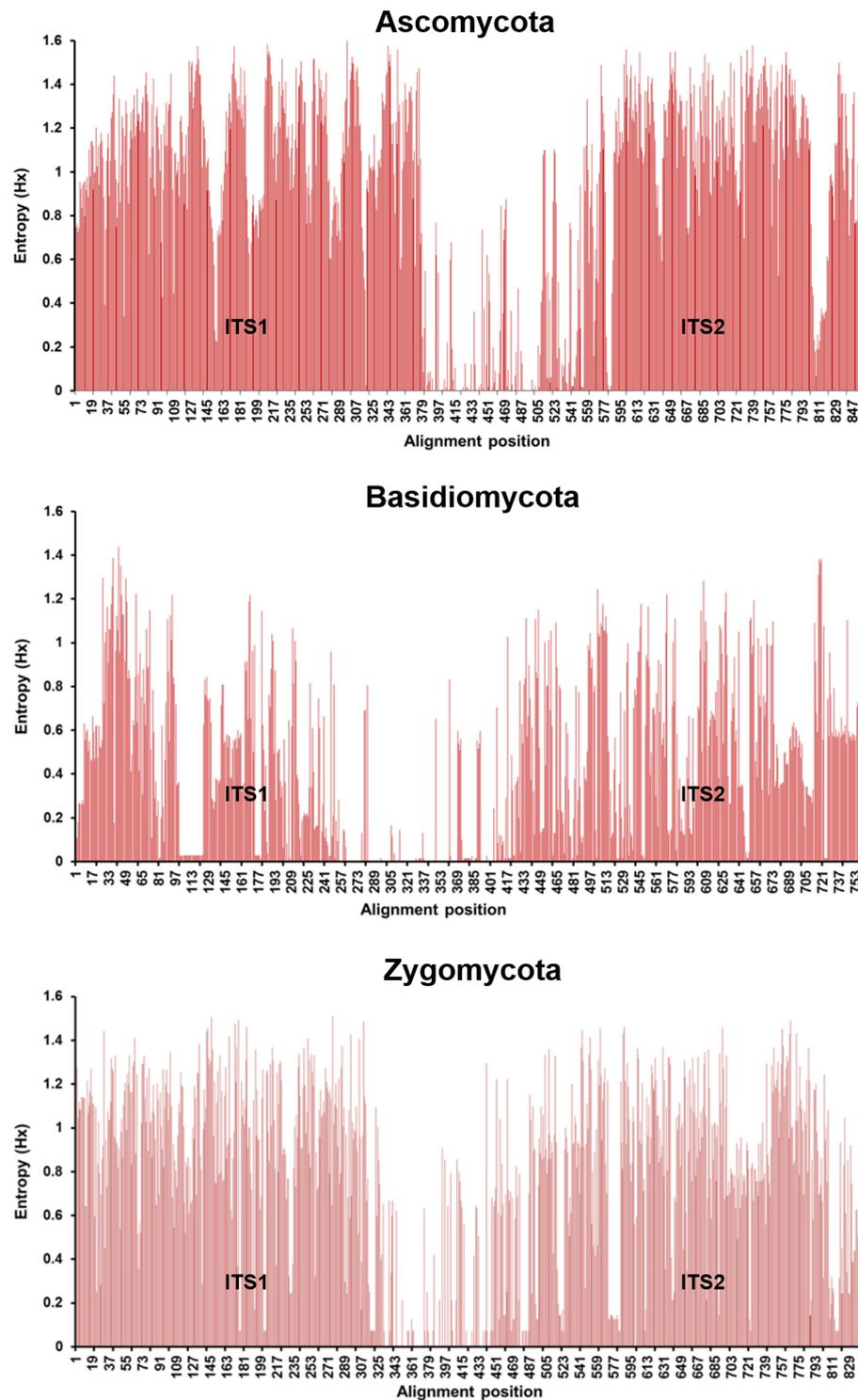


FIGURE 2 | Nucleotide entropy plots showing the variations in the ITS1/2 region for Ascomycota, Basidiomycota, and Zygomycota. The higher the bar is the greater the variation at that position in the nucleotide sequence. The variation of ITS1/2 is higher than that of the ITS1 or ITS2 regions alone and so indicates higher discriminatory power for identification. Entropy values were calculated using BioEdit 7.0.

et al., 2016). The MinION™ is a palm sized portable single molecule sequencing device capable of generating high-throughput, ultra-long sequence reads in real-time at relatively low cost.

The sequencing chemistry of ONT's sequencers has undergone rapid development with the latest flow cell released at the time of writing being the R10.4. However, the basis of nanopore

sequencing remains consistent, relying on biological nanopores embedded in solid state membranes within disposable flow cells (Leggett and Clark, 2017). Different sequencing library preparation methods are available depending on the intended use and data requirements. However, in all library preparation methods, a leader adapter and motor protein are ligated to the end of dsDNA fragments. This library is then loaded into the flow cells and the motor protein functions to unzip dsDNA and process the DNA strand throughout the pore at a specific speed. After one strand is processed, the pore is available to sequence the next available strand. Nucleotides are processed through the biological pores in groups of four called k-mers and as these travel through the pore, a change in an ionic current across the membrane occurs. These changes are then recorded as signal traces and during base calling, are converted to nucleotide sequences (Jain et al., 2016).

ONT's MinION flow cell contains 4x512 nanopore channels, each containing one nanopore, with 512 channels being accessed at a given time. These are run on both the MinION, which supports one flow cell at a time, and the GridION, which supports five flow cells. The PromethION is ONT's production-scale sequencer and provides 3,000 nanopore channels for each flow cell (Oxford Nanopore Technologies, 2021). All instruments can be run until the flow cells no longer have any pores available for sequencing and can be stopped when the intended sequencing depth is achieved. Additionally, the Flongle is a MinION adapter that enables the use of Flongle flow cells that contain 126 nanopore channels at a reduced cost (Oxford Nanopore Technologies, 2019).

The primary appeal of ONT sequencers over other current sequencing technologies is that read length is mostly dependent on sample DNA length, besides high DNA quantity and purity. It has been previously demonstrated to generate read lengths of over 2 million base pairs (Payne et al., 2019). ONT sequencers also offer reduced initial investment costs for sequencers when considering the MinION and the Flongle adapter although library preparation costs remain high (Tedersoo et al., 2021). Per sample sequencing costs depend on multiplexing, sequencing coverage, and re-use of flow cells after washing with a DNase. This makes sequencing available to a wider range of small routine laboratories when other sequencers may only be available for large laboratories or sequencing center. The availability of higher throughput instruments allows nanopore sequencing to be scaled to many different purposes. Multiplexing kits of up to 96 samples are available to further scale projects down if less in-depth sequencing is needed or a lower cost-per sample is necessary. ONT sequencers are highly suited for use in unconventional settings, as field library preparation kits are available that reduce library preparation time to 10 min with limited laboratory equipment. The sequencers are portable as the MinION and Flongle adapters are palm sized, allowing sequencing to be done outside the laboratory, such as at disease outbreak hotspots, war zones or for real-time environmental monitoring. This has previously been demonstrated through its use in space and resource poor settings (Quick et al., 2016; Castro-Wallace et al., 2017). The major drawback of nanopore sequencing has been the low read accuracy being reported

with 90% with the R9.4 flow cell and earlier models (Lu et al., 2016). The introduction of new flow cells has led to strong improvements, with average accuracies of 87–98% being commonly reported (Loit et al., 2019; Logsdon et al., 2020). The highest accuracy ever reported after bioinformatics polishing, which might be computation-intensive and may take weeks, stands at >99% (Ashikawa et al., 2018; Morrison et al., 2020). In addition, new bioinformatic tools to improve the accuracy of nanopore reads are now readily available (Loman et al., 2015; Koren et al., 2017; Salmela et al., 2017; Xiao et al., 2017; Hu et al., 2021). As sequencing quality and length depends on the quality of the applied DNA/RNA samples a major drawback is that impurities within the loaded library may block pores and render them unable to sequence.

Overall, nanopore sequencing is by now a broadly accessible long-read sequencing platform with read length limited only by the DNA input making it a fundamentally different approach for sequencing.

HOW CAN LONG-READ SEQUENCING BE APPLIED TO CLINICAL DIAGNOSIS

As long-read sequencing platforms have the potential to accurately identify pathogenic species and simultaneously provide additional genomic information impacting patient outcome, we are focusing the remaining part of the review on the ONT and PacBio platforms for their use in clinical diagnostics.

The read length of SMRT sequencers has advanced to where it is comparable to those generated by ONT although read length is determined by input DNA for both sequencers. When using these directly on patient samples without prior amplification, as in metagenomics, long-reads provide genetic information spanning the whole informative regions of pathogen genomes. For example, DNA barcoding regions, which are between 500 and 800bp, cannot be covered by short-read sequencing read lengths but are now being able to be fully sequenced by the long-read sequencing instruments. It would also be possible to sequence antimicrobial genes present in the genomes of resistant microbes, informing treatment strategies to improve patient outcome. Long-reads may also allow accurate sequencing of loci important in surveillance or strain typing. In samples where pathogen material is low in abundance, a prior amplification step of important genetic regions would ensure the sequencing of these important target regions although potentially missing other information.

The application of metagenomics to clinical diagnosis requires reliable and highly accurate sequencing and both ONT and PacBio have advertised over 99% per read accuracy for their latest sequencing chemistries. For PacBio, this has been reflected in many recent studies, e.g., Wenger et al., 2019. The accuracy of ONT depends on the type of sequencing chemistry and base-calling algorithms used. The accuracy of ONT technologies has been significantly improving and high accuracy levels are targeted to be achieved in the near future. Although nanopore sequencing may not yet be suitable for

applications, such as single-nucleotide variant identification or genotyping, which requires high coverage to exclude false positives, however, the accuracy required for pathogen identification can already be achieved (Greig et al., 2019; Feng et al., 2021).

Sequencing requirements of diagnostic laboratories cannot be guaranteed to be constant and so this may result in different throughput requirements. As such, investments into high-throughput sequencing instruments, such as the PacBio instruments, are not suitable for routine diagnostic laboratories. However, an alternative may be the implementation of core diagnostic centers that service multiple smaller laboratories and the multiplexing of samples, which would provide an effective way to reduce the cost-per sample. However, this would result in waiting for an adequate number of samples to be pooled for sequencing, which is contradicting the fact that fast diagnosis being the most important factor for the best patient outcome. As such, the smaller scale sequencing offered by ONT with the Flongle flow cells allows for sequencing of individual samples when needed, which is more suitable for general routine laboratories. Additionally, the initial capital investment in sequencing instruments, such as the ones produced by PacBio, may not be feasible for smaller diagnostic laboratories, which only could be overcome by pooling resources into larger diagnostic centers. The smaller scale ONT sequencers are generally affordable, making them more suitable in low throughput settings, such as it is the case in the diagnostic setting for mycoses.

The proposed workflow of the use of long-read sequencers in clinical diagnosis illustrates the need for timely identification (Figure 3). The time from sample collection to identification must be as rapid as possible and so the selection of a sequencer and its sample preparation and sequencing time must be considered. PacBio sequencing reports library preparation time of 1 day, while the slowest ONT library preparation kit, the direct cDNA Sequencing Kit, has an estimated preparation time of 270 min. The shortest library preparation option for

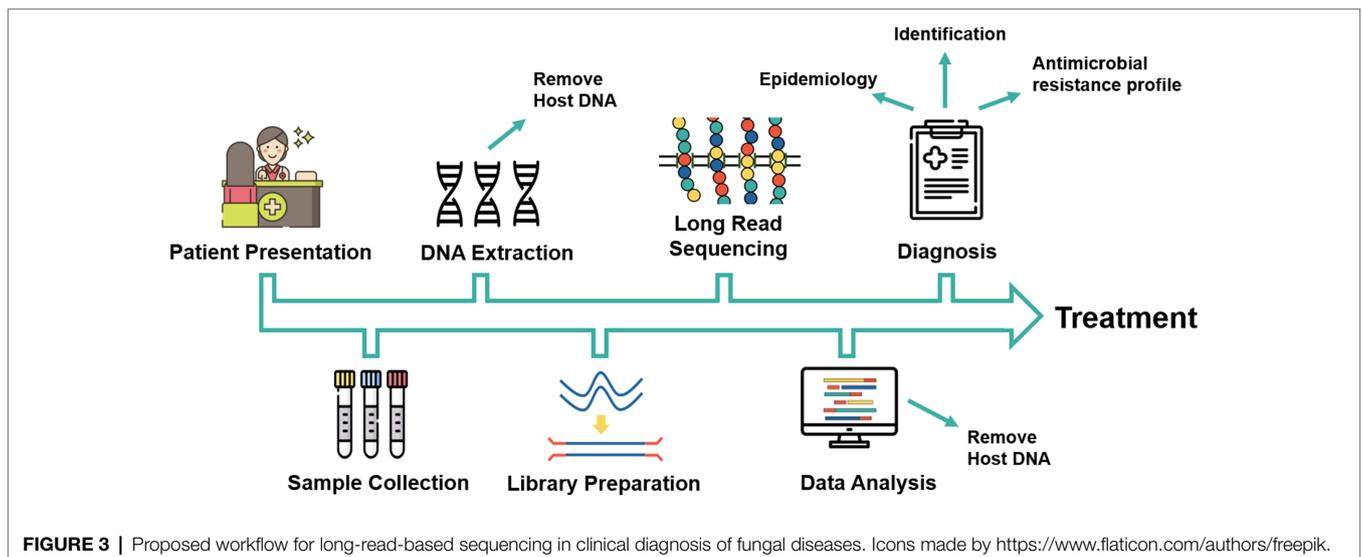
ONT sequencing is the Field Sequencing Kit (SQK-LRK001) and the Rapid Sequencing Kit (SQK-RAD004), which have a preparation time of only 10 min although these kits produce a lower throughput (Kafetzopoulou et al., 2018). There are also further efforts to reduce turnaround time by reducing sample preparation and sequencing times (Loit et al., 2019). Further sample preparation may be added, which increases preparation times, such as a prior amplification step and sample multiplexing. Both PacBio (Sequel IIe) and ONT provide the option for real-time sequencing, thus eliminating the need to wait for the sequencing to be completed before base-calling can occur. This allows for a substantial reduction of the time from sample collection to pathogen identification to less than 24h, enabling a faster introduction of the correct treatment, and improving patient outcome.

Before long-read sequencing can be applied to routine diagnostic testing, further work is required to determine the accuracy, specificity, and precision, of long-read sequencers in a clinical setting. The limit of detection must also be determined before it can be reliably implemented. In addition, adequate DNA isolation, library preparation, and loading according to standardized procedures performed by trained personnel are required to take full advantage of long-read technologies for routine diagnostics laboratories.

Long-read sequencers have many features that make them an attractive alternative for the use in clinical diagnosis of infectious disease agents.

LONG-READ SEQUENCING APPLIED TO INFECTIOUS DISEASE INVESTIGATIONS

The application of long-read sequencers to bacterial and viral species identification within a clinical setting is more advanced than that for fungal infections. The general smaller size of bacterial and viral genomes enables to use whole-genome



sequencing studies instead of or in parallel with metagenomics studies, with both employing shotgun sequencing methods without prior amplification. As such, whole-genome sequencing studies will be discussed below alongside metagenomics and metabarcoding studies using long-read sequencers.

Bacterial Infections

The ability of PacBio to identify bacteria using the full length 16S region to the species level was demonstrated by the study by Earl et al. in which 100% of a mock community of 20 bacterial species was identified to the species level (Earl et al., 2018). This was then extended to apply sequencing of a mock community of 250 bacterial species, of which over 90% were accurately identified to the species level. Finally, it was successfully applied to characterize the microbiome of six sinonasal sites of 12 subjects (Earl et al., 2018). The use of PacBio in identification of bacterial species through sequencing of the full length 16S rRNA gene was further demonstrated through generation of microbiome profiles for vaginal samples and sputum samples from patients with cystic fibrosis (Hahn et al., 2016; Wagner et al., 2016). PacBio long-reads have also been used in the generation of whole-genome reference sequences of clinical samples indicating its potential use for future metagenomic studies to identify pathogenic bacteria (Kanamori et al., 2016; El-Rami et al., 2019).

Metagenomics and metabarcoding studies using nanopore sequencers for bacterial samples have been widely used in clinical and research settings. Rapid and accurate diagnosis of a *Capnocytophaga canimorsus* infection in a 62-year-old female patient using metagenomic nanopore sequencing has been reported (Bialasiewicz et al., 2019). This study demonstrated a positive identification was achieved by nanopore sequencing within 19h, whereas conventional diagnostic techniques required 6.5 days to achieve the same. This short turnaround time from beginning of the sample preparation to pathogen identification is reflected in other studies, estimating a 6–8-h turnaround time (Schmidt et al., 2017; Charalampous et al., 2019; Gargis et al., 2019; Greig et al., 2019), and even using as little as 30 min of sequencing (Sakai et al., 2019; Taxt et al., 2020). Gu et al. (2021) also found a turnaround time of 6h for nanopore sequencing compared to 24h to Illumina sequencing. The study also demonstrated accurate identification of pathogenic bacterial species from a variety of clinical samples using metagenomic nanopore and Illumina sequencing. These samples also ranged from culture and PCR negative to PCR positive. In cases where NGS was unable to identify the causative pathogen, the cause was likely due to low pathogen titer or high levels of host background (Gu et al., 2021). An appeal of the use of metagenomics in clinical diagnostics is the potential for additional genomic information, such as antimicrobial resistance coding genes, to be obtained during sequencing. This has been achieved in studies for common bacterial pathogens, such as *Escherichia coli*, *Klebsiella pneumoniae*, *Mycobacterium tuberculosis*, and *Staphylococcus aureus* (Schmidt et al., 2017; Votintseva et al., 2017; Cheng et al., 2018; Charalampous et al., 2019; Gonzalez-Escalona et al., 2019; Taxt et al., 2020). Metagenomic sequencing of clinical bacterial samples has also revealed accurate

epidemiological information that can be used for surveillance and outbreak tracing (Votintseva et al., 2017; Gorrie et al., 2018; Imai et al., 2020). ONT sequencers have also been used in metabarcoding studies involving the full length 16S rRNA gene to characterize complex mock communities and identify causative pathogens of clinical samples (Kai et al., 2019; Moon et al., 2019; Nakagawa et al., 2019; Winand et al., 2019).

Many published studies applying long-read sequencing to the identification of pathogenic bacterial species in clinical samples strongly suggest its potential incorporation into routine bacterial diagnosis and encouraging its potential application in fungal diagnoses.

Viral Infections

PacBio sequencing has been used for whole-genome sequencing of a variety of human pathogenic viruses. A limitation of sequencing of viral genomes from clinical samples is the low viral load within cells. As such, reaching the required coverage to ensure an accurate consensus is a challenge. The small size of viral genomes lends themselves to overcome this issue through the amplification of the viral genome, which has been successfully used in whole-genome sequencing studies using PacBio, where the read lengths are able to span the whole amplicon length (Ocwieja et al., 2012; Dilernia et al., 2015; Bull et al., 2016; Nakano et al., 2017). PacBio sequencing has been demonstrated to be capable of the identification of viral pathogens from clinical samples but has not yet been used in routine viral diagnosis due to the high costs associated (Ho et al., 2017; Li et al., 2020). Identification of norovirus from patient samples has been performed to simulate diagnosis during an outbreak and demonstrated PacBio achieved sequencing of the entire viral pathogen, Norovirus GII, with high coverage of 99.11%; however, this was greatly limited by the long turnaround time (66h; Li et al., 2020). PacBio reads have also been utilized in sequencing SARS-CoV-2 from patient samples to study the evolution and progression of disease and inform antiviral treatment (Ko et al., 2021). PacBio has additionally been used to achieve deep sequencing of genomic regions of viruses that are unable to be resolved using short-read sequencers (Bergfors et al., 2016; Huang et al., 2016; Ho et al., 2017; Lui et al., 2019; Takeda et al., 2019; Yamashita et al., 2020). Although these studies do not aim to identify the pathogens, the regions targeted for deep sequencing inform on resistance to antiviral treatments.

ONT has been widely used for the identification of viral pathogens in disease diagnostics and surveillance. Like the PacBio studies, nanopore sequencing of viral genomes often first involves either the amplification of the full genome or the generation of overlapping pooled amplicons that span the entirety of the genome (Quick et al., 2017; Deng et al., 2020). Nanopore sequencing has been demonstrated to be a useful tool for the identification of major pathogens during outbreaks, such as for the Ebola epidemic in West Africa from 2014 to 2016 and to generate additional genomic information on geographic, climate, and demographic factors (Hoenen et al., 2016; Quick et al., 2016; Dudas et al., 2017). Nanopore sequencing

has also been used in the recent COVID-19 pandemic, e.g., for a prospective study of the complex transmission within healthcare settings through the generation of 747 SARS-CoV-2 genomes from PCR-positive diagnostic samples within 24h of sample collection (Meredith et al., 2020). Epidemiological information of the spread of a variant of SARS-CoV-2 was generated using nanopore sequencing to inform on further action needed to control its spread (Washington et al., 2021). ONT sequencers have additionally been used for whole-genome sequencing during the Zika virus epidemic (Faria et al., 2017), Yellow fever virus epidemic (Faria et al., 2018), and the Lassa fever outbreak in Nigeria (Kafetzopoulou et al., 2019). Whole-genome sequencing from clinical samples has also been demonstrated for human metapneumovirus (Xu et al., 2020), Chikungunya and Dengue viruses (Kafetzopoulou et al., 2018), Influenza (Lewandowski et al., 2019), and Hepatitis C virus (Greninger et al., 2015).

Long-read sequencing technologies have demonstrated their ability to be used to diagnose viral infections in the clinical setting and have a proven track record for analysis of samples during epidemics and a pandemic to produce valuable public health information.

Fungal Infections

The application of long-read sequencing to fungal identification in a clinical setting is currently limited.

Whole genomes of major fungal pathogens have been generated with the use of long-read sequencers often in combination with short-read sequencing, so called as hybrid assemblies (Cuomo et al., 2017; Luo et al., 2017; Vale-Silva et al., 2017; Panthee et al., 2018; Rhodes et al., 2018; Morand et al., 2019; Schultzhuis et al., 2019; Pchelin et al., 2020). The use of whole-genome sequencing in clinical practice is not yet feasible, as the cost and data processing requirements are currently too high for routine use.

In clinical applications, PacBio sequencers have been used to characterize the fungal composition of fecal samples from 14 healthy individuals (Motooka et al., 2017). The study identified multiple fungal species present within the healthy gut mycobiota and characterized two major mycobiota types. This study by Motooka et al. (2017) additionally found PacBio was able to accurately identify all 26 fungal species in a mock community through metabarcoding of the ITS1 region, where short-read sequencers could not. Additionally, PacBio most accurately estimated the abundance of species within the mock community. The predominant reason PacBio outperformed the short-read sequencers was the use of long-reads, ensuring complete and accurate coverage of the entire 300–800bp long ITS1 region (Motooka et al., 2017). PacBio has also been used for metabarcoding analysis of environmental fungal samples using the shorter ITS2 minibarcode and the large ribosomal subunit (28S), which are often used by short-read barcode studies (Cline and Zak, 2015; Kyaschenko et al., 2017). Sequencing of the full length ITS1/2 DNA barcoding region by PacBio sequencers has been demonstrated on a wide variety of fungal species (James et al., 2016; Walder et al., 2017; Heeger et al.,

2018; Tedersoo et al., 2018; Purahong et al., 2019; Tedersoo and Anslan, 2019). A comparison between PacBio and nanopore sequencers found PacBio to be more efficient in the metabarcoding of complex environmental fungal samples due to errors found in nanopore sequencing (Loit et al., 2019). Additionally, sample to identification time was reduced to 2.5h through metagenomic nanopore sequencing of plant samples (Loit et al., 2019). These studies indicate the further potential for long-read PacBio sequencing to be used in metabarcoding to identify causative pathogens in clinical samples.

The use of ONT in metabarcoding and metagenomics is similarly limited; however, direct clinical use has been previously demonstrated (Irinnyi et al., 2020; Gu et al., 2021). Three patient samples, positive for *P. jirovecii* infection, alongside three samples without infection underwent metagenomic sequencing using ONT's MinION flow cells (Irinnyi et al., 2020). The resulting reads were assigned using two different analysis tools, ONT's own "What's in my pot" and nucleotide BLAST, and the resulting assignments were compared. Sequencing of all three positive patient samples produced *P. jirovecii* reads while surprisingly, *Pneumocystis* reads were detected in negative samples when using the "What's in my pot" analysis tool; however, this is attributed to issues applying the analysis tool to fungal identification. *Homo sapiens* sequences accounted for 70–95% of all reads besides one outlier (10%). Additionally, the sample sites were bronchiolar lavage and induced sputum samples which are non-sterile sites and so multiple other fungal species were identified alongside the known causative pathogen. Notably, *Paracoccidioides* spp. were identified within the sample. However, this indicates a clear false positive, as *Paracoccidioides* spp. are geographically restricted to the tropical areas of Latin America, which did not align with the patient's travel history. This indicated that fungal diagnosis solely based off metagenomic sequencing using ONT has currently a number of limiting factors, including the low host:pathogen ratio, lack of comprehensive reference sequence databases, and the lack of appropriate bioinformatic tool, which need to be overcome before it can be applied in a routine diagnostic setting (Irinnyi et al., 2020).

Gu et al. (2021) evaluated the diagnostic accuracy and performance of nanopore and Illumina sequencing on clinical body fluid samples (Gu et al., 2021). This study sequenced 87 patient samples by both sequencing technologies to compare the sensitivity and specificity. For fungal pathogen detection, the sensitivity and specificity of nanopore sequencing were found to be 90.9 and 100%, respectively, for nanopore sequencing, compared to 90.6 and 89.0% for Illumina sequencing. Additionally, five patient samples with negative culturing and PCR testing results were analyzed using NGS and three samples were found to be positive for fungal pathogens. For two samples, NGS did not detect any pathogens. However, the fungal pathogens *Cryptococcus neoformans* and *Sporothrix schenckii* were identified in these samples through further testing, most likely due to low pathogen titers or high host background DNA in the sample. This study demonstrated that mNGS can be applied to infectious disease diagnosis and would be a valuable tool in the routine diagnostic

process (Gu et al., 2021). Even though this study achieved the identification of causative pathogens using ONT sequencing, it also highlighted again many issues currently associated with practical implementation into routine diagnostics, including the importance of the quality of the input DNA, the proportion of pathogen DNA to host DNA, and the lack of comprehensive reference databases (Gu et al., 2021). Furthermore,

Metabarcoding for the identification of fungal pathogens from clinical samples using nanopore sequencing has also been performed. The ITS1/2 region was amplified in a patient sample previously confirmed having a *Candida albicans* infection and the resulting amplicons were sequenced on the ONT MinION (Ashikawa et al., 2018). This resulted in 97.3% of the first 4,000 reads and 99.1% of reads generated after 48h being assigned to *C. albicans*, confirming the previous diagnosis. Another study used metabarcoding by nanopore sequencing to identify 16 members of an artificial fungal community using the ITS1/2 region (Mafune et al., 2020). The ITS1/2 region was also used to identify 87% of the fungal species present in 43 respiratory samples and identified others not detected through routine testing (Chan et al., 2020). Metabarcoding of 1,312 clinical respiratory samples with the bacterial 16S gene and fungal ITS1/2 region identified pathogens in 51.5% of samples with a turnaround time of <24h. Nanopore sequence alone identified 31% of samples, while 0.5% were identified with culture-based techniques only (Wang et al., 2020). Long amplicon nanopore sequencing of the full ITS region and the whole fungal operon successfully identified all members of a mock community and microbial cultures (D'Andrea et al., 2021). The full ITS region was also used to characterize the fungal microbiota of healthy and potentially infected clinical canine samples (D'Andrea et al., 2021). The fungal intergenic spacer region has also been demonstrated to identify members from the *Cryptococcus gattii* and *Cryptococcus neoformans* species complexes by multiplexing 24 strains on a single MinION flow cell. Sequencing errors in regions with homopolymers were observed. However, with the R10.3 flow cell, this error rate was reduced by 57% and the sequence identity increased to 99.83% (Morrison et al., 2020).

These preliminary studies indicate the potential for the application of ONT and PacBio sequencers to the clinical diagnosis of fungal infections if the above-mentioned limitations are overcome to develop a workflow for clinical diagnosis.

CURRENT ISSUES WITH LONG-READ SEQUENCING

Long-read sequencers have been demonstrated to effectively identify bacteria, viruses, and fungi, from clinical samples using both metagenomics and metabarcoding approaches. Although these studies have illustrated the potential for long-read sequencing in clinical diagnosis, a range of current challenges facing their implementation has also been highlighted.

DNA Extraction

Sequencing success depends on sample DNA quantity and quality. Pathogens are often present at low concentrations during infection and so high DNA concentrations are optimal to ensure adequate sequencing coverage of the causative pathogen for identification. Metagenomics and metabarcoding sequencing are aimed to represent everything within a sample and so ideally genomic material will be directly representative of what is present. As such, DNA extraction methods must be able to effectively lyse all cell types and preserve all genomic material.

For fungal pathogens, this presents a challenge, as fungal species have thick cell walls, containing polysaccharides and high content of lipids, that are difficult to lyse, which impedes the release of nucleic acids. As such fungal DNA extraction methods often involve a manual disruption step, such as bead bashing or crushing with liquid nitrogen to break open the fungal cell walls. This often leads to DNA shearing, greatly reducing the lengths of DNA fragments, defeating the purpose of long-read sequencing. A number of methods describing the extraction of DNA from fungal material for the use of nanopore and PacBio sequencing have reported to produce DNA fragments with an average fragment size above 35 kbp, although this has not yet been applied to clinical samples (Feehan et al., 2017; Schwessinger and Rathjen, 2017; Inglis et al., 2018). The development of better extraction methods is important to improve the application of long-read sequencers to accurate clinical diagnosis.

Ratio of Host to Pathogen DNA

A major problem in metagenomics analysis is the high amount of host (human) DNA in the sample, which may shadow the detection of pathogen DNA. One possible solution would be to increase the proportion of pathogen DNA in comparison with host DNA within the sample. This may be achieved through the depletion of human DNA in the clinical sample, which still is suboptimal even though there are several commercial kits and laboratory protocols for this purpose available (Marotz et al., 2018; Charalampous et al., 2019; Helmersen and Aamot, 2020; Heravi et al., 2020).

An alternative to this would be to increase the abundance of pathogen DNA, which can be achieved through amplification of important genomic regions, such as in metabarcoding studies, or amplification of overlapping DNA regions as done in viral whole-genome sequencing (Quick et al., 2017; Deng et al., 2020). An improved host to pathogen DNA ratio would greatly improve identification using long-read sequencing as it would enable higher coverage of the causative pathogen, leading to more accurate identification and the possibility of further genomic information. An increase in pathogen reads would also make long-read sequencing more cost-effective since more relevant data would be generated, possibly allowing for multiplexing of samples.

Another way to enrich pathogen DNA and deplete host DNA is the application of adaptive sequencing with nanopore sequencers. DNA molecules are selected for by identifying unwanted sequences and reversing the voltage across the membrane, thereby stopping

sequencing. Bioinformatic tools have been developed to implement adaptive sequencing for the selection of specific chromosomes and the depletion of bacterial genomes to enrich the remaining species in a metagenomic population (Kovaka et al., 2020; Payne et al., 2020). For diagnosis of mycoses, fungal DNA would ideally be selected for while human DNA is rejected; however, this is yet to be explored. The improvement of host to pathogen DNA ratio would greatly increase the amount of relevant data for clinical diagnosis and further work is required to implement these methods to routine diagnostics.

Lack of Reference Genomes

The major appeal of long-read sequencing in a shotgun approach is the vast genomic information that can be gained from the data generated. In metagenomic studies, genomic information from throughout the genome is generated; however, there is a lack of reference genomes for most of the pathogenic fungal species (Figure 4) and as such a proportion of data generated cannot be mapped to an appropriate reference genome either in FungiDB⁵ (Basenko et al., 2018) and GenBank⁶ (Sayers et al., 2020).

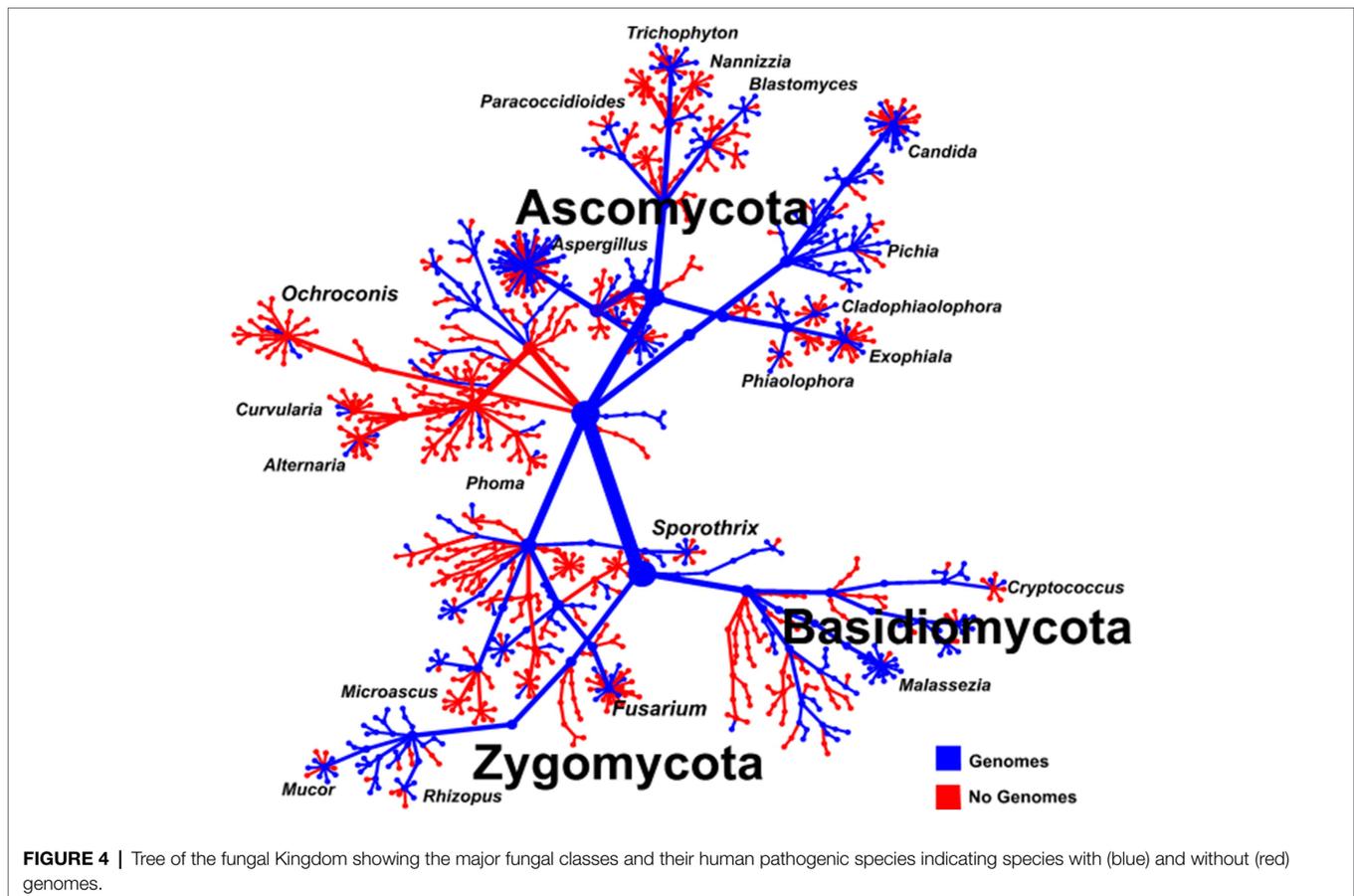
The basis for accurate genomic-based identification is the availability of high-quality, curated reference whole-genome

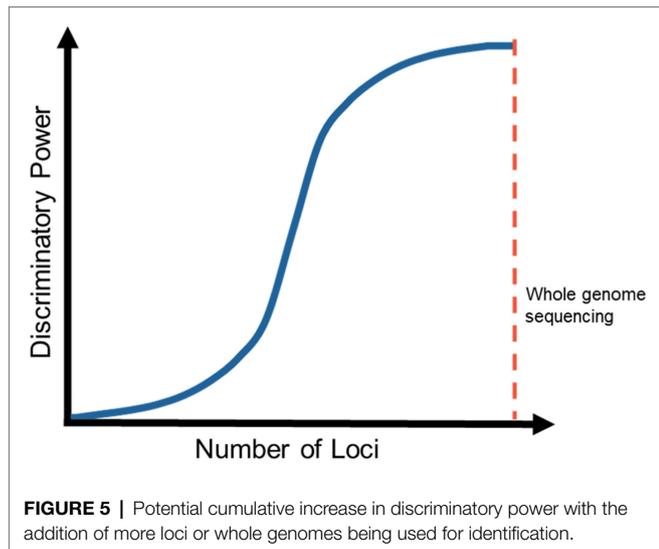
databases. It has been demonstrated that inaccurate reference sequences have caused miss-mapping of the generated sequence reads leading to clear misidentification that potentially causes confusion surrounding diagnosis in clinical practice (Bialasiewicz et al., 2019; Irinyi et al., 2020). In addition, it is causing an unproductive strain on bioinformatic tools used and lengthening the time for identification. Diagnostic whole-genome database entries must have accurate sequences, metadata, and phylogenetic coverage to prevent false positive and negative identifications.

It is predicted that with the implementation of a reliable and accurate whole-genome reference database, the accuracy of fungal identification will dramatically improve (Figure 5). The principle is demonstrated through the developing taxonomy of the *Fusarium solani* species complex. A previous study examined four loci of these species and proposed the change in nomenclature of the *F. solani* species complex to be placed in the *Neocosmospora* genus (Sandoval-Denis et al., 2019). However, a recent study has analyzed 19 loci and has maintained that the members of the *F. solani* species complex should remain in the genus *Fusarium* (Geiser et al., 2020; O'Donnell et al., 2020). As such, the increased loci analyzed generated more data and provided more insight into the taxonomy of the *F. solani* species complex and if whole-genome data were available, the most accurate taxonomic information would be revealed, which if applied to all (pathogenic) fungal species

⁵<https://fungidb.org/>

⁶<https://www.ncbi.nlm.nih.gov/genbank/>





would greatly increase the discriminatory power of long-read sequencing in clinical diagnosis.

This highlights that global efforts are urgently needed to improve the relevant reference genome databases (FungiDB and GenBank) as well as the antifungal resistance database MARDy⁷ (Nash et al., 2018), to be able to open up the full potential of the metagenomic data generated by long-read sequencers.

Bioinformatics

Currently, there is a wide variety of bioinformatics tools for the analysis of long-read sequences (Makalowski and Shabardina, 2020). Some such bioinformatics tools used for species identification in metagenomic studies are “What’s In My Pot” (Juil et al., 2015), SUPRI real-time (Gu et al., 2021), MetaMaps (Dilthey et al., 2019), BugSeq (Fan et al., 2021), and NGSspeciesID (Vasiljevic et al., 2021).

For use in a clinical setting, a simple bioinformatics workflow would be necessary for practical implementation into routine diagnosis as it would be best to avoid requiring highly specialized personnel for analysis. A study comparing the ability of different bioinformatics tools to identify a wide array of pathogens from long-read sequencing dataset would be necessary to determine the most suitable bioinformatics tools.

Data interpretation is a major issue in metagenomics and metabarcoding analysis. Both methods generate large amounts of data and for the purpose of clinical diagnosis, the interpretation can be complex. When samples are taken from sterile sites, the expected sequencing results would consist of predominantly *Homo sapiens* reads and reads assigned the causative pathogen. However, if any contamination was to occur and reads were assigned to multiple microbes, this would complicate identification of the causative pathogen. This issue is magnified when patient samples are taken from non-sterile sites. Sequencing

results will have reads assigned to multiple microbes and although it would be intuitive to assume the microbe with the highest abundance would be the causative pathogen, this is not always the case as has been observed in Irinyi et al. (2020). To overcome this, the application of machine-learning algorithms helping to distinguish true pathogens from common colonizers or environmental contaminants may be one of the ways forward. The large proportion of *H. sapiens* reads also complicates data interpretation as percentages of reads assigned to microorganisms is comparatively small. However, if bioinformatics tools were designed to remove host reads, microorganism abundance can be more easily compared and may further illuminate the causative pathogen.

THE FUTURE OF LONG-READ SEQUENCING AND FUNGAL DIAGNOSIS

The introduction of long-read sequencers from PacBio and ONT has allowed sequencing of genetic material in a way previously not possible. The first applications of long-read sequencing to identify agents of infectious disease has demonstrated its potential in clinical diagnosis.

The current state of DNA extraction technologies, the lack of comprehensive reference databases, and simple bioinformatics pipelines requires further development before long-read sequencers can be implemented in routine diagnosis. Additionally, the identification of multiple microbes within a clinical specimen may not directly lead to a specific diagnosis, but in conjunction with the patient’s presentation and further testing should seriously be considered as a way to achieve a fast and reliable diagnosis. Further advancements in the clinical diagnostic metagenomics workflow must be made to make its implementation in routine clinical diagnostic feasible. With the rapid decline of sequencing costs and the exponential improvement in sequencing capabilities, long-read sequencing is an important new technology which may eventually reduce the impact of pathogenic fungi on human health, by drastically reducing the turnaround time for diagnosis, supporting better treatment choices, reducing morbidity and mortality, and massively reduce associated healthcare costs, through its use in clinical diagnosis.

AUTHOR CONTRIBUTIONS

WM, LI, and BS conceived the review. WM coordinated and supervised the review. MH, LI, YH, BS, and WM wrote and corrected the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by a National Health and Medical Research Council of Australia (NH&MRC) grant (no. APP1121936) to WM.

⁷<http://mardy.dide.ic.ac.uk/>

REFERENCES

- Angeletti, S. (2017). Matrix assisted laser desorption time of flight mass spectrometry (MALDI-TOF MS) in clinical microbiology. *J. Microbiol. Methods* 138, 20–29. doi: 10.1016/j.mimet.2016.09.003
- Ardui, S., Ameer, A., Vermeesch, J. R., and Hestand, M. S. (2018). Single molecule real-time (SMRT) sequencing comes of age: applications and utilities for medical diagnostics. *Nucleic Acids Res.* 46, 2159–2168. doi: 10.1093/nar/gky066
- Armanhi, J. S., de Souza, R. S., de Araujo, L. M., Okura, V. K., Mieczkowski, P., Imperial, J., et al. (2016). Multiplex amplicon sequencing for microbe identification in community-based culture collections. *Sci. Rep.* 6:29543. doi: 10.1038/srep29543
- Ashikawa, S., Tarumoto, N., Imai, K., Sakai, J., Kodana, M., Kawamura, T., et al. (2018). Rapid identification of pathogens from positive blood culture bottles with the MinION nanopore sequencer. *J. Med. Microbiol.* 67, 1589–1595. doi: 10.1099/jmm.0.000855
- Austin, B. (2017). The value of cultures to modern microbiology. *Antonie Van Leeuwenhoek* 110, 1247–1256. doi: 10.1007/s10482-017-0840-8
- Basenko, E. Y., Pulman, J. A., Shanmugasundram, A., Harb, O. S., Crouch, K., Starns, D., et al. (2018). FungiDB: an integrated bioinformatic resource for fungi and oomycetes. *J. Fungi* 4, 1–28. doi: 10.3390/jof4010039
- Bergfors, A., Leenheer, D., Bergqvist, A., Ameer, A., and Lennerstrand, J. (2016). Analysis of hepatitis C NS5A resistance associated polymorphisms using ultra deep single molecule real time (SMRT) sequencing. *Antivir. Res.* 126, 81–89. doi: 10.1016/j.antiviral.2015.12.005
- Bharti, R., and Grimm, D. G. (2021). Current challenges and best-practice protocols for microbiome analysis. *Brief. Bioinform.* 22, 178–193. doi: 10.1093/bib/bbz155
- Bialasiewicz, S., Duarte, T. P., Nguyen, S. H., Sukumaran, V., Stewart, A., Appleton, S., et al. (2019). Rapid diagnosis of *Capnocytophaga canimorsus* septic shock in an immunocompetent individual using real-time Nanopore sequencing: a case report. *BMC Infect. Dis.* 19:660. doi: 10.1186/s12879-019-4173-2
- Brandt, M. E., Padhye, A. A., Mayer, L. W., and Holloway, B. P. (1998). Utility of random amplified polymorphic DNA PCR and TaqMan automated detection in molecular identification of *Aspergillus fumigatus*. *J. Clin. Microbiol.* 36, 2057–2062. doi: 10.1128/JCM.36.7.2057-2062.1998
- Bull, R. A., Eltahla, A. A., Rodrigo, C., Koekkoek, S. M., Walker, M., Pirozian, M. R., et al. (2016). A method for near full-length amplification and sequencing for six hepatitis C virus genotypes. *BMC Genomics* 17:247. doi: 10.1186/s12864-016-2575-8
- Carvalho, A., Costa-De-Oliveira, S., Martins, M. L., Pina-Vaz, C., Rodrigues, A. G., Ludovico, P., et al. (2007). Multiplex PCR identification of eight clinically relevant *Candida* species. *Med. Mycol.* 45, 619–627. doi: 10.1080/13693780701501787
- Castro-Wallace, S. L., Chiu, C. Y., John, K. K., Stahl, S. E., Rubins, K. H., McIntyre, A. B. R., et al. (2017). Nanopore DNA sequencing and genome assembly on the international space station. *Sci. Rep.* 7:18022. doi: 10.1038/s41598-017-18364-0
- Chan, W. S., Au, C. H., Leung, S. M., Ho, D. N., Wong, E. Y. L., To, M. Y., et al. (2020). Potential utility of targeted nanopore sequencing for improving etiologic diagnosis of bacterial and fungal respiratory infection. *Diagn. Pathol.* 15:41. doi: 10.1186/s13000-020-00960-w
- Charalampous, T., Kay, G. L., Richardson, H., Aydin, A., Baldan, R., Jeanes, C., et al. (2019). Nanopore metagenomics enables rapid clinical diagnosis of bacterial lower respiratory infection. *Nat. Biotechnol.* 37, 783–792. doi: 10.1038/s41587-019-0156-5
- Chen, S. C.-A., Meyer, W., and Sorrell, T. C. (2014). *Cryptococcus gattii* infections. *Clin. Microbiol. Rev.* 27, 980–1024. doi: 10.1128/CMR.00126-13
- Cheng, J., Hu, H., Kang, Y., Chen, W., Fang, W., Wang, K., et al. (2018). Identification of pathogens in culture-negative infective endocarditis cases by metagenomic analysis. *Ann. Clin. Microbiol. Antimicrob.* 17:43. doi: 10.1186/s12941-018-0294-5
- Chin, C. S., Alexander, D. H., Marks, P., Klammer, A. A., Drake, J., Heiner, C., et al. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* 10, 563–569. doi: 10.1038/nmeth.2474
- Church, G., Deamer, D.W., Branton, D., Baldarelli, R., and Kasianowicz, J. (1998). *Characterization of Individual Polymer Molecules Based on Monomer-Interface Interactions*. US Patent No. 5795782. Available at: <https://patents.google.com/patent/US5795782A/en> (Accessed 23 February 2016).
- Clancy, C. J., Pappas, P. G., Vazquez, J., Judson, M. A., Kontoyiannis, D. P., Thompson, G. R. 3rd, et al. (2018). Detecting infections rapidly and easily for candidemia trial, part 2 (DIRECT2): a prospective, multicenter study of the T2Candida panel. *Clin. Infect. Dis.* 66, 1678–1686. doi: 10.1093/cid/cix1095
- Clark, A. E., Kaleta, E. J., Arora, A., and Wolk, D. M. (2013). Matrix-assisted laser desorption ionization-time of flight mass spectrometry: a fundamental shift in the routine practice of clinical microbiology. *Clin. Microbiol. Rev.* 26, 547–603. doi: 10.1128/CMR.00072-12
- Cline, L. C., and Zak, D. R. (2015). Initial colonization, community assembly and ecosystem function: fungal colonist traits and litter biochemistry mediate decay rate. *Mol. Ecol.* 24, 5045–5058. doi: 10.1111/mec.13361
- Colabella, C., Casagrande Pierantoni, D., Corte, L., Roscini, L., Conti, A., Bassetti, M., et al. (2021). Single strain high-depth NGS reveals high rDNA (ITS-LSU) variability in the four prevalent pathogenic species of the genus *Candida*. *Microorganisms* 9:302. doi: 10.3390/microorganisms9020302
- Cornell, B. A., Braach-Maksvytis, V. L. B., King, L. G., Osman, P. D. J., Raguse, B., Wiczorek, L., et al. (1997). A biosensor that uses ion-channel switches. *Nature* 387, 580–583. doi: 10.1038/42432
- Couto, N., Schuele, L., Raangs, E. C., Machado, M. P., Mendes, C. I., Jesus, T. F., et al. (2018). Critical steps in clinical shotgun metagenomics for the concomitant detection and typing of microbial pathogens. *Sci. Rep.* 8:13767. doi: 10.1038/s41598-018-31873-w
- Cuomo, C. A., Shea, T., Yang, B., Rao, R., and Forche, A. (2017). Whole genome sequence of the heterozygous clinical isolate *Candida krusei* 81-B-5. *G3* 7, 2883–2889. doi: 10.1534/g3.117.043547
- D'Andreano, S., Cusco, A., and Francino, O. (2021). Rapid and real-time identification of fungi up to species level with long amplicon nanopore sequencing from clinical samples. *Biol. Methods Protoc.* 6:bpaa026. doi: 10.1093/biomethods/bpaa026
- de Heer, K., Gerritsen, M. G., Visser, C. E., and Leeflang, M. M. (2019). Galactomannan detection in broncho-alveolar lavage fluid for invasive aspergillosis in immunocompromised patients. *Cochrane Database Syst. Rev.* 5:CD012399. doi: 10.1002/14651858.CD012399.pub2
- De Hoog, G., Guarro, J., Gené, J., Ahmed, S., Al-Hatmi, A., Figueras, M., et al. (2019). *Atlas of Clinical Fungi* 2019. Reus: Utrecht, The Netherlands.
- De Maio, N., Shaw, L. P., Hubbard, A., George, S., Sanderson, N. D., Swann, J., et al. (2019). Comparison of long-read sequencing technologies in the hybrid assembly of complex bacterial genomes. *Microb. Genom.* 5, 1–12. doi: 10.1099/mgen.0.000294
- Deamer, D., Akeson, M., and Branton, D. (2016). Three decades of nanopore sequencing. *Nat. Biotechnol.* 34, 518–524. doi: 10.1038/nbt.3423
- Dendis, M., Horvath, R., Michalek, J., Ruzicka, F., Grijalva, M., Bartos, M., et al. (2003). PCR-RFLP detection and species identification of fungal pathogens in patients with febrile neutropenia. *Clin. Microbiol. Infect.* 9, 1191–1202. doi: 10.1111/j.1469-0691.2003.00719.x
- Deng, X., Achari, A., Federman, S., Yu, G., Somasekar, S., Bártolo, I., et al. (2020). Metagenomic sequencing with spiked primer enrichment for viral diagnostics and genomic surveillance. *Nat. Microbiol.* 5, 443–454. doi: 10.1038/s41564-019-0637-9
- Dilernia, D. A., Chien, J.-T., Monaco, D. C., Brown, M. P., Ende, Z., Deymier, M. J., et al. (2015). Multiplexed highly-accurate DNA sequencing of closely-related HIV-1 variants using continuous long reads from single molecule, real-time sequencing. *Nucleic Acids Res.* 43:e129. doi: 10.1093/nar/gkv630
- Dilthey, A. T., Jain, C., Koren, S., and Phillippy, A. M. (2019). Strain-level metagenomic assignment and compositional estimation for long reads with MetaMaps. *Nat. Commun.* 10:3066. doi: 10.1038/s41467-019-10934-2
- Dudas, G., Carvalho, L. M., Bedford, T., Tatem, A. J., Baele, G., Faria, N. R., et al. (2017). Virus genomes reveal factors that spread and sustained the Ebola epidemic. *Nature* 544, 309–315. doi: 10.1038/nature22040
- Earl, J. P., Adappa, N. D., Krol, J., Bhat, A. S., Balashov, S., Ehrlich, R. L., et al. (2018). Species-level bacterial community profiling of the healthy sinonasal microbiome using pacific biosciences sequencing of full-length 16S rRNA genes. *Microbiome* 6:190. doi: 10.1186/s40168-018-0569-2

- Edouard, S., Million, M., Bachar, D., Dubourg, G., Michelle, C., Ninove, L., et al. (2018). The nasopharyngeal microbiota in patients with viral respiratory tract infections is enriched in bacterial pathogens. *Eur. J. Clin. Microbiol. Infect. Dis.* 37, 1725–1733. doi: 10.1007/s10096-018-3305-8
- Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., et al. (2009). Real-time DNA sequencing from single polymerase molecules. *Science* 323, 133–138. doi: 10.1126/science.1162986
- Eisenstz, M. (2012). Oxford nanopore announcement sets sequencing sector abuzz. *Nat. Biotechnol.* 30, 295–296. doi: 10.1038/nbt0412-295
- El-Rami, F. E., Zielke, R. A., Wi, T., Sikora, A. E., and Unemo, M. (2019). Quantitative proteomics of the 2016 WHO *Neisseria gonorrhoeae* reference strains surveys vaccine candidates and antimicrobial resistance determinants. *Mol. Cell. Proteomics* 18, 127–150. doi: 10.1074/mcp.RA118.001125
- Fan, J., Huang, S., and Chorlton, S. D. (2021). BugSeq: a highly accurate cloud platform for long-read metagenomic analyses. *BMC Bioinform.* 22:160. doi: 10.1186/s12859-021-04089-5
- Faria, N. R., Kraemer, M. U. G., Hill, S. C., Goes de Jesus, J., Aguiar, R. S., Iani, F. C. M., et al. (2018). Genomic and epidemiological monitoring of yellow fever virus transmission potential. *Science* 361, 894–899. doi: 10.1126/science.aat7115
- Faria, N. R., Quick, J., Claro, I. M., Theze, J., de Jesus, J. G., Giovanetti, M., et al. (2017). Establishment and cryptic transmission of Zika virus in Brazil and the Americas. *Nature* 546, 406–410. doi: 10.1038/nature22401
- Feehan, J. M., Scheibel, K. E., Bourras, S., Underwood, W., Keller, B., and Somerville, S. C. (2017). Purification of high molecular weight genomic DNA from powdery mildew for long-read sequencing. *J. Vis. Exp.* e55463, 1–8. doi: 10.3791/55463
- Feng, Z., Clemente, J. C., Wong, B., and Schadt, E. E. (2021). Detecting and phasing minor single-nucleotide variants from long-read sequencing data. *Nat. Commun.* 12:3032. doi: 10.1038/s41467-021-23289-4
- Fenselau, C., and Demirev, P. A. (2001). Characterization of intact microorganisms by MALDI mass spectrometry. *Mass Spectrom. Rev.* 20, 157–171. doi: 10.1002/mas.10004
- Forbes, J. D., Knox, N. C., Ronholm, J., Pagotto, F., and Reimer, A. (2017). Metagenomics: the next culture-independent game changer. *Front. Microbiol.* 8:1069. doi: 10.3389/fmicb.2017.01069
- Ganley, A. R., and Kobayashi, T. (2007). Highly efficient concerted evolution in the ribosomal DNA repeats: total rDNA repeat variation revealed by whole-genome shotgun sequence data. *Genome Res.* 17, 184–191. doi: 10.1101/gr.5457707
- Gargis, A. S., Cherney, B., Conley, A. B., McLaughlin, H. P., and Sue, D. (2019). Rapid detection of genetic engineering, structural variation, and antimicrobial resistance markers in bacterial bioterror pathogens by nanopore sequencing. *Sci. Rep.* 9:13501. doi: 10.1038/s41598-019-49700-1
- Garrido-Cardenas, J. A., and Manzano-Aguilario, F. (2017). The metagenomics worldwide research. *Curr. Genet.* 63, 819–829. doi: 10.1007/s00294-017-0693-8
- Geiser, D. M., Al-Hatmi, A., Aoki, T., Arie, T., Balmas, V., Barnes, I., et al. (2020). Phylogenomic analysis of a 55.1 kb 19-gene dataset resolves a monophyletic *Fusarium* that includes the *Fusarium solani* species complex. *Phytopathology* 111, 1064–1079. doi: 10.1094/PHYTO-08-20-0330-LE
- Goldstein, E. J., Tyrrell, K. L., and Citron, D. M. (2015). *Lactobacillus* species: taxonomic complexity and controversial susceptibilities. *Clin. Infect. Dis.* 60(Suppl. 2), S98–S107. doi: 10.1093/cid/civ072
- Gonzalez-Escalona, N., Allard, M. A., Brown, E. W., Sharma, S., and Hoffmann, M. (2019). Nanopore sequencing for fast determination of plasmids, phages, virulence markers, and antimicrobial resistance genes in Shiga toxin-producing *Escherichia coli*. *PLoS One* 14:e0220494. doi: 10.1371/journal.pone.0220494
- Gorrie, C. L., Mirceta, M., Wick, R. R., Judd, L. M., Wyres, K. L., Thomson, N. R., et al. (2018). Antimicrobial-resistant *Klebsiella pneumoniae* carriage and infection in specialized geriatric care wards linked to acquisition in the referring hospital. *Clin. Infect. Dis.* 67, 161–170. doi: 10.1093/cid/ciy027
- Greig, D. R., Jenkins, C., Gharbia, S., and Dallman, T. J. (2019). Comparison of single-nucleotide variants identified by illumina and oxford nanopore technologies in the context of a potential outbreak of Shiga toxin-producing *Escherichia coli*. *GigaScience* 8, 1–12. doi: 10.1093/gigascience/giz104
- Greninger, A. L., Naccache, S. N., Federman, S., Yu, G., Mbala, P., Bres, V., et al. (2015). Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis. *Genome Med.* 7:99. doi: 10.1186/s13073-015-0220-9
- Gu, W., Deng, X., Lee, M., Sucu, Y. D., Arevalo, S., Stryke, D., et al. (2021). Rapid pathogen detection by metagenomic next-generation sequencing of infected body fluids. *Nat. Med.* 27, 115–124. doi: 10.1038/s41591-020-1105-z
- Guarner, J., and Brandt, M. E. (2011). Histopathologic diagnosis of fungal infections in the 21st century. *Clin. Microbiol. Rev.* 24, 247–280. doi: 10.1128/CMR.00053-10
- Gustafson, E., Bakotic, W., Bennett, L., Page, L., and McCarthy, L. (2019). DNA-based detection for onychomycosis correlates better to histopathology than does fungal culture. *Dermatol. Online J.* 25, 1–9.
- Hahn, A., Sanyal, A., Perez, G. F., Colberg-Poley, A. M., Campos, J., Rose, M. C., et al. (2016). Different next generation sequencing platforms produce different microbial profiles and diversity in cystic fibrosis sputum. *J. Microbiol. Methods* 130, 95–99. doi: 10.1016/j.mimet.2016.09.002
- Hamad, I., Abou Abdallah, R., Ravaux, I., Mokhtari, S., Tissot-Dupont, H., Michelle, C., et al. (2018). Metabarcoding analysis of eukaryotic microbiota in the gut of HIV-infected patients. *PLoS One* 13:e0191913. doi: 10.1371/journal.pone.0191913
- Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J., and Goodman, R. M. (1998). Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem. Biol.* 5, R245–R249. doi: 10.1016/S1074-5521(98)90108-9
- Heather, J. M., and Chain, B. (2016). The sequence of sequencers: the history of sequencing DNA. *Genomics* 107, 1–8. doi: 10.1016/j.ygeno.2015.11.003
- Hebert, P. D., Cywinska, A., Ball, S. L., and Dewaard, J. R. (2003). Biological identifications through DNA barcodes. *Proc. Biol. Sci.* 270, 313–321. doi: 10.1098/rspb.2002.2218
- Heeger, F., Bourne, E. C., Baschien, C., Yurkov, A., Bunk, B., Sproer, C., et al. (2018). Long-read DNA metabarcoding of ribosomal RNA in the analysis of fungi from aquatic environments. *Mol. Ecol. Resour.* 18, 1500–1514. doi: 10.1111/1755-0998.12937
- Helmersen, K., and Aamot, H. V. (2020). DNA extraction of microbial DNA directly from infected tissue: an optimized protocol for use in nanopore sequencing. *Sci. Rep.* 10:2985. doi: 10.1038/s41598-020-59957-6
- Heravi, F. S., Zakrzewski, M., Vickery, K., and Hu, H. (2020). Host DNA depletion efficiency of microbiome DNA enrichment methods in infected tissue samples. *J. Microbiol. Methods* 170:105856. doi: 10.1016/j.mimet.2020.105856
- Ho, C. K., Raghwan, J., Koekkoek, S., Liang, R. H., Van der Meer, J. T., Van Der Valk, M., et al. (2017). Characterization of hepatitis C virus (HCV) envelope diversification from acute to chronic infection within a sexually transmitted HCV cluster by using single-molecule, real-time sequencing. *J. Virol.* 91, 1–14. doi: 10.1128/JVI.02262-16
- Hoenen, T., Groseth, A., Rosenke, K., Fischer, R. J., Hoenen, A., Judson, S. D., et al. (2016). Nanopore sequencing as a rapidly deployable Ebola outbreak tool. *Emerg. Infect. Dis.* 22:331. doi: 10.3201/eid2202.151796
- Hu, K., Huang, N., Zou, Y., Liao, X., and Wang, J. (2021). MultiNanopolish: refined grouping method for reducing redundant calculations in nanopolish. *Bioinformatics* 37, 2757–2760. doi: 10.1093/bioinformatics/btab078
- Huang, D. W., Raley, C., Jiang, M. K., Zheng, X., Liang, D., Rehman, M. T., et al. (2016). Towards better precision medicine: PacBio single-molecule long reads resolve the interpretation of HIV drug resistant mutation profiles at explicit quasispaces (haplotype) level. *J. Data Min. Genom. Proteom.* 7, 1–16. doi: 10.4172/2153-0602.1000182
- Illumina (2021). Focused Power on the MiSeq System [Online]. Available: <https://sapac.illumina.com/systems/sequencing-platforms/miseq.html> (Accessed April 11, 2021).
- Im, K., Mareninov, S., Diaz, M. F. P., and Yong, W. H. (2019). An introduction to performing immunofluorescence staining. *Methods Mol. Biol.* 1897, 299–311. doi: 10.1007/978-1-4939-8935-5_26
- Imai, K., Nemoto, R., Kodana, M., Tarumoto, N., Sakai, J., Kawamura, T., et al. (2020). Rapid and accurate species identification of mitis group streptococci using the MinION nanopore sequencer. *Front. Cell. Infect. Microbiol.* 10:11. doi: 10.3389/fcimb.2020.00011
- Inglis, P. W., Pappas, M. C. R., Resende, L. V., and Grattapaglia, D. (2018). Fast and inexpensive protocols for consistent extraction of high quality DNA and RNA from challenging plant and fungal samples for high-throughput SNP genotyping and sequencing applications. *PLoS One* 13:e0206085. doi: 10.1371/journal.pone.0206085

- Irinyi, L., Hu, Y., Hoang, M. T. V., Pasic, L., Halliday, C., Jayawardena, M., et al. (2020). Long-read sequencing based clinical metagenomics for the detection and confirmation of *Pneumocystis jirovecii* directly from clinical specimens: a paradigm shift in mycological diagnostics. *Med. Mycol.* 58, 650–660. doi: 10.1093/mmy/myz109
- Irinyi, L., Lackner, M., de Hoog, G. S., and Meyer, W. (2016). DNA barcoding of fungi causing infections in humans and animals. *Fungal Biol.* 120, 125–136. doi: 10.1016/j.funbio.2015.04.007
- Irinyi, L., Serena, C., Garcia-Hermoso, D., Arabatzis, M., Desnos-Ollivier, M., Vu, D., et al. (2015). International Society of Human and Animal Mycology (ISHAM)-ITS reference DNA barcoding database—the quality controlled standard tool for routine identification of human and animal pathogenic fungi. *Med. Mycol.* 53, 313–337. doi: 10.1093/mmy/myv008
- Jain, M., Olsen, H. E., Paten, B., and Akeson, M. (2016). The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol.* 17:239. doi: 10.1186/s13059-016-1103-0
- James, T. Y., Marino, J. A., Perfecto, I., and Vandermeer, J. (2016). Identification of putative coffee rust mycoparasites via single-molecule DNA sequencing of infected pustules. *J. Appl. Environ. Microbiol.* 82, 631–639. doi: 10.1128/AEM.02639-15
- Janda, J. M., and Abbott, S. L. (2007). 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. *J. Clin. Microbiol.* 45, 2761–2764. doi: 10.1128/JCM.01228-07
- Jumpponen, A. (2003). Soil fungal community assembly in a primary successional glacier forefront ecosystem as inferred from rDNA sequence analyses. *New Phytol.* 158, 569–578. doi: 10.1046/j.1469-8137.2003.00767.x
- Juul, S., Izquierdo, E., Hurst, A., Dai, X., Wright, A., Kulesha, E., et al. (2015). What's in my pot? Real-time species identification on the MinION™. bioRxiv. doi: 10.1101/030742
- Kafetzopoulou, L. E., Efthymiadis, K., Lewandowski, K., Crook, A., Carter, D., Osborne, J., et al. (2018). Assessment of metagenomic Nanopore and Illumina sequencing for recovering whole genome sequences of chikungunya and dengue viruses directly from clinical samples. *Euro Surveill.* 23, 1–13. doi: 10.2807/1560-7917.ES.2018.23.50.1800228
- Kafetzopoulou, L., Pullan, S., Lemey, P., Suchard, M., Ehichioya, D., Pahlmann, M., et al. (2019). Metagenomic sequencing at the epicenter of the Nigeria 2018 Lassa fever outbreak. *Science* 363, 74–77. doi: 10.1126/science.aau9343
- Kai, S., Matsuo, Y., Nakagawa, S., Kryukov, K., Matsukawa, S., Tanaka, H., et al. (2019). Rapid bacterial identification by direct PCR amplification of 16S rRNA genes using the MinION™ nanopore sequencer. *FEBS Open Bio* 9, 548–557. doi: 10.1002/2211-5463.12590
- Kanamori, H., Parobek, C. M., Weber, D. J., van Duin, D., Rutala, W. A., Cairns, B. A., et al. (2016). Next-generation sequencing and comparative analysis of sequential outbreaks caused by multidrug-resistant *Acinetobacter baumannii* at a large academic burn center. *Antimicrob. Agents Chemother.* 60, 1249–1257. doi: 10.1128/AAC.02014-15
- Karageorgopoulos, D. E., Vouloumanou, E. K., Ntziora, F., Michalopoulos, A., Rafailidis, P. I., and Falagas, M. E. (2011). Beta-D-glucan assay for the diagnosis of invasive fungal infections: a meta-analysis. *Clin. Infect. Dis.* 52, 750–770. doi: 10.1093/cid/ciq206
- Kasianowicz, J. J., Brandin, E., Branton, D., and Deamer, D. W. (1996). Characterization of individual polynucleotide molecules using a membrane channel. *Proc. Natl. Acad. Sci. U. S. A.* 93, 13770–13773. doi: 10.1073/pnas.93.24.13770
- Kidd, S. E., Chen, S. C., Meyer, W., and Halliday, C. L. (2019). A new age in molecular diagnostics for invasive fungal disease: are we ready? *Front. Microbiol.* 10:2903. doi: 10.3389/fmicb.2019.02903
- Klingspor, L., and Jalal, S. (2006). Molecular detection and identification of *Candida* and *Aspergillus* spp. from clinical samples using real-time PCR. *Clin. Microbiol. Infect.* 12, 745–753. doi: 10.1111/j.1469-0691.2006.01498.x
- Ko, S. H., Bayat Mokhtari, E., Mudvari, P., Stein, S., Stringham, C. D., Wagner, D., et al. (2021). High-throughput, single-copy sequencing reveals SARS-CoV-2 spike variants coincident with mounting humoral immunity during acute COVID-19. *PLoS Pathog.* 17:e1009431. doi: 10.1371/journal.ppat.1009431
- Koljalg, U., Larsson, K. H., Abarenkov, K., Nilsson, R. H., Alexander, I. J., Eberhardt, U., et al. (2005). UNITE: a database providing web-based methods for the molecular identification of ectomycorrhizal fungi. *New Phytol.* 166, 1063–1068. doi: 10.1111/j.1469-8137.2005.01376.x
- Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., and Phillippy, A. M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* 27, 722–736. doi: 10.1101/gr.215087.116
- Kovaka, S., Fan, Y., Ni, B., Timp, W., and Schatz, M. C. (2020). Targeted nanopore sequencing by real-time mapping of raw electrical signal with UNCALLED. *Nat. Biotechnol.* 29, 431–441. doi: 10.1038/s41587-020-0731-9
- Kulik, T., Fordoński, G., Pszczółkowska, A., Płodzień, K., and Lapiński, M. (2004). Development of PCR assay based on ITS2 rDNA polymorphism for the detection and differentiation of *Fusarium sporotrichioides*. *FEMS Microbiol. Lett.* 239, 181–186. doi: 10.1016/j.femsle.2004.08.037
- Kurtzman, C., Fell, J. W., and Boekhout, T. (2011). *The Yeasts: A Taxonomic Study*. Amsterdam, Netherlands: Elsevier.
- Kyaschenko, J., Clemmensen, K. E., Hagenbo, A., Karlton, E., and Lindahl, B. D. (2017). Shift in fungal communities and associated enzyme activities along an age gradient of managed *Pinus sylvestris* stands. *ISME J.* 11, 863–874. doi: 10.1038/ismej.2016.184
- Lackner, M., de Hoog, G. S., Yang, L., Moreno, L. F., Ahmed, S. A., Andreas, F., et al. (2014). Proposed nomenclature for *Pseudallescheria*, *Scedosporium* and related genera. *Fungal Divers.* 67, 1–10. doi: 10.1007/s13225-014-0295-4
- Laudadio, I., Fulci, V., Palone, F., Stronati, L., Cucchiara, S., and Carissimi, C. (2018). Quantitative assessment of shotgun metagenomics and 16S rDNA amplicon sequencing in the study of human gut microbiome. *OMICS* 22, 248–254. doi: 10.1089/omi.2018.0013
- Leggett, R. M., and Clark, M. D. (2017). A world of opportunities with nanopore sequencing. *J. Exp. Bot.* 68, 5419–5429. doi: 10.1093/jxb/erx289
- Lewandowski, K., Xu, Y., Pullan, S. T., Lumley, S. F., Foster, D., Sanderson, N., et al. (2019). Metagenomic nanopore sequencing of influenza virus direct from clinical respiratory samples. *J. Clin. Microbiol.* 58, 1–15. doi: 10.1128/JCM.00963-19
- Li, Y., He, X. Z., Li, M. H., Li, B., Yang, M. J., Xie, Y., et al. (2020). Comparison of third-generation sequencing approaches to identify viral pathogens under public health emergency conditions. *Virus Genes* 56, 288–297. doi: 10.1007/s11262-020-01746-4
- Lieckfeldt, E., Meyer, W., and Borner, T. (1993). Rapid identification and differentiation of yeasts by DNA and PCR fingerprinting. *J. Basic Microbiol.* 33, 413–425. doi: 10.1002/jobm.3620330609
- Lindsley, M. D., Hurst, S. F., Iqbal, N. J., and Morrison, C. J. (2001). Rapid identification of dimorphic and yeast-like fungal pathogens using specific DNA probes. *J. Clin. Microbiol.* 39, 3505–3511. doi: 10.1128/JCM.39.10.3505-3511.2001
- Logsdon, G. A., Vollger, M. R., and Eichler, E. E. (2020). Long-read human genome sequencing and its applications. *Nat. Rev. Genet.* 21, 597–614. doi: 10.1038/s41576-020-0236-x
- Loit, K., Adamson, K., Bahram, M., Puusepp, R., Anslan, S., Kiiker, R., et al. (2019). Relative performance of MinION (Oxford Nanopore technologies) versus sequel (Pacific biosciences) third-generation sequencing instruments in identification of agricultural and Forest fungal pathogens. *J. Appl. Environ. Microbiol.* 85, 1–20. doi: 10.1128/AEM.01368-19
- Loman, N. J., Quick, J., and Simpson, J. T. (2015). A complete bacterial genome assembled de novo using only nanopore sequencing data. *Nat. Methods* 12, 733–735. doi: 10.1038/nmeth.3444
- Lu, H., Giordano, F., and Ning, Z. (2016). Oxford nanopore MinION sequencing and genome assembly. *Genom. Proteom. Bioinform.* 14, 265–279. doi: 10.1016/j.gpb.2016.05.004
- Lui, W. Y., Yuen, C. K., Li, C., Wong, W. M., Lui, P. Y., Lin, C. H., et al. (2019). SMRT sequencing revealed the diversity and characteristics of defective interfering RNAs in influenza A (H7N9) virus infection. *Emerg. Microbes Infect.* 8, 662–674. doi: 10.1080/22221751.2019.1611346
- Luo, R., Zimin, A., Workman, R., Fan, Y., Pertea, G., Grossman, N., et al. (2017). First draft genome sequence of the pathogenic fungus *lomentospora prolificans* (formerly *scedosporium prolificans*). *G3* 7, 3831–3836. doi: 10.1534/g3.117.300107
- Mafune, K. K., Godfrey, B. J., Vogt, D. J., and Vogt, K. A. (2020). A rapid approach to profiling diverse fungal communities using the MinION nanopore sequencer. *BioTechniques* 68, 72–78. doi: 10.2144/btn-2019-0072
- Makalowski, W., and Shabardina, V. (2020). Bioinformatics of nanopore sequencing. *J. Hum. Genet.* 65, 61–67. doi: 10.1038/s10038-019-0659-4
- Mannarelli, B. M., and Kurtzman, C. P. (1998). Rapid identification of *Candida albicans* and other human pathogenic yeasts by using short oligonucleotides

- in a PCR. *J. Clin. Microbiol.* 36, 1634–1641. doi: 10.1128/JCM.36.6.1634-1641.1998
- Mantere, T., Kersten, S., and Hoischen, A. (2019). Long-read sequencing emerging in medical genetics. *Front. Genet.* 10:426. doi: 10.3389/fgene.2019.00426
- Marotz, C. A., Sanders, J. G., Zuniga, C., Zaramela, L. S., Knight, R., and Zengler, K. (2018). Improving saliva shotgun metagenomics by chemical host DNA depletion. *Microbiome* 6:42. doi: 10.1186/s40168-018-0426-3
- Martin, C., Roberts, D., van der Weide, M., Rossau, R., Jannes, G., Smith, T., et al. (2000). Development of a PCR-based line probe assay for identification of fungal pathogens. *J. Clin. Microbiol.* 38, 3735–3742. doi: 10.1128/JCM.38.10.3735-3742.2000
- Meredith, L. W., Hamilton, W. L., Warne, B., Houldcroft, C. J., Hosmillo, M., Jahun, A. S., et al. (2020). Rapid implementation of SARS-CoV-2 sequencing to investigate cases of health-care associated COVID-19: a prospective genomic surveillance study. *Lancet Infect. Dis.* 20, 1263–1272. doi: 10.1016/S1473-3099(20)30562-4
- Meyer, W., Irinyi, L., Hoang, M. T. V., Robert, V., Garcia-Hermoso, D., Desnos-Ollivier, M., et al. (2019). Database establishment for the secondary fungal DNA barcode *translational elongation factor 1 α* (*TEF1 α*). *Genome* 62, 160–169. doi: 10.1139/gen-2018-0083
- Miller, S., and Chiu, C. (2018). “Metagenomic next-generation sequencing for pathogen detection and identification,” in *Advanced Techniques in Diagnostic Microbiology: Volume 2: Applications*. eds. Y.-W. Tang and C. W. Stratton (Cham: Springer International Publishing), 617–632.
- Mobasherzadeh, S., Shojaei, H., Havaei, S. A., Mostafavizadeh, K., Davoodabadi, F., Khorvash, F., et al. (2016). Application of the random amplified polymorphic DNA (RAPD) fingerprinting to analyze genetic variation in community associated-methicillin resistant *staphylococcus Aureus* (CA-MRSA) isolates in Iran. *Global J. Health Sci.* 8:53822. doi: 10.5539/gjhs.v8n8p185
- Moon, J., Kim, N., Kim, T. J., Jun, J. S., Lee, H. S., Shin, H. R., et al. (2019). Rapid diagnosis of bacterial meningitis by nanopore 16S amplicon sequencing: a pilot study. *Int. J. Med. Microbiol.* 309:151338. doi: 10.1016/j.ijmm.2019.151338
- Morand, S. C., Bertignac, M., Iltis, A., Kolder, I., Pirovano, W., Jourdain, R., et al. (2019). Complete genome sequence of *Malassezia restricta* CBS 7877, an opportunist pathogen involved in dandruff and seborrheic dermatitis. *Microbiol. Resour. Announc.* 8, 1–3. doi: 10.1128/MRA.01543-18
- Morrison, G. A., Fu, J., Lee, G. C., Wiederhold, N. P., Canete-Gibas, C. F., Bunnik, E. M., et al. (2020). Nanopore sequencing of the fungal intergenic spacer sequence as a potential rapid diagnostic assay. *J. Clin. Microbiol.* 58, 1–16. doi: 10.1128/JCM.01972-20
- Motooka, D., Fujimoto, K., Tanaka, R., Yaguchi, T., Gotoh, K., Maeda, Y., et al. (2017). Fungal ITS1 deep-sequencing strategies to reconstruct the composition of a 26-species community and evaluation of the gut mycobiota of healthy Japanese individuals. *Front. Microbiol.* 8:238. doi: 10.3389/fmicb.2017.00238
- Mylonakis, E., Clancy, C. J., Ostrosky-Zeichner, L., Garey, K. W., Alangaden, G. J., Vazquez, J. A., et al. (2015). T2 magnetic resonance assay for the rapid diagnosis of candidemia in whole blood: a clinical trial. *Clin. Infect. Dis.* 60, 892–899. doi: 10.1093/cid/ciu959
- Nakagawa, S., Inoue, S., Kryukov, K., Yamagishi, J., Ohno, A., Hayashida, K., et al. (2019). Rapid sequencing-based diagnosis of infectious bacterial species from meningitis patients in Zambia. *Clin. Transl. Immunol.* 8:e01087. doi: 10.1002/cti2.1087
- Nakano, K., Shiroma, A., Shimoji, M., Tamotsu, H., Ashimine, N., Ohki, S., et al. (2017). Advantages of genome sequencing by long-read sequencer using SMRT technology in medical area. *Hum. Cell* 30, 149–161. doi: 10.1007/s13577-017-0168-8
- Nash, A., Sewell, T., Farrer, R. A., Abdolrasouli, A., Shelton, J. M. G., Fisher, M. C., et al. (2018). MARDy: mycology antifungal resistance database. *Bioinformatics* 34, 3233–3234. doi: 10.1093/bioinformatics/bty321
- Neely, L. A., Audeh, M., Phung, N. A., Min, M., Suchocki, A., Plourde, D., et al. (2013). T2 magnetic resonance enables nanoparticle-mediated rapid detection of candidemia in whole blood. *Sci. Transl. Med.* 5:182ra154. doi: 10.1126/scitranslmed.3005377
- Nilsson, R. H., Ryberg, M., Abarenkov, K., Sjökvist, E., and Kristiansson, E. (2009). The ITS region as a target for characterization of fungal communities using emerging sequencing technologies. *FEMS Microbiol. Lett.* 296, 97–101. doi: 10.1111/j.1574-6968.2009.01618.x
- Ocwieja, K. E., Sherrill-Mix, S., Mukherjee, R., Custers-Allen, R., David, P., Brown, M., et al. (2012). Dynamic regulation of HIV-1 mRNA populations analyzed by single-molecule enrichment and long-read sequencing. *Nucleic Acids Res.* 40, 10345–10355. doi: 10.1093/nar/gks753
- O'Donnell, K., Al-Hatmi, A. M. S., Aoki, T., Brankovics, B., Cano-Lira, J. F., Coleman, J. J., et al. (2020). No to neocosmospora: phylogenomic and practical reasons for continued inclusion of the *Fusarium solani* species complex in the genus *Fusarium*. *mSphere* 5, 1–7. doi: 10.1128/mSphere.00810-20
- O'Donnell, K., and Cigelnik, E. (1997). Two divergent intragenomic rDNA ITS2 types within a monophyletic lineage of the fungus *Fusarium* are nonorthologous. *Mol. Phylogenet. Evol.* 7, 103–116. doi: 10.1006/mpev.1996.0376
- Oxford Nanopore Technologies (2019). Oxford Nanopore launches Flongle, for Rapid, Smaller DNA/RNA Sequencing Tests in Any Environment [Online]. Available: <https://nanoporetech.com/about-us/news/oxford-nanopore-launches-flongle-rapid-smaller-dnarna-sequencing-tests-any> (Accessed October 28, 2021).
- Oxford Nanopore Technologies (2021). Products [Online]. Available: <https://nanoporetech.com/products> (Accessed October 28, 2021).
- Pacific Biosciences (2020). Pacific Biosciences Launches the Sequel IIE System to Accelerate Adoption of Highly Accurate HiFi Sequencing [Online]. Available: https://www.pacb.com/press_releases/pacific-biosciences-launches-the-sequel-iie-system-to-accelerate-adoption-of-highly-accurate-hifi-sequencing/ (Accessed October 28, 2021).
- Pacific Biosciences (2021a). PacBio Sequel Systems [Online]. Available: <https://www.pacb.com/products-and-services/sequel-system/> (Accessed October 28, 2021).
- Pacific Biosciences (2021b). What Can You Do with One SMRT Cell? [Online]. Available: <https://www.pacb.com/wp-content/uploads/Application-Brochure-What-Can-You-Do-with-One-SMRT-Cell.pdf> (Accessed October 28, 2021).
- Panthee, S., Hamamoto, H., Ishijima, S. A., Paudel, A., and Sekimizu, K. (2018). Utilization of hybrid assembly approach to determine the genome of an opportunistic pathogenic fungus, *Candida albicans* TIMM 1768. *Genome Biol. Evol.* 10, 2017–2022. doi: 10.1093/gbe/evy166
- Parker, J., Helmstetter, A. J., Devey, D., Wilkinson, T., and Papadopoulos, A. S. (2017). Field-based species identification of closely-related plants using real-time nanopore sequencing. *Sci. Rep.* 7, 1–8. doi: 10.1038/s41598-017-08461-5
- Patel, T. S., Kaakeh, R., Nagel, J. L., Newton, D. W., and Stevenson, J. G. (2017). Cost analysis of implementing matrix-assisted laser desorption/ionization-time of flight mass spectrometry plus real-time antimicrobial stewardship intervention for bloodstream infections. *J. Clin. Microbiol.* 55, 60–67. doi: 10.1128/JCM.01452-16
- Payne, A., Holmes, N., Clarke, T., Munro, R., Debebe, B. J., and Loose, M. (2020). Readfish enables targeted nanopore sequencing of gigabase-sized genomes. *Nat. Biotechnol.* 35, 2193–2198. doi: 10.1038/s41587-020-00746-x
- Payne, A., Holmes, N., Rakyar, V., and Loose, M. (2019). BulkVis: a graphical viewer for Oxford nanopore bulk FAST5 files. *Bioinformatics* 35, 2193–2198. doi: 10.1093/bioinformatics/bty841
- Pchelin, I. M., Azarov, D. V., Churina, M. A., Ryabinin, I. A., Vibornova, I. V., Apalko, S. V., et al. (2020). Whole genome sequence of first *Candida auris* strain, isolated in Russia. *Med. Mycol.* 58, 414–416. doi: 10.1093/mmy/myz078
- Perfect, J. R. (2013). Fungal diagnosis: how do we do it and can we do better? *Curr. Med. Res. Opin.* 29(Suppl 4), 3–11. doi: 10.1185/03007995.2012.761134
- Pfeiffer, C. D., Fine, J. P., and Safdar, N. (2006). Diagnosis of invasive aspergillosis using a galactomannan assay: a meta-analysis. *Clin. Infect. Dis.* 42, 1417–1727. doi: 10.1086/503427
- Pollard, M. O., Gurdasani, D., Mentzer, A. J., Porter, T., and Sandhu, M. S. (2018). Long reads: their purpose and place. *Hum. Mol. Genet.* 27, R234–R241. doi: 10.1093/hmg/ddy177
- Polz, M. F., and Cavanaugh, C. M. (1998). Bias in template-to-product ratios in multitemplate PCR. *J. Appl. Environ. Microbiol.* 64, 3724–3730. doi: 10.1128/AEM.64.10.3724-3730.1998
- Pryce, T. M., Fau, P. S., Kay, I. D., and Coombs, G. W. (2003). Rapid identification of fungi by sequencing the ITS1 and ITS2 regions using an automated capillary electrophoresis system. *Med. Mycol.* 41, 369–381. doi: 10.1080/13693780310001600435
- Purahong, W., Mapook, A., Wu, Y. T., and Chen, C. T. (2019). Characterization of the *Castanopsis carlesii* deadwood mycobiome by PacBio sequencing of the full-length fungal nuclear ribosomal internal transcribed spacer (ITS). *Front. Microbiol.* 10:983. doi: 10.3389/fmicb.2019.00983

- Quick, J., Grubaugh, N. D., Pullan, S. T., Claro, I. M., Smith, A. D., Gangavarapu, K., et al. (2017). Multiplex PCR method for MinION and illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat. Protoc.* 12, 1261–1276. doi: 10.1038/nprot.2017.066
- Quick, J., Loman, N. J., Duraffour, S., Simpson, J. T., Severi, E., Cowley, L., et al. (2016). Real-time, portable genome sequencing for Ebola surveillance. *Nature* 530, 228–232. doi: 10.1038/nature16996
- Ratnasingham, S., and Hebert, P. D. (2007). Bold: the barcode of life data system. *Mol. Ecol. Notes* 7, 355–364. doi: 10.1111/j.1471-8286.2007.01678.x
- Ravi, R. K., Walton, K., and Khosroheidari, M. (2018). MiSeq: a next generation sequencing platform for genomic analysis. *Methods Mol. Biol.* 1706, 223–232. doi: 10.1007/978-1-4939-7471-9_12
- Rhoads, A., and Au, K. F. (2015). PacBio sequencing and its applications. *Genom. Proteom. Bioinform.* 13, 278–289. doi: 10.1016/j.gpb.2015.08.002
- Rhodes, J., Abdolrasouli, A., Farrer, R. A., Cuomo, C. A., Aanensen, D. M., Armstrong-James, D., et al. (2018). Genomic epidemiology of the UK outbreak of the emerging human fungal pathogen *Candida auris*. *Emerg. Infect. Dis.* 7:43. doi: 10.1038/s41426-018-0045-x
- Riesenfeld, C. S., Schloss, P. D., and Handelsman, J. (2004). Metagenomics: genomic analysis of microbial communities. *Annu. Rev. Genet.* 38, 525–552. doi: 10.1146/annurev.genet.38.072902.091216
- Romanelli, A. M., Sutton, D. A., Thompson, E. H., Rinaldi, M. G., and Wickes, B. L. (2010). Sequence-based identification of filamentous basidiomycetous fungi from clinical specimens: a cautionary note. *J. Clin. Microbiol.* 48, 741–752. doi: 10.1128/JCM.01948-09
- Sakai, J., Tarumoto, N., Kodana, M., Ashikawa, S., Imai, K., Kawamura, T., et al. (2019). An identification protocol for ESBL-producing gram-negative bacteria bloodstream infections using a MinION nanopore sequencer. *J. Med. Microbiol.* 68, 1219–1226. doi: 10.1099/jmm.0.001024
- Salmela, L., Walve, R., Rivals, E., and Ukkonen, E. (2017). Accurate self-correction of errors in long reads using de Bruijn graphs. *Bioinformatics* 33, 799–806. doi: 10.1093/bioinformatics/btw321
- Salzberg, S. L., and Yorke, J. A. (2005). Beware of mis-assembled genomes. *Bioinformatics* 21, 4320–4321. doi: 10.1093/bioinformatics/bti769
- Sandhu, G. S., Kline, B. C., Stockman, L., and Roberts, G. D. (1995). Molecular probes for diagnosis of fungal infections. *J. Clin. Microbiol.* 33, 2913–2919. doi: 10.1128/jcm.33.11.2913-2919.1995
- Sandoval-Denis, M., Lombard, L., and Crous, P. W. (2019). Back to the roots: a reappraisal of *Neocosmospora*. *Persoonia* 43, 90–185. doi: 10.3767/persoonia.2019.43.04
- Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* 74, 5463–5467. doi: 10.1073/pnas.74.12.5463
- Sayers, E. W., Cavanaugh, M., Clark, K., Ostell, J., Pruitt, K. D., and Karsch-Mizrachi, I. (2020). GenBank. *Nucleic Acids Res.* 48, D84–D86. doi: 10.1093/nar/gkz956
- Schelenz, S., Barnes, R. A., Barton, R. C., Cleverley, J. R., Lucas, S. B., Kibbler, C. C., et al. (2015). British society for medical mycology best practice recommendations for the diagnosis of serious fungal diseases. *Lancet Infect. Dis.* 15, 461–474. doi: 10.1016/S1473-3099(15)70006-X
- Schmidt, K., Mwaigwisya, S., Crossman, L. C., Doumith, M., Munroe, D., Pires, C., et al. (2017). Identification of bacterial pathogens and antimicrobial resistance directly from clinical urines by nanopore-based metagenomic sequencing. *J. Antimicrob. Chemother.* 72, 104–114. doi: 10.1093/jac/dkw397
- Schoch, C. L., Robbertse, B., Robert, V., Vu, D., Cardinali, G., Irinyi, L., et al. (2014). Finding needles in haystacks: linking scientific names, reference specimens and molecular data for fungi. *Database* 2014, 1–21. doi: 10.1093/database/bau061
- Schoch, C. L., Seifert, K. A., Huhndorf, S., Robert, V., Spouge, J. L., Levesque, C. A., et al. (2012). Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for fungi. *Proc. Natl. Acad. Sci. U. S. A.* 109, 6241–6246. doi: 10.1073/pnas.1117018109
- Schultzhause, Z., Cuomo, C. A., and Wang, Z. (2019). Genome sequence of the black yeast *Exophiala lecanii-corni*. *Microbiol. Resour. Announc.* 8, 1–2. doi: 10.1128/MRA.01709-18
- Schwessinger, B., and Rathjen, J. P. (2017). Extraction of high molecular weight DNA from fungal rust spores for long read sequencing. *Methods Mol. Biol.* 1659, 49–57. doi: 10.1007/978-1-4939-7249-4_5
- Shokralla, S., Gibson, J. F., Nikbakht, H., Janzen, D. H., Hallwachs, W., and Hajibabaei, M. (2014). Next-generation DNA barcoding: using next-generation sequencing to enhance and accelerate DNA barcode capture from single specimens. *Mol. Ecol. Resour.* 14, 892–901. doi: 10.1111/1755-0998.12236
- Simon, U. K., and Weiss, M. (2008). Intragenomic variation of fungal ribosomal genes is higher than previously thought. *Mol. Biol. Evol.* 25, 2251–2254. doi: 10.1093/molbev/msn188
- Springer, J., and Löffler, J. (2017). Genus- and species-specific PCR detection methods. *Methods Mol. Biol.* 1508, 267–279. doi: 10.1007/978-1-4939-6515-1_15
- Stielow, J. B., Levesque, C. A., Seifert, K. A., Meyer, W., Irinyi, L., Smits, D., et al. (2015). One fungus, which genes? Development and assessment of universal primers for potential secondary fungal DNA barcodes. *Persoonia* 35:242. doi: 10.3767/003158515X689135
- Takeda, H., Yamashita, T., Ueda, Y., and Sekine, A. (2019). Exploring the hepatitis C virus genome using single molecule real-time sequencing. *World J. Gastroenterol.* 25:4661. doi: 10.3748/wjg.v25.i32.4661
- Taxt, A. M., Avershina, E., Frye, S. A., Naseer, U., and Ahmad, R. (2020). Rapid identification of pathogens, antibiotic resistance genes and plasmids in blood cultures by nanopore sequencing. *Sci. Rep.* 10:7622. doi: 10.1038/s41598-020-64616-x
- Tedersoo, L., Albertsen, M., Anslan, S., and Callahan, B. (2021). Perspectives and benefits of high-throughput long-read sequencing in microbial ecology. *Appl. Environ. Microbiol.* 87:e0062621. doi: 10.1128/AEM.00626-21
- Tedersoo, L., and Anslan, S. (2019). Towards PacBio-based pan-eukaryote metabarcoding using full-length ITS sequences. *Environ. Microbiol. Rep.* 11, 659–668. doi: 10.1111/1758-2229.12776
- Tedersoo, L., Tooming-Klunderud, A., and Anslan, S. (2018). PacBio metabarcoding of fungi and other eukaryotes: errors, biases and perspectives. *New Phytol.* 217, 1370–1385. doi: 10.1111/nph.14776
- Teng, J. L., Yeung, M. L., Chan, E., Jia, L., Lin, C. H., Huang, Y., et al. (2017). PacBio but not Illumina technology can achieve fast, accurate and complete closure of the high GC, complex *Burkholderia pseudomallei* two-chromosome genome. *Front. Microbiol.* 8:1448. doi: 10.3389/fmicb.2017.01448
- Tia, T., Putaporntip, C., Kosuwini, R., Kongpolprom, N., Kawkitinarong, K., and Jongwutiwes, S. (2012). A highly sensitive novel PCR assay for detection of *Pneumocystis jirovecii* DNA in bronchoalveolar lavage specimens from immunocompromised patients. *Clin. Microbiol. Infect.* 18, 598–603. doi: 10.1111/j.1469-0691.2011.03656.x
- Treangen, T. J., and Salzberg, S. L. (2011). Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat. Rev. Genet.* 13, 36–46. doi: 10.1038/nrg3117
- Tsang, C. C., Teng, J. L., Lau, S. K. P., and Woo, P. C. Y. (2021). Rapid genomic diagnosis of fungal infections in the age of next-generation sequencing. *J. Fungi* 7, 1–15. doi: 10.3390/jof7080636
- Vale-Silva, L., Beaudoin, E., Tran, V. D. T., and Sanglard, D. (2017). Comparative genomics of two sequential *Candida glabrata* clinical isolates. *G3* 7, 2413–2426. doi: 10.1534/g3.117.042887
- van Dijk, E. L., Auger, H., Jaszczyszyn, Y., and Thermes, C. (2014). Ten years of next-generation sequencing technology. *Trends Genet.* 30, 418–426. doi: 10.1016/j.tig.2014.07.001
- van Dijk, E. L., Jaszczyszyn, Y., Naquin, D., and Thermes, C. (2018). The third revolution in sequencing technology. *Trends Genet.* 34, 666–681. doi: 10.1016/j.tig.2018.05.008
- Vasiljevic, N., Lim, M., Humble, E., Seah, A., Kratzer, A., Morf, N. V., et al. (2021). Developmental validation of Oxford nanopore technology MiniION sequence data and the NGSspeciesID bioinformatic pipeline for forensic genetic species identification. *Forensic Sci. Int. Genet.* 53:102493. doi: 10.1016/j.fsigen.2021.102493
- Votintseva, A. A., Bradley, P., Pankhurst, L., Del Ojo Elias, C., Loose, M., Nilgiriwala, K., et al. (2017). Same-day diagnostic and surveillance data for tuberculosis via whole-genome sequencing of direct respiratory samples. *J. Clin. Microbiol.* 55, 1285–1298. doi: 10.1128/JCM.02483-16
- Wagner, A., Chavez, V., Huang, R. S. P., Wahed, A., Actor, J. K., and Dasgupta, A. (2017). “Chapter 5: biochemical tests and staining techniques for microbial identification,” in *Microbiology and Molecular Diagnosis in Pathology A Comprehensive Review for Board Preparation, Certification and Clinical Practice*. ed. T. Broderick and K. Padilla (Amsterdam, Netherlands: Elsevier), 61–73.
- Wagner, J., Coupland, P., Browne, H. P., Lawley, T. D., Francis, S. C., and Parkhill, J. (2016). Evaluation of PacBio sequencing for full-length bacterial

- 16S rRNA gene classification. *BMC Microbiol.* 16:274. doi: 10.1186/s12866-016-0891-4
- Walder, F., Schlaeppli, K., Wittwer, R., Held, A. Y., Vogelgsang, S., and van der Heijden, M. G. (2017). Community profiling of *Fusarium* in combination with other plant-associated fungi in different crop species using SMRT sequencing. *Front. Plant Sci.* 8:2019. doi: 10.3389/fpls.2017.02019
- Wang, M., Fu, A., Hu, B., Shen, G., Liu, R., Zhao, W., et al. (2020). Same-day simultaneous diagnosis of bacterial and fungal infections in clinical practice by nanopore targeted sequencing. medRxiv. doi: 10.1101/2020.04.08.20057604
- Wang, W. L., Xu, S. Y., Ren, Z. G., Tao, L., Jiang, J. W., and Zheng, S. S. (2015). Application of metagenomics in the human gut microbiome. *World J. Gastroenterol.* 21, 803–814. doi: 10.3748/wjg.v21.i3.803
- Washington, N.L., Gangavarapu, K., Zeller, M., Bolze, A., Cirulli, E.T., Schiabor Barrett, K.M., et al. (2021). Genomic epidemiology identifies emergence and rapid transmission of SARS-CoV-2 B.1.1.7 in the United States. medRxiv. doi: 10.1101/2021.02.06.21251159
- Wenger, A. M., Peluso, P., Rowell, W. J., Chang, P. C., Hall, R. J., Concepcion, G. T., et al. (2019). Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nat. Biotechnol.* 37, 1155–1162. doi: 10.1038/s41587-019-0217-9
- White, P. L., Backx, M., and Barnes, R. A. (2017). Diagnosis and management of *Pneumocystis jirovecii* infection. *Expert Rev. Anti-Infect. Ther.* 15, 435–447. doi: 10.1080/14787210.2017.1305887
- Wickes, B. L., and Wiederhold, N. P. (2018). Molecular diagnostics in medical mycology. *Nat. Commun.* 9:5135. doi: 10.1038/s41467-018-07556-5
- Winand, R., Bogaerts, B., Hoffman, S., Lefevre, L., Delvoeye, M., Braekel, J. V., et al. (2019). Targeting the 16s rRNA gene for bacterial identification in complex mixed samples: comparative evaluation of second (Illumina) and third (Oxford Nanopore technologies) generation sequencing technologies. *Int. J. Mol. Sci.* 21, 1–22. doi: 10.3390/ijms21010298
- Xiao, C. L., Chen, Y., Xie, S. Q., Chen, K. N., Wang, Y., Han, Y., et al. (2017). MECAT: fast mapping, error correction, and de novo assembly for single-molecule sequencing reads. *Nat. Methods* 14, 1072–1074. doi: 10.1038/nmeth.4432
- Xu, Y., Lewandowski, K., Jeffery, K., Downs, L. O., Foster, D., Sanderson, N. D., et al. (2020). Nanopore metagenomic sequencing to investigate nosocomial transmission of human metapneumovirus from a unique genetic group among haematology patients in the United Kingdom. *J. Inf. Secur.* 80, 571–577. doi: 10.1016/j.jinf.2020.02.003
- Yamashita, T., Takeda, H., Takai, A., Arasawa, S., Nakamura, F., Mashimo, Y., et al. (2020). Single-molecular real-time deep sequencing reveals the dynamics of multi-drug resistant haplotypes and structural variations in the hepatitis C virus genome. *Sci. Rep.* 10:2651. doi: 10.1038/s41598-020-59397-2

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Hoang, Irinyi, Hu, Schwessinger and Meyer. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.