Oxford NANOPORE Technologies

Metagenomic sequencing with Oxford Nanopore

GETTING STARTED

nanoporetech.com

Introduction

Metagenomics can be defined as the analysis of the community of genomes present within an isolated sample, and is a term predominantly applied to the detection and analysis of microorganisms.

The rapid and significant decrease in sequencing costs has accelerated the field of metagenomics in recent years, enabling substantial progress in our understanding across a wide range of applications, from clinical research and outbreak surveillance, to biofuel development and crop selection **(Figure 1)**. The development of metagenomic sequencing has allowed the rapid identification and analysis of culturable microorganisms, and, importantly, has made possible the analysis of those microbes that cannot be cultured.

Oxford Nanopore sequencing technology provides a number of key benefits for metagenomics research **(Table 1)**. There is no upper read length with nanopore sequencing — reads of any length are produced, from short to ultra-long. Long sequencing reads enhance genome assembly, enabling more accurate analysis of known and novel microbes, and precise differentiation of closely-related microbes. Reads exceeding 4 Mb have been generated with nanopore technology[1], meaning that entire microbial genomes can be obtained in single reads, or with minimal contigs (uninterrupted stretches of overlapping DNA) **(Figure 2)**. Long reads also improve the resolution of repeat sequences and structural variants, further enhancing genome assembly and antimicrobial resistance (AMR) gene analysis.

Table 1. Advantages of nanopore technology for metagenomic sequencing

Easier assembly
Long sequencing reads mean fewer fragments to assemble

Accurate species identification
Long sequencing reads enable differentiation between closely-related organisms

Structural variant and repeat resolution
Long reads can span entire structural variants and repeat segments in single reads

Real-time analysis
Sequencing reads can be basecalled and analysed as sequencing progresses, enabling rapid time to result

Extensive analysis solutions
From simple point-and-click workflows for microbial identification and AMR detection to routine workflows for metagenomic analysis (see page 10)

Cost-effective and scalable
A range of sequencing devices are available to suit all project sizes and experimental goals

Portable
Sequence at sample source with portable MinION™ and Flongle™ devices

Figure 1: Nanopore metagenomic sequencing has been used across a range of applications and environments, including in-field sequencing to understand microbiomes in extreme environments

Images courtesy of: (from left to right) Arwyn Edwards, University of Aberystwyth; Ken McGrath, Australian Genome Research Facility; NASA's Johnson Space Center.

1. Oxford Nanopore Technologies (2022). Ultra-Long DNA Sequencing Kit. Available at: store.nanoporetech.com/ultra-long-dna-sequencing-kit-v14.html [Accessed: 15 May 2023]

Nanopore sequencing technology provides rapid turnaround times with real-time sequencing, ideal for, for example, outbreak scenarios where rapid detection and response are required for effective pathogen surveillance. Metagenomic samples can be analysed anywhere, even in extreme, isolated environments, with the portable MinION, MinION Mk1C, and Flongle devices. With low start-up costs ($1,000 for a MinION Starter Pack, and Flongle Flow Cells for $90 each) and the facility for sample multiplexing, low-cost sequencing is readily achievable. Nanopore technology is scalable and flexible, so you can sequence according to your needs — the modular GridION™ (5 x MinION or Flongle Flow Cells) and PromethION™ range (2, 24, or 48 PromethION Flow Cells) enable on-demand sequencing for higher throughput requirements.

For more information on how long nanopore sequencing reads benefit genome assembly, and the scalability of nanopore sequencing devices, view the 'Whole-genome sequencing: small genomes' Getting Started guide: **nanoporetech.com/resource-centre/guide-whole-genome-sequencing-small-genomes**

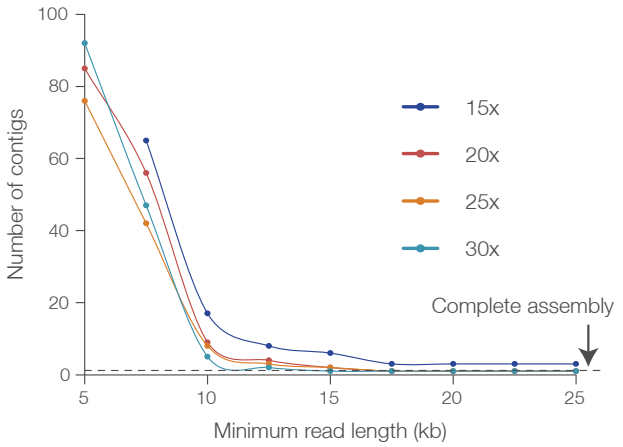


Figure 2. Longer reads facilitate assembly completion from lower-coverage data

Should I use 16S or WGS?

Although, by definition, metagenomics involves whole-genome analysis, 16S rRNA sequencing should also be considered in this context, being a widely used method of microbial identification since the advent of high-throughput next-generation sequencing.

There are a number of differences between 16S and whole-genome sequencing (WGS) approaches to consider, including the degree of taxonomic resolution, cost, and amount of data required for analysis **(Table 2)**.

Another alternative is sequencing the entire rRNA operon, which provides better taxonomic resolution than 16S alone[2], but due to a lack of rRNA sequence databases for subsequent identification, and primer bias across this region, some species present may be missed, and measures of relative abundance could be distorted. Moreover, the rRNA genes may be separated across the genome, rather than linked within a single operon, as is estimated to be the case in 41% of soil microbial genomes; therefore, the degree of success of this approach may depend on sample origin[3].

For identifying organisms of interest, 16S rRNA analysis may be sufficient, although the extent of species-level resolution is limited. Nanopore technology offers full-length 16S rRNA gene analysis, which has been shown to provide superior taxonomic resolution to traditional short-read approaches that analyse just a small region of the 16S gene. For truly unbiased detection, rare pathogen identification, up to strain-level resolution, and AMR and variant analyses, WGS is recommended.

Incorporating a negative control into a 16S or WGS metagenomic experiment should also be considered, to avoid false-positive results from microbial contaminants which may be obtained during sample collection and/or processing.

Table 2. Comparison between 16S and whole-genome sequencing (WGS) approaches

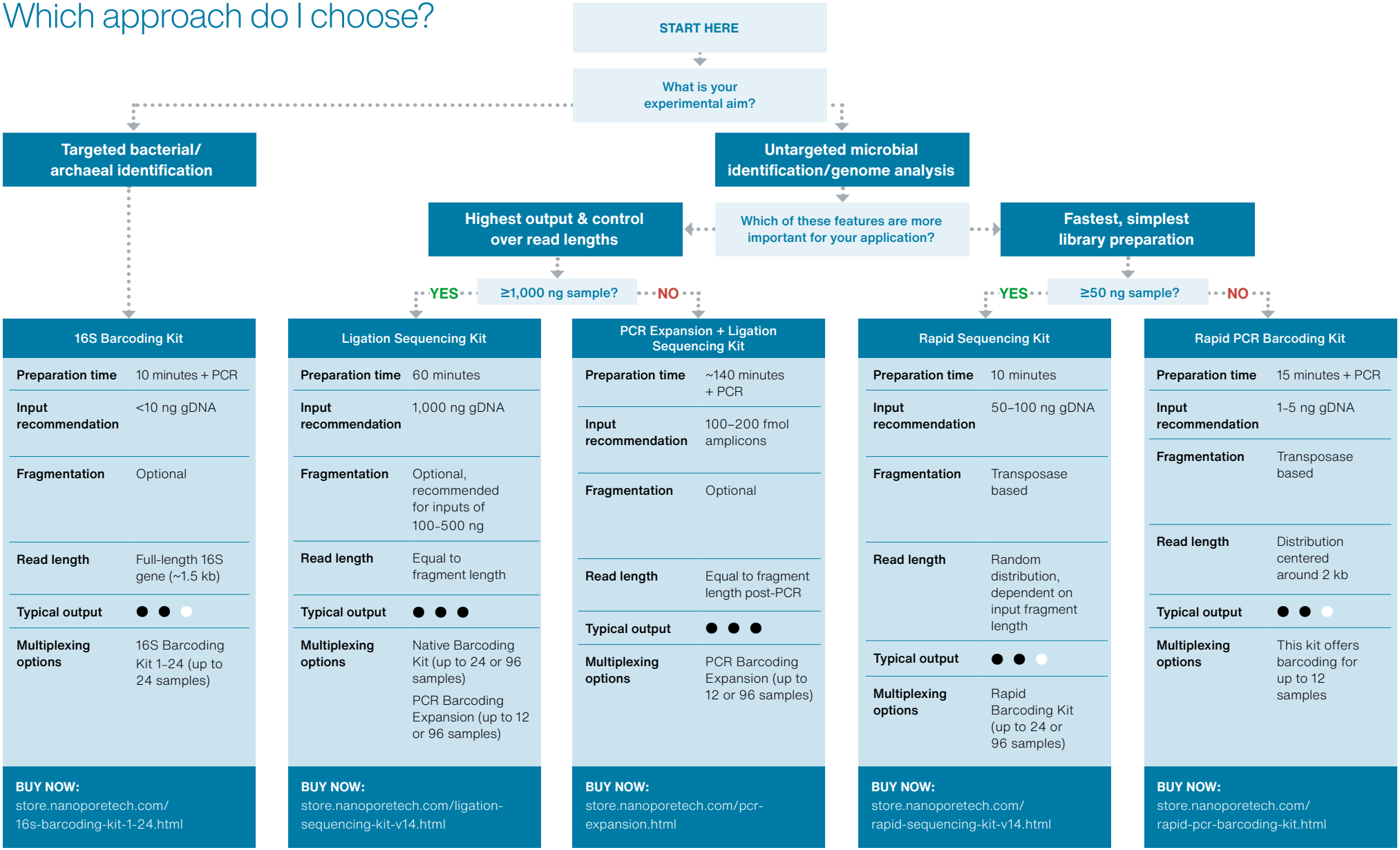
	16S	WGS
Taxonomic resolution	Phylum > genus*	Species > strain
Advantages	• Enriches over background (e.g. host DNA) • Lower cost than WGS; less sequencing data required per sample • No culture required	• Provides additional information (e.g. antimicrobial resistance) • Can identify different microorganisms beyond bacteria (e.g. viruses, fungi, protozoa) • No culture required • PCR not required; unbiased approach
Disadvantages	• PCR required; primers may not work across all bacteria/archaea; possible PCR artefacts • Limited to bacterial and archaeal ID • No additional information obtained (e.g. antimicrobial resistance)	• Does not enrich over background (e.g. host DNA) • More expensive than 16S ID; more sequencing data required per sample • High data requirements for low microbial counts
Cost	$	$ $ $
Multiplexing capacity	24	96 (with or without PCR)
Sequence data needed	Mb → Gb	Mb → 10s of Gb†

* Genus and even species level may be possible in a limited number of cases.
† Application dependent — e.g., identification of prevalent organism vs. *de novo* assembly of genomes present.

2. Cusco, C. et al. Microbiota profiling with long amplicons using nanopore sequencing: full-length 16S rRNA gene and the 16S-ITS-23S of the rrn operon. *F1000 Research*. 7:1755, DOI: doi.org/10.12688/f1000research.16817.2 (2019).

3. Brewer, T. E. et al. Unlinked rRNA genes are widespread among bacteria and archaea. *The ISME Journal*. 14:597–608, DOI: doi.org/10.1038/s41396-019-0552-3 (2019).

Which approach do I choose?



16S Barcoding Kit	
Preparation time	10 minutes + PCR
Input recommendation	<10 ng gDNA
Fragmentation	Optional
Read length	Full-length 16S gene (~1.5 kb)
Typical output	● ● ○
Multiplexing options	16S Barcoding Kit 1-24 (up to 24 samples)

BUY NOW:
store.nanoporetech.com/16s-barcoding-kit-1-24.html

Ligation Sequencing Kit	
Preparation time	60 minutes
Input recommendation	1,000 ng gDNA
Fragmentation	Optional, recommended for inputs of 100–500 ng
Read length	Equal to fragment length
Typical output	● ● ●
Multiplexing options	Native Barcoding Kit (up to 24 or 96 samples) PCR Barcoding Expansion (up to 12 or 96 samples)

BUY NOW:
store.nanoporetech.com/ligation-sequencing-kit-v14.html

PCR Expansion + Ligation Sequencing Kit	
Preparation time	~140 minutes + PCR
Input recommendation	100–200 fmol amplicons
Fragmentation	Optional
Read length	Equal to fragment length post-PCR
Typical output	● ● ●
Multiplexing options	PCR Barcoding Expansion (up to 12 or 96 samples)

BUY NOW:
store.nanoporetech.com/pcr-expansion.html

Rapid Sequencing Kit	
Preparation time	10 minutes
Input recommendation	50–100 ng gDNA
Fragmentation	Transposase based
Read length	Random distribution, dependent on input fragment length
Typical output	● ● ○
Multiplexing options	Rapid Barcoding Kit (up to 24 or 96 samples)

BUY NOW:
store.nanoporetech.com/rapid-sequencing-kit-v14.html

Rapid PCR Barcoding Kit	
Preparation time	15 minutes + PCR
Input recommendation	1–5 ng gDNA
Fragmentation	Transposase based
Read length	Distribution centered around 2 kb
Typical output	● ● ○
Multiplexing options	This kit offers barcoding for up to 12 samples

BUY NOW:
store.nanoporetech.com/rapid-pcr-barcoding-kit.html

From sample to answer

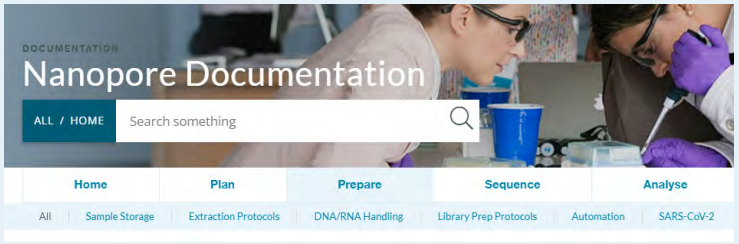
QUESTION

SET-UP

This is my first nanopore sequencing experiment. Where do I start?

Firstly, you will need to set up your sequencing device, download the required software, and then prepare and run a control sequencing experiment. This checks that everything is working as it should and helps to familiarise users with our library prep and sequencing workflow. Our step-by-step guides take you through this entire process, with easy-to-follow instructions for every step of the way.

View our step-by-step guides:
community.nanoporetech.com/getting_started

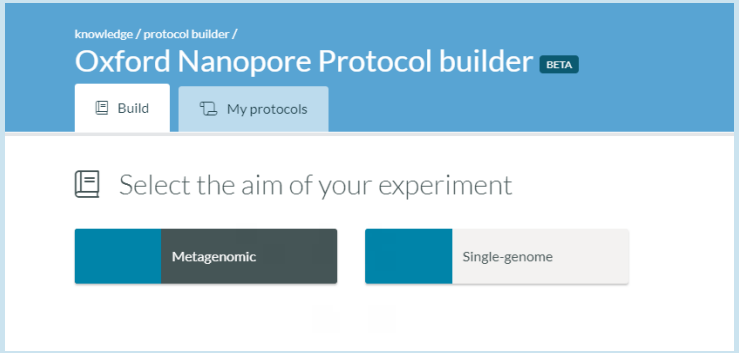


PLANNING

How do I design my protocol?

The Oxford Nanopore Protocol Builder is an interactive tool that enables you to generate your own end-to-end protocol, with application-specific advice encompassing DNA extraction, library prep, sequencing, and data analysis.

Create your bespoke metagenomic sequencing protocol:
community.nanoporetech.com/knowledge/protocol_builder



From sample to answer

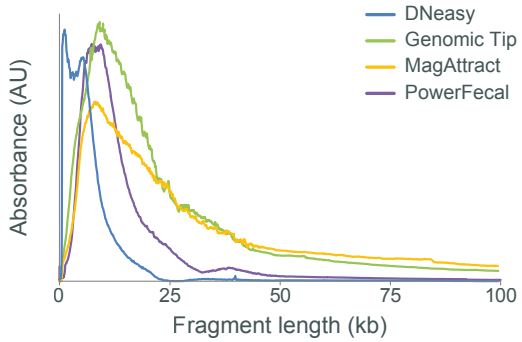
EXTRACTION

How can I best extract high-quality DNA from my sample?

The 'Prepare' Documentation section of the Nanopore Community features recommended DNA extraction methods and comparisons for a wide range of sample types, such as stool, soil, and environmental water samples, plus data on the effects of carryover of contaminants, such as phenol and ethanol, on library prep efficiency. The resource also includes guidance on size selection; if performing metagenomic assembly, we would recommend size selecting for long fragments prior to sequencing to facilitate the downstream assembly process.

Read more about recommended extraction methods for your sample:
community.nanoporetech.com/docs/prepare

Fragment length comparison of methods for DNA extraction from stool



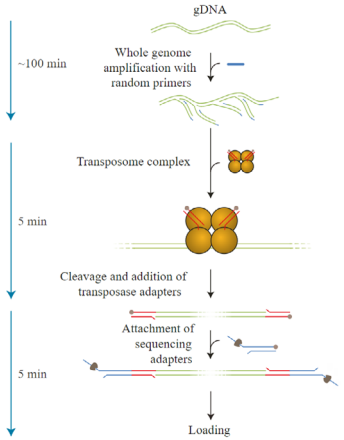
DNA SAMPLE

How much DNA do I need?

Our kit selection workflow (see page 5) displays sample input recommendations, with both PCR-based and PCR-free options depending on how much starting material is available. When starting with less than 1 ng of DNA, whole-genome amplification (WGA) can be performed to generate sufficient sample for sequencing using the rapid whole-genome amplification protocol. This protocol uses the QIAGEN REPLI-g UltraFast Mini Kit to amplify from >1 ng starting DNA, or directly from cells, in under two hours; the library is then prepared in 10 minutes with the Rapid Sequencing Kit.

View our PCR-based, PCR-free, and WGA protocols:
community.nanoporetech.com/protocols

Rapid whole-genome amplification workflow



From sample to answer

EXTRACTION

Should I perform host depletion?

Depending on your sample type and experimental aims, you may choose to perform host depletion. Clinical research samples (e.g. respiratory specimens and swabs) are likely to have significant host contamination; this is less of a problem for environmental samples (e.g. wastewater).

If performing 16S sequencing, host depletion is likely to be unnecessary. In contrast, as >95% of whole-genome sequencing data from clinical research samples may be host-derived, depletion will reduce costs and greatly increase the amount of relevant sequencing data. Metagenomic genome assembly, variant calling, and unbiased microbial identification all benefit hugely from host depletion. Methods commonly used for host depletion include: saponin, MolYsis kits, and rRNA depletion kits (for RNA viral metagenomic sequencing).

Find more depletion methods used by the Nanopore Community in recent publications:
nanoporetech.com/resource-centre

LIBRARY PREPARATION

Can I multiplex my samples?

Sample multiplexing increases the cost efficiency of sequencing, but will also reduce depth of coverage per genome, so the rarity of organisms, sample complexity, and presence of host DNA should be considered prior to choosing this approach (see page 10).

Barcoding options are available for ligation-based and rapid library preps, for both PCR-based and PCR-free protocols:

Ligation:

PCR-free:	Native Barcoding Kit (up to 24 or 96 samples)
PCR-based:	Ligation Sequencing Kit + PCR Barcoding Expansion (up to 12 or 96 samples)

Rapid:

PCR-free:	Rapid Barcoding Kit (up to 24 or 96 samples)
PCR-based:	Rapid PCR Barcoding Kit (up to 12 samples)

For 16S rRNA kit barcoding options, refer to the 'Which kit do I choose?' workflow on page 5.

View the nanopore sequencing kits and expansion packs:
nanoporetech.com/products/kits

From sample to answer

SEQUENCING

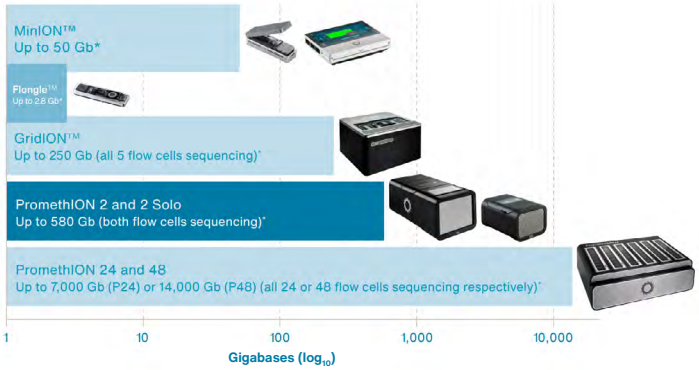
Which device should I choose?

The chemistry underpinning nanopore sequencing can be scaled up or down; our devices make full use of this. The portable MinION devices are ideal for sequencing at sample source, even in the most extreme environments: the MinION can be operated from a laptop, whilst the MinION Mk1C is an all-in-one sequencer with an LCD touchscreen and integrated compute. The high-throughput GridION provides on-demand benchtop sequencing, for up to five MinION or Flongle Flow Cells. For high to ultra-high-throughput, the PromethION can accommodate up to two (PromethION 2 and PromethION 2 Solo), 24 (PromethION 24) or 48 (PromethION 48) high-output PromethION Flow Cells. Lastly, the Flongle ('Flow Cell Dongle') device adapts MinION or GridION for use with Flongle Flow Cells.

Flongle is ideal if you want to perform low-pass sequencing to determine the dominant species in a metagenomic sample, or for performing an initial library QC run. However, for genome assembly, the MinION and PromethION Flow Cells, which provide higher sequencing output are recommended.

For more information on choosing a sequencing device that fits your application, visit nanoporetech.com/applications.

Output per device over 72 hours



*=Theoretical maximum output of device when run at 420 bases/second

View and compare nanopore sequencing devices:
nanoporetech.com/products/specifications

From sample to answer

SEQUENCING

How much sequencing data do I need?

The recommended sequencing data yield depends on your experimental aims and the sample being analysed (e.g. rarity of organism; complexity of sample; presence of host DNA).

For WGS metagenomics, we suggest the following as rough guidelines for sequencing depth of coverage (per organism):

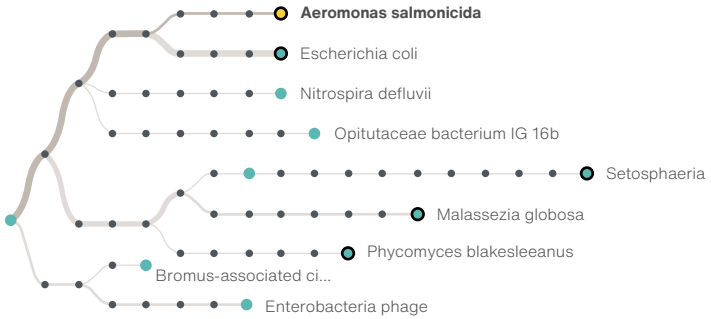
- Confirm presence of organism of interest: 10x
- Species-level ID: 20x
- AMR gene analysis: 20x
- Assembly: 30x
- Variant calling: 100x

Microbial identification with 16S vs. whole-genome sequencing

For 20x depth of the ~1.5 kb 16S locus, 30 kb of sequencing data is required, per microorganism. Therefore, bacteria/archaea with an abundance of ≥0.003% can be detected with 1 Gb of data.

For 20x depth of a 3 Mb genome, 60 Mb of sequencing data is needed. From 1 Gb of data, organisms with ≥6% abundance can therefore be identified. Sensitivity can be increased or decreased by increasing/decreasing the amount of data obtained, respectively. Sample multiplexing will decrease sensitivity but increase cost effectiveness.

Cumulative output: species-level taxonomic classification



View our end-to-end metagenomic assembly workflow:
nanoporetech.com/resource-centre/literature/metagenomic-assembly-workflow

DATA ANALYSIS

How can I analyse my data?

Oxford Nanopore provides a range of solutions for analysing metagenomic sequence data, covering all levels of bioinformatics expertise.

For cloud-based or local analysis, our EPI2ME solutions offer simple, point-and-click workflows via a user-friendly interface for routine metagenomic sample analysis **(Table 3)**.

Third-party tools can be found in our Resource Centre. We recommend MetaFlye as a highly accurate metagenomic assembly tool. Classification tools include Centrifuge and MetaMaps.

Find out more about nanopore sequencing analysis:
nanoporetech.com/analyse

Table 3. EPI2ME tools for metagenomic analysis

Workflow	Description	Output
16S	Real-time family>genus-level identification of bacteria and archaea, using the 16S rRNA gene	Classification report
What's in my pot?	Rapid species-level identification of fungi, bacteria, viruses, or archaea, from metagenomic samples	Real-time building of a taxonomic tree
ARMA	Builds on the 'What's in my pot?' workflow, with real-time antimicrobial resistance profiling	Report detailing the resistance genes found

Case studies

Case study 1: Nanopore-only assembly of reference-quality bacterial genomes from wastewater samples

Wastewater treatment plants employ a variety of processes for treating sewage. One method is the activated sludge process, in which organic products and toxins are broken down by a complex microbial community. The microbes in activated sludge may contain functional strains for degradation of waste compounds or pathogenic microorganisms; of particular concern is the presence of antimicrobial resistance (AMR) genes[4]. Assembling metagenomes from these complex communities using traditional short-read sequencing is challenging, and fails to resolve related species and strains which may contain long sequences of near-identical DNA[5].

Professor Albertson and colleagues, based at Aalborg University, Denmark, assessed the performance of R9 and the more recent R10 nanopore chemistries in assembling genomes from mixed microbial communities[5]. They noted that, in the past, researchers have often performed polishing of nanopore sequencing data via reference polishing or using short-read data when generating microbial assemblies. First sequencing a mock microbial community and then moving on to an activated sludge sample, they demonstrated that, with R10.4 nanopore chemistry, they were able to obtain *'near-finished'* bacterial genome assemblies using nanopore sequencing alone. They concluded that '*Oxford Nanopore R10.4 enables the generation of near-finished microbial genomes from pure cultures or metagenomes at coverages of 40-fold without short-read polishing*'.

Read the publication:
nanoporetech.com/resource-centre/high-quality-low-cost-nanopore-only-bacterial-genome-sequences

Case study 2: Assembly-free nanopore sequencing analysis of viral metagenomes

To gain a greater understanding of the diversity and nature of ocean viruses, Beaulaurier and colleagues investigated the genomes of ocean phages[6]. Water samples were taken at three different depths (15 m, 117 m and 250 m) and filtration was used to capture viral particles. Prepared viral genomic libraries were sequenced on the GridION device and an assembly-free analysis pipeline was developed to isolate and polish full-length phage genomes. The analysis method maintained complex repeat structures, such as direct terminal repeats, contrasting with typical short-read metagenomic assembly approaches, for which repeats tend to be collapsed into single sequences.

Read lengths of up to 254 kb were obtained, which almost entirely covered the typical size range of dsDNA viruses that infect bacteria and archaea (~3–300 kb). In total, over 1,000 high-quality, full-length draft virus genomes were generated. These genomes could not be fully recovered using short-read assembly approaches. Interestingly, as sampling depth increased, the proportion of phage genomes that were well characterised taxonomically decreased, demonstrating that deep-sea viral populations have been significantly less well defined[6].

Read the publication:
nanoporetech.com/resource-centre/assembly-free-single-molecule-nanopore-sequencing-recovers-complete-virus-genomes

4. Cheng, X. et al. Metagenomic profiling of antibiotic resistance and virulence removal: Activated sludge vs. algal wastewater treatment system. *J Environ Manage.* 295 (1), DOI: doi.org/10.1016/j.jenvman.2021.113129 (2021).
5. Sereika, M. et al. Oxford Nanopore R10.4 long-read sequencing enables the generation of near-finished bacterial genomes from pure cultures and metagenomes without short-read or reference polishing. *Nat Methods.* 19, 823-826, DOI: doi.org/10.1038/s41592-022-01539-7 (2022).
6. Beaulaurier, J. et al. Assembly-free single-molecule sequencing recovers complete virus genomes from natural microbial communities. *Genome Res.* 30(3):437-446, DOI: doi.org/10.1101/gr.251686.119 (2020).

Oxford Nanopore Technologies
Phone: +44 (0)845 034 7900
Email: support@nanoporetech.com
Twitter: @nanopore

www.nanoporetech.com

GS_1059(EN)_V3_31May2023