Advantages of nanopore sequencing in microbiome research
<table>
<thead>
<tr>
<th>Contents</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1</strong> Challenges of microbiome research</td>
</tr>
<tr>
<td><strong>2</strong> From microbiome to metatranscriptome and beyond</td>
</tr>
<tr>
<td><strong>3</strong> Kits and analysis workflows</td>
</tr>
<tr>
<td><strong>4</strong> Case studies</td>
</tr>
<tr>
<td><strong>5</strong> Summary</td>
</tr>
<tr>
<td><strong>6</strong> About Oxford Nanopore Technologies</td>
</tr>
<tr>
<td><strong>7</strong> References</td>
</tr>
</tbody>
</table>
Introduction

The study of microbiomes — the genetic material of all microorganisms in a given sample — has recently attracted considerable attention, mainly due to the realisation that the microbial composition of our bodies and environment can have a profound effect on our health.

The composition of the human microbiome, for example, has been connected to obesity\(^1\), immunity\(^2\) and psychiatric conditions\(^3\). Of note, there has also been a surge in research efforts focusing on analysing the microbiome of extreme environments\(^4\), water\(^5\), soil\(^6\), buildings\(^7\), etc.

Classical microbiome research relies on culturing, which is associated with long sample-to-result time and biases related to the different susceptibility of microorganisms to laboratory handling\(^8\).

The advent of modern sequencing technologies has brought significant advantages to the field. Speed and accuracy of microbial analysis (i.e. species identification and abundance) have increased substantially, with culturing no longer being necessary. Nonetheless, some challenges remain with traditional sequencing approaches.

This review will explore the challenges of microbiome research, providing real-world examples of how they are being overcome through the use of nanopore sequencing technology.
Challenges of microbiome research

Accurate microbe identification

There are two general approaches for the identification of microbes in a biological sample. The metagenomic sequencing approach results in the whole genome of any microbe present in the sample being sequenced. This unbiased method avoids the need for culturing and thus allows the identification of currently unculturable organisms. Nonetheless, traditional short-read sequencing technologies struggle to reliably resolve repetitive regions, making the taxonomic classification and genome assembly of closely-related species in such samples challenging. This and other concerns, such as time to result and cost-effectiveness, led to the introduction of 16S rRNA gene sequencing.

The 16S rRNA gene sequence is commonly used because of its combination of conserved and highly variable regions, allowing for accurate species distinction. However, the 16S primers typically used in these experiments do not cover the whole 16S gene region, which can reduce the resolution of species identification as some taxonomically informative regions fall outside of the analysed amplicon.

Nanopore technology allows for the design of primers covering the whole length of the 16S gene or even whole ribosomal operons.

The 16S rRNA gene sequence is commonly used because of its combination of conserved and highly variable regions, allowing for accurate species distinction. However, the 16S primers typically used in these experiments do not cover the whole 16S gene region, which can reduce the resolution of species identification as some taxonomically informative regions fall outside of the analysed amplicon.

Nanopore long reads are much more suitable for the assembly of repetitive genomic regions, resulting in improved precision of metagenomic species identification. Furthermore, in 16S-based studies, nanopore technology allows for the design of primers covering the whole 16S gene or even whole ribosomal operons, often leading to nanopore sequencing outperforming traditional sequencing platforms in the number of species accurately distinguished.

Analysis at point of collection

For the most accurate understanding of microbiome composition, samples should be analysed quickly — ideally at the site of collection. However, due to the large size and infrastructure requirements of traditional sequencing machines, researchers often collect samples, store them for long periods of time (risking loss of material or change in sample composition while in storage) and transport them to centralised laboratories.

Oxford Nanopore’s MinION™, weighs less than 100 g, is easily transportable in aeroplane luggage and is powered to sequence DNA or RNA using the USB port on a laptop or the IT accessory, MinIT™, making it suitable for mobile research setups. As a result, samples do not need to be frozen and stored for weeks or even days prior to analysis. For example, the time from sample collection to sequencing data generation with the MinION in an Arctic environment was reported to be just under 40 hours. Oxford Nanopore is also developing Flongle™, a flow cell adapter designed to provide even more cost-effective analysis of smaller, more frequently performed tests and...
experiments. Furthermore, portable library preparation, requiring the bare minimum of laboratory equipment, is possible using Oxford Nanopore’s VolTRAX™ system. Together, these devices will deliver the flexibility to collect and analyse samples onsite, which not only reduces time to result, but also allows the number of samples collected to be informed by the data generated, preventing under or over sampling and further streamlining the project workflow.

Streamlined workflows coupled with real-time analysis enable metagenomic species identification in as little as 20 minutes — from sample to result15.

Reduced time to result
When monitoring changes in the microbiome and handling many samples, being able to process samples quickly and easily has many advantages. For example, in clinical research and disease surveillance applications, the potential for accurate and rapid sample analysis could, in the future, lead to faster implementation of appropriate therapeutic intervention or pathogen containment strategies14.

Nanopore sequencing allows sample analysis both in the lab or field, with the most recent transposase-based library preparation kits from Oxford Nanopore requiring just 10 minutes of hands-on time (after genomic DNA extraction).

In combination with real-time data analysis, these streamlined workflows enable applications such as species identification to be achieved in as little as 20 minutes — from sample to result15. The facility for real-time sequencing also allows the sequencing run to be stopped as soon as sufficient data has been generated or a particular species identified, further highlighting the benefits of nanopore sequencing for the delivery of rapid and potentially actionable information.
The advent of high-throughput sequencing techniques has not only enabled researchers to identify species and construct genomes from microbiome samples, but also to investigate gene expression. Such metatranscriptomic studies can provide more detailed insight into the interaction of complex microbial communities and their response to specific environments.

Expanding upon this, the role of DNA and RNA modifications (e.g. methylation) in microbial communities is also of increasing interest; however, such modifications are cumbersome to detect with traditional technologies due to the need for bisulfite treatment or antibody pull-down sample pre-processing.

Nanopore platforms are capable of sequencing genomic DNA and cDNA from a metagenomic or target-enriched sample, generating long reads suitable for accurate taxonomic classification and transcriptomics investigation.

The technology is also capable of performing direct RNA sequencing without introducing reverse transcription- or PCR amplification-related biases. Furthermore, the facility for direct DNA and RNA sequencing enables the detection of chemical modifications, such as methylation, without the need for bisulfite treatment or antibody-based assays.

Accordingly, nanopore sequencing technology facilitates more complete characterisation of microbial communities than was previously possible on a single platform, dramatically reducing the time and cost associated with such studies.
**Kits and analysis workflows**

Oxford Nanopore offers easy-to-follow workflows for both metagenomic and 16S-based sequencing approaches used in microbiome research.

**16S analysis**

Oxford Nanopore has released a 16S library preparation kit with barcoding for cost-effective species identification from multiple samples (Figure 4). The whole workflow can be performed in just 40 minutes following genomic extraction, with less than 20 minutes hands-on time\(^2\). Additionally, a bioinformatics workflow for the analysis of 16S data is available, delivering a report that is updated in real-time as the sequencing progresses. The report is based on the result of comparing basecalled sequences (via BLAST) against the NCBI 16S bacterial database.

**Metagenomic analysis**

Oxford Nanopore provides a range of DNA library preparation kits, including low-input and barcoding options to suit all experimental requirements. Library preparation can be undertaken in as little as 10 minutes. Data analysis is achieved using the WIMP bioinformatics workflow, which allows the identification of bacteria, fungi, archaea and viruses in a metagenomic sample in real-time as the sequencing progresses (Figure 5). WIMP uses the Centrifuge software tool\(^2\), which is capable of accurately identifying reads when using databases containing multiple highly similar reference genomes, such as different strains of a bacterial species. The WIMP workflow also allows quantification of the microorganisms present in the sample.

**Figure 4**
Rapid microorganism identification using 16S sequencing with real-time analysis.
Nanopore sequencing library preparation kits are available for both direct and indirect (cDNA) analysis of RNA. Full-length cDNA analysis can be achieved from just 50 ng input RNA, with library preparation taking less than 1.5 hours.

Direct RNA sequencing, allows simultaneous analysis of epigenetic modifications and nucleotide sequence from full-length RNA transcripts. Researchers have developed a number of analysis pipelines that offer read alignment, isoform identification and quantification.
4

Case studies
Necrotising enterocolitis (NEC) is one of the most devastating gastro-intestinal diseases affecting pre-term and low-birth-weight babies. Microbial dysbiosis (imbalance) has been established as the most likely underlying cause of the disease. Previous studies have revealed that pre-term babies suffering from NEC had more potentially pathogenic bacteria and less of the beneficial Bifidobacterium bifidum.

Dr. Matthew Clark and Dr. Richard Leggett’s teams at the Earlham Institute in Norwich, UK, are utilising metagenomic nanopore sequencing to profile pre-term infant gut microbiota, identifying not just species but also quantifying them and profiling their antimicrobial resistance (AMR) genes. Their study is being conducted as part of a clinical trial for probiotic treatment with Dr. Lindsay Hall at the Quadram Institute of Biosciences, Norwich, UK.

Through a time-course experiment, the team was able to observe the impact of probiotic and antibiotic treatments on microbiota composition. As the probiotic was administered, the amount of B. bifidum (identified to the species level) in the baby’s gut increased, with both long-read nanopore sequencing and short-read platforms giving similar profiles.

Interestingly, the researchers also noted that the nanopore metagenomic sequencing approach allowed more detailed taxonomic assignment when compared with 16S rRNA analysis using a short-read sequencing platform, which failed to differentiate members of the Enterobacteriaceae family.

Utilising real-time analysis of a faecal sample from a critically ill infant, the team were able to confidently call the presence of the pathogen Klebsiella pneumoniae along with corresponding AMR profiles within 1 hour of sequencing.

Nanopore sequencing data was found to be of similar quality to that provided by short-read sequencing technology but, critically, it could be generated in a fraction of the time, requiring just 5 hours as opposed to 39 hours for the short-read technology.

In conclusion, the researchers highlighted how nanopore technology could, in the future, be applied to clinical applications: “Our results demonstrate that MinION sequencers offer the ability to progress from clinical samples to a potential tailored patient antimicrobial treatment in just a few hours.”

Case study 1
Baby-Associated MicroBiota of the Intestine (BAMBI) — an example from the clinic

Nanopore sequencing allowed pathogen identification and AMR profiling within 1 hour of sequencing.
Case study 2

The advantage of long reads in 16S-based microbiome studies

Mouse gut microbiome

16S-based species identification is the most commonly used method in microbiome studies. Shin et al. examined the composition of mouse gut microbiota using a 16S approach and either short-read sequencing technology or long-read nanopore sequencing.

A key difference between the two approaches was that, for short-read experiments, the hypervariable V3-V4 16S region was used for primer design and amplicon generation (mean read length – 447 bp), while the full length of the 16S gene was used for the nanopore experiments (mean read length – 1393 bp).

Sequencing data derived from the two platforms were highly concordant (R²=0.8-0.9) at all taxonomic levels apart from the species level. It was shown that nanopore data allowed the identification of more species (i.e., *Bifidobacterium animalis* and *Bifidobacterium pseudolongum*, well-known members of the gut microbiome), resulting in the construction of a fuller microbiota representation.

The study illustrates that the common use of the V3-V4 16S region for species identification is not as sensitive as using the full length of the 16S gene, as allowed by nanopore sequencing.

In conclusion, nanopore sequencing’s long read advantage was instrumental for the generation of a more accurate profile of the mouse gut microbiome compared to short-read technology.

Figure 6

Variability in the 16S rRNA gene sequence. Detected variants between the 16S rRNA gene sequences are represented as vertical lines on the 16S rDNA sequences. The black and red arrows indicate the binding positions of primer sets for the amplification of V3-V4 regions (for short-read sequencing) and nearly full-length regions of 16S rDNA sequences (nanopore sequencing), respectively. Image adapted from Shin et al.
Case study 3
Using ribosomal operons for species identification

Professor Lee Kerkhof from Rutgers University employed nanopore sequencing for bacterial microbiome profiling using whole rRNA operon sequences instead of only 16S\textsuperscript{12,18}. The forward primer was generated in the 16S rRNA gene sequence and the reverse primer was within the 23S rRNA gene, resulting in a 4.2 kb fragment (Figure 7).

The team sequenced 6 environmental samples, each comprised of farm soil DNA and bioreactor DNA mixed in varying proportions. Twelve hours of sequencing generated over 1000 operational taxonomic units (OTUs)*, with good concordance between replicates. The team was able to build consensus operons where each rRNA operon contained a 16S and a 23S gene.

The study concluded that nanopore sequencing of operons allowed better taxonomic resolution than standard 16S sequencing using short-read technology. The team also reported that their approach yielded accurate quantification of OTUs.

In summary, Professor Kerkhof commented that: \textit{‘Our analysis demonstrated that the MinION has the ability to provide rRNA operon sequence data of sufficient quality for characterising the microbiota of complex environmental samples and provided results that are reproducible, quantitative, and consistent’}\textsuperscript{28}. The team are now extending their research to elucidate the impact of external factors (e.g. toxin exposure) on the composition of the mouse microbiome\textsuperscript{12}.

Figure 7
Amplicon structure for ribosomal operon sequencing (4.2 kb), containing almost full length 16S and 23S sequences. Image adapted from Kerkhof\textsuperscript{12}.

* Organisms are clustered into a single OTU based on similarity of DNA sequences above a pre-determined threshold.
Case study 4
Portable shotgun metagenomic sequencing of supraglacial microbiota

Glaciers have rich microbiota, which play an active role in maintaining the glacier’s properties and determining melting speed\textsuperscript{13}. As such, the study of changing glacial microbiomes is important for understanding the effects of global warming. Dr. Arwyn Edwards from the Aberystwyth University (UK) pioneered the establishment of a portable lab setup in Svalbard (between Norway and the North Pole) using Oxford Nanopore’s MinION to study glacial microbiomes\textsuperscript{13}.

Dr. Edwards’ team set out to study the microbiomes of a cryoconite (windblown deposits comprised of rock particles and microbes) and a red snow algal bloom \textit{in situ} using a metagenomic approach\textsuperscript{13} (Figure 8). All of the equipment necessary for this study was transported in 2 standard sized airplane pieces of luggage. The team was able to obtain raw sequencing data within 36 hours of sample collection. The WIMP bioinformatics workflow used to resolve microbe composition of the sequenced metagenomic samples revealed a dominant presence of Proteobacteria, as expected from previous studies using short-read technology. The majority of the remaining microbes identified also followed prior expectations.

Recently, the team have assessed the utility of 16S nanopore sequencing to understand the impact of different ecological conditions on microbial communities in the field, with initial results showing significant changes in microbiome composition\textsuperscript{29}.

Dr. Edwards’ results demonstrate the utility of nanopore sequencing for accurate microbiome analysis in remote and extreme, resource-limited environments.

\textbf{Figure 8}
Sample collection from a cryoconite (a) and red snow algal bloom (b) in Svalbard. Images courtesy of Dr. Arwyn Edwards, Aberystwyth University, UK.
Case study 5

Pink microbiome

Lake Hillier in Western Australia is one of several sites targeted for microbiome investigation by the eXtreme Microbiome Project (XMP - extrememicrobiome.org). The lake’s water is about 10 times saltier than sea water and has a bright-pink colour4 (Figure 9).

The unusual colouration was thought to be due to a type of red algae (*Dunaliella salina*), but no formal investigations had been performed until recently. Dr. Ken McGrath’s team from the Australian Genome Research Facility collected water samples from the top and lower levels of the lake, as well as sediment samples, for metagenomic analysis.

Traditional culture-based analysis managed to resolve 13 different microorganisms, with a culture-dependent bias towards bacterial species4. A 16S-based microbial profile using short-read sequencing technology built a more complete taxonomic representation, but species-level identification was problematic. In contrast, a whole-shotgun DNA sequencing approach using nanopore technology resulted in species-level taxonomic classification, with a more complete profile starting to emerge after just 2 hours of sequencing. The final results were available after 24 hours of sequencing, with successful identification of algae, archaea, bacteria and viruses.

The XMP found that the red bacterium *Salinibacter ruber* dominated the metagenomics profile of the lake, whereas *D. salina* was present at relatively low levels, and concluded that the pink colour of the lake is produced by the bacteria, and not the algae as previously thought.

The Lake Hillier microbiome study, as well as the rest of the XMP work, has put nanopore sequencing at the forefront of microbiome research in extreme understudied environments.

**Figure 9**

Lake Hillier. Image courtesy of Dr. Ken McGrath, Australian Genome Research Facility, Australia.
Case study 6
Direct sequencing of 16S rRNA

The majority of microbiome studies to date have focused on the analysis of DNA to elucidate microbial composition and genome structure; however, advances in sequencing technology and workflows, have allowed researchers to investigate RNA, providing insights into gene expression in complex bacterial communities. The analysis of epigenetic modifications adds a further level of characterisation to microbial communities and their environmental responses; however, such modifications are erased using traditional sequencing techniques.

In a proof-of-principle study, Dr. Andrew Smith and colleagues at the University of California, Santa Cruz, utilised nanopore sequencing to directly analyse full-length (1.5 kb) 16S ribosomal RNA (rRNA) of *Escherichia coli* to characterise both nucleotide and epigenetic composition in a single sequencing run. 16S rRNA is a core constituent of the small ribosomal subunit which is expressed in all living cells and plays an essential role in RNA translation. Numerous antibiotics target prokaryotic ribosomes which can acquire resistance via nucleotide substitutions, or by gain or loss of base modifications, making their analysis increasingly important.

The team developed an enrichment strategy that allowed the analysis of just 5 pg of *E. coli* 16S rRNA in a background of 4.5 µg total human RNA. Furthermore, the team reported that analysis could be performed in a potentially clinically actionable timeframe of just 2 hours from sample to result. In fact, the real-time analysis provided by nanopore technology enabled 16S rRNA reads to be generated within 20 seconds of starting the sequencing run.

Examination of the data allowed the identification of 7-methyl guanosine (m7G) at known positions in the 16S rRNA (Figure 10) and indicated the presence of additional epigenetic modifications, including pseudouridine.

In summary, the team suggested that: ‘This [nanopore] direct RNA sequencing technology has promise for rapid identification of microbes in the environment and in patient samples.’

* Nanopore devices are currently for research use only.
The advent of sequencing has introduced significant improvements in the field of microbiome research, with metagenomic and 16S-based experiments achieving faster and more accurate microbiota profiling than previous culture-based methods\(^4,8\); however, a number of challenges remain to fully analysing microbiome composition.

Nanopore sequencing is now being used to address these challenges, delivering portable and scalable options for microbiome research, with rapid, streamlined workflows, incorporating automated library preparation and real-time data analysis.

Furthermore, the facility for direct DNA and RNA sequencing allows the detection of epigenetic modifications alongside sequence data — adding further depth to our understanding of microbiomes and the complex interactions between microorganisms and their host environment. Importantly, long nanopore reads allow complete 16S rRNA genomic regions\(^10\) or even whole ribosomal operons\(^12\) to be resolved, leading to a more detailed, species-level classification\(^10\).

Long reads also provide a more accurate representation of species diversity in metagenomic samples\(^4,11\).
About Oxford Nanopore Technologies

By introducing the MinION — the first portable, real-time, long-read nanopore sequencing machine — Oxford Nanopore Technologies has expanded the capabilities of microbiome researchers. These specifications make the technology suitable for on-site analysis of microbiomes with streamlined workflows and real-time analysis delivering rapid access to results. Nanopore technology is fully scalable, with the GridION™ X5 and PromethION™ devices providing flexible sequencing capacity for high-throughput applications.

For the latest news, products, application information and developments in nanopore sequencing, visit: www.nanoporetech.com


