Oxford
NANOPORE
Technologies

METAGENOMIC SEQUENCING
WITH OXFORD NANOPORE

Getting Started

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 detection, 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 which 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 of up to 2 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.

Figure 1. Nanopore metagenomic sequencing has been used across a range of applications and environments, including in-field outbreak surveillance, and understanding 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.

Reference

1. Payne, A., Holmes, N., Rakyan, V., and Loose, M. BulkVis: a graphical viewer for Oxford nanopore bulk FAST5 files. Bioinformatics. 35(13):2193–2198 (2019).

Nanopore sequencing platforms provide rapid sample-to-answer with real-time sequencing, ideal for clinical research and outbreak scenarios where rapid detection and response are required. Metagenomic samples can be analysed anywhere, even in extreme, isolated environments, with the MinION™ and Flongle™ portable sequencing 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™ (24/48 PromethION Flow Cells) enable on-demand sequencing for higher throughput requirements.

For more information on how nanopore long sequencing reads benefit genome assembly, and the scalability of nanopore sequencing platforms, please see the Whole genome sequencing: small genomes Getting Started guide, found in the Resource Centre on our website.

Table 1. Advantages of nanopore technology for metagenomic sequencing

- **Easier assembly** *long sequencing reads mean fewer fragments to assemble*

- **Easier species differentiation** *long sequencing reads enhance differentiation of 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*

- **Analysis solutions** *simple microbial identification and AMR analysis workflows are available via EPI2ME (see page 6)*

- **Cost-effective and scalable** *a range of sequencing platforms are available to suit all project sizes*

- **Portable** *sequence at sample source with portable MinION and Flongle devices*

Figure 2. Longer reads allow complete assembly from lower-coverage data
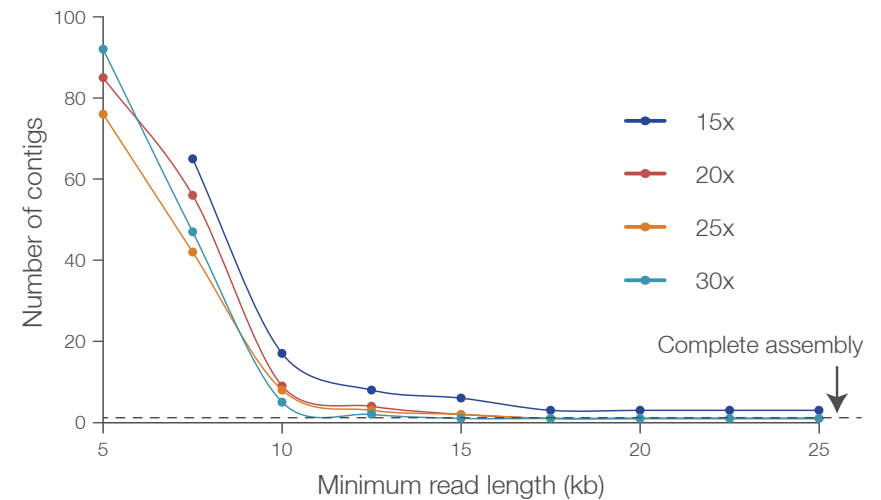
Should I use 16S or WGS?

Although by definition metagenomics involves whole genome analyses, 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 extent of data required for subsequent 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 may be the case in 41% of soil microbial genomes; therefore, the degree of success of this approach may depend on sample origin[3].

For unbiased detection, rare pathogen identification, species-level resolution, and AMR and variant analyses, WGS is recommended. For identifying organisms of interest, 16S may be sufficient, although the extent of species-level resolution is highly limited.

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.

References

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 (2019).

3. Brewer, T. E. *et al*. Unlinked rRNA genes are widespread among bacteria and archaea. The ISME Journal. 14:597–608 (2019).

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

	16S	WGS
Taxonomic resolution	Phyla > 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 PCR), 24 (without PCR)
Sequence data needed	Mb → Gb	Mb → 10s of Gb

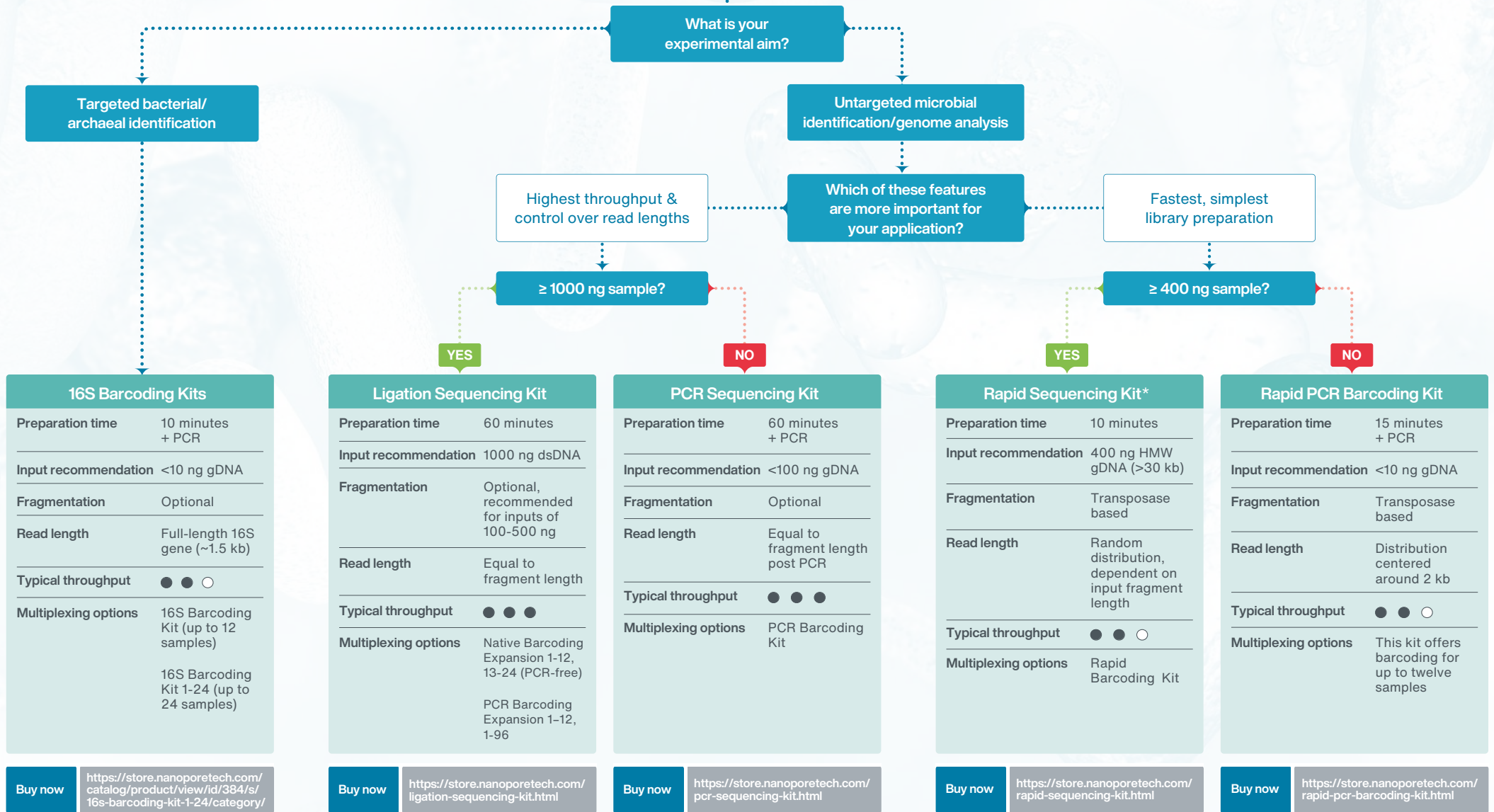* Genus and even species level may be possible in a limited number of cases.
† Application dependent e.g. identification of prevalent organism vs. *de novo* assembly of genomes present.

Which kit do I choose?

Start here

What is your experimental aim?

→ **Targeted bacterial/archaeal identification**

→ **Untargeted microbial identification/genome analysis**

Which of these features are more important for your application?

→ **Highest throughput & control over read lengths**

→ **Fastest, simplest library preparation**

≥ 1000 ng sample?

≥ 400 ng sample?

YES → Ligation Sequencing Kit
NO → PCR Sequencing Kit
YES → Rapid Sequencing Kit*
NO → Rapid PCR Barcoding Kit

16S Barcoding Kits

Preparation time	10 minutes + PCR
Input recommendation	<10 ng gDNA
Fragmentation	Optional
Read length	Full-length 16S gene (~1.5 kb)
Typical throughput	● ● ○
Multiplexing options	16S Barcoding Kit (up to 12 samples) 16S Barcoding Kit 1-24 (up to 24 samples)

Buy now — https://store.nanoporetech.com/catalog/product/view/id/384/s/16s-barcoding-kit-1-24/category/

Ligation Sequencing Kit

Preparation time	60 minutes
Input recommendation	1000 ng dsDNA
Fragmentation	Optional, recommended for inputs of 100-500 ng
Read length	Equal to fragment length
Typical throughput	● ● ●
Multiplexing options	Native Barcoding Expansion 1-12, 13-24 (PCR-free) PCR Barcoding Expansion 1–12, 1-96

Buy now — https://store.nanoporetech.com/ligation-sequencing-kit.html

PCR Sequencing Kit

Preparation time	60 minutes + PCR
Input recommendation	<100 ng gDNA
Fragmentation	Optional
Read length	Equal to fragment length post PCR
Typical throughput	● ● ●
Multiplexing options	PCR Barcoding Kit

Buy now — https://store.nanoporetech.com/pcr-sequencing-kit.html

Rapid Sequencing Kit*

Preparation time	10 minutes
Input recommendation	400 ng HMW gDNA (>30 kb)
Fragmentation	Transposase based
Read length	Random distribution, dependent on input fragment length
Typical throughput	● ● ○
Multiplexing options	Rapid Barcoding Kit

Buy now — https://store.nanoporetech.com/rapid-sequencing-kit.html

Rapid PCR Barcoding Kit

Preparation time	15 minutes + PCR
Input recommendation	<10 ng gDNA
Fragmentation	Transposase based
Read length	Distribution centered around 2 kb
Typical throughput	● ● ○
Multiplexing options	This kit offers barcoding for up to twelve samples

Buy now — https://store.nanoporetech.com/rapid-pcr-barcoding-kit.html

*The Field Sequencing Kit (SQK-LRK001), a lyophilised version of the Rapid Sequencing Kit, is also available; the kit does not require a cold chain, and can be stored at up to 30ºC for 1 month or 2-8ºC for 3 months (unopened), for ultimate portability.

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 both checks that everything is working as it should and helps to familiarise users with our library prep and sequencing workflow. Our Getting Started Guides take you through this entire process, with easy-to-follow instructions for every step of the way.

Step by step guides

These guides provide detailed instructions for running a control experiment on a MinION.

Choose a guide

View our Getting Started Guides here

https://community.nanoporetech.com/getting_started

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 whole genome sequencing protocol

https://community.nanoporetech.com/knowledge/protocol_builder

BIOLOGICAL SAMPLE

How do I design my protocol?

EXTRACTION

How can I best extract high molecular weight DNA from my sample?

Our Extraction Methods resource page features recommended DNA extraction methods and comparisons for a wide range of sample types, such as stool, soil, and a variety of plant and animal tissues, 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, with info sheets for three methods: SPRI selection for DNA fragments >1.5-2 kb; the Circulomics Short Read Eliminator (SRE) Kit for fragments >10 kb; and the BluePippin instrument, capable of enriching for molecules >40 kb.
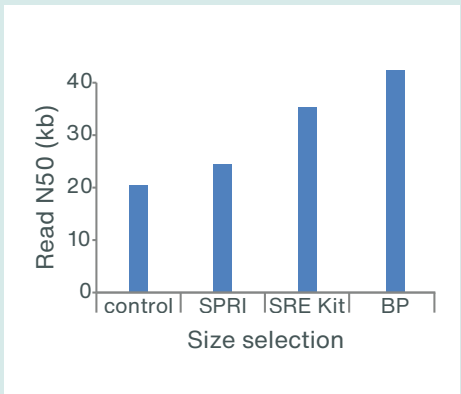
Read more about recommended extraction methods for your sample

https://community.nanoporetech.com/extraction_methods
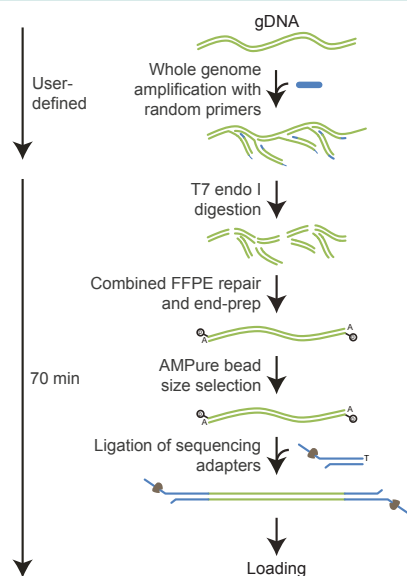
Our kit selection guide (see page 5) displays sample input recommendations, with both PCR and PCR-free options depending on how much starting material is available. When starting with very low quantities of DNA, whole genome amplification can be performed to generate sufficient sample for sequencing:

- **Premium whole genome amplification:** incorporates the Qiagen REPLI-g Midi Kit into the Ligation Sequencing protocol, enabling users to start from just 10 pg gDNA.

- **Rapid whole genome amplification:** uses the Qiagen REPLI-g UltraFast Mini Kit to amplify from >1 ng starting DNA, or directly from cells, in under 2 hours; the library is then prepared in 10 minutes with the Rapid Sequencing Kit.

Find our PCR-free, PCR and WGA protocols in the Protocol Library

https://community.nanoporetech.com/protocols

DNA SAMPLE

How much DNA do I need?

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).

EXTRACTION

Should I perform host depletion?

PUBLICATION

Optimizing DNA extraction methods for Nanopore sequencing of Neisseria gonorrhoeae direct from urine samples

18 December 2019
Journal of Clinical Microbiology

PUBLICATION

Metagenomic identification of severe pneumonia pathogens in mechanically-ventilated patients: a feasibility and clinical validity study

27 November 2019
Respiratory Research

PUBLICATION

Nanopore metagenomics enables rapid clinical diagnosis of bacterial lower respiratory infection

24 June 2019
Nature Biotechnology

Find more depletion methods used by the Nanopore Community in recent publications

https://nanoporetech.com/resource-centre

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

Ligation:

PCR-free:	Ligation Sequencing Kit + Native Barcoding Expansion Packs (12 or 24 samples)
PCR:	PCR Barcoding Kit (12 samples); Ligation Sequencing Kit + PCR Barcoding Expansion Pack (12 or 96 samples)

Rapid:

PCR-free:	Rapid Barcoding Kit (12 samples)
PCR:	Rapid PCR Barcoding Kit (12 samples)

For more information on the methods of multiplexing available, watch our Barcoding Knowledge Exchange webinar.

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

View the Barcoding Knowledge Exchange webinar here

https://community.nanoporetech.com/posts/3598

LIBRARY PREPARATION

How do I multiplex my samples?

SEQUENCING

Which device do I need?

Throughput per device over 48 hours

Gigabases in 48 hours[‡] (log_{10})
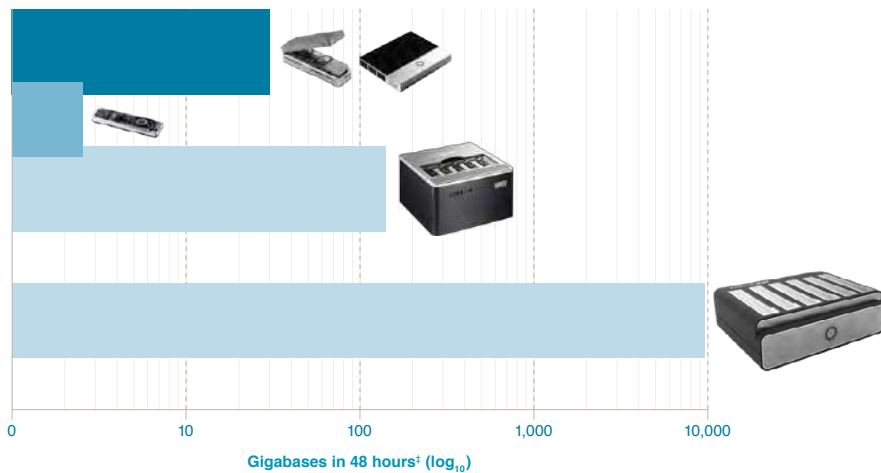
The chemistry underpinning nanopore sequencing can be scaled up or down; our devices make full use of this. The miniature, portable Flongle and MinION Mk1B are ideal for sequencing at sample source, and the MinION Mk1C is an all-in-one device providing both sequencing and analysis. The mid-throughput GridION (5 x MinION or Flongle Flow Cells), and high-throughput PromethION platforms offer flexible, on-demand benchtop sequencing. To help choose a sequencing device that fits your application, visit the product comparison page.

Full product details may be found here

https://nanoporetech.com/products/comparison

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
- AMR gene analysis: 20x
- Variant calling: 100x
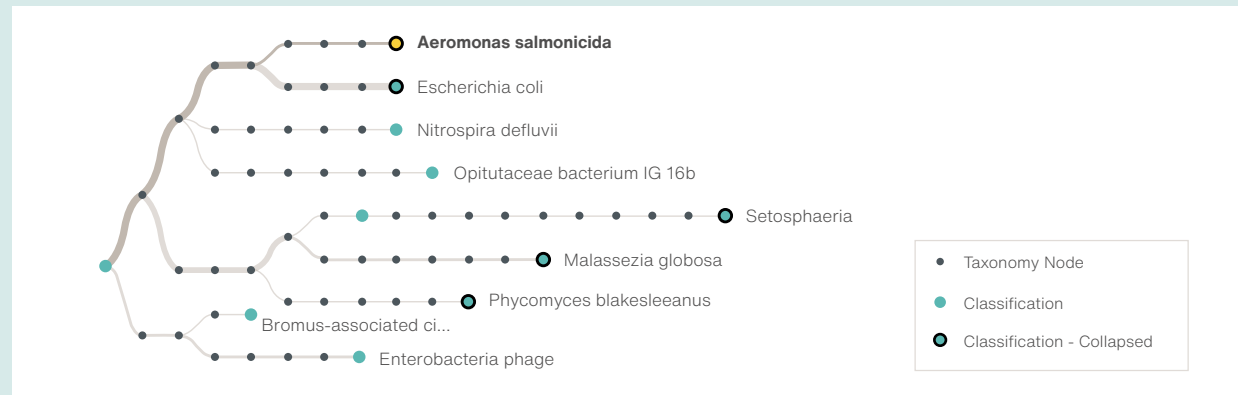- Species-level ID: 20x
- Assembly: 30x

Microbial identification with 16S

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.

Microbial identification with WGS

For 30x depth of a 3 Mb genome (assuming an aim of genome assembly), 90 Mb of sequencing data is needed. From 1 Gb of data, organisms with ≥9% abundance can therefore be identified. Sensitivity can be increased or decreased by increasing/decreasing the amount of data obtained, respectively. Sample multiplexing decreases sensitivity but increases cost effectiveness.

Cumulative Output: Species-level taxonomic classification

- Aeromonas salmonicida
- Escherichia coli
- Nitrospira defluvii
- Opitutaceae bacterium IG 16b
- Setosphaeria
- Malassezia globosa
- Phycomyces blakesleeanus
- Bromus-associated ci...
- Enterobacteria phage

- Taxonomy Node
- Classification
- Classification - Collapsed

Find out more about using nanopore sequencing for metagenomics analysis

https://nanoporetech.com/applications/metagenomics

SEQUENCING

How much sequencing data do I need?

DATA ANALYSIS

How can I begin analysing my data?

Table 3. EPI2ME tools for metagenomics 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 full, real-time antibiotic resistance profiling	Report detailing the resistance genes found and gene overviews

Find out more about nanopore sequencing analysis

https://nanoporetech.com/nanopore-sequencing-data-analysis

The cloud-based analysis platform EPI2ME offers three workflows tailored to metagenomic analysis **(Table 3)**. EPI2ME provides a user-friendly interface; experience with the command line is not needed.

Third party tools can be found in our Resource Centre. We recommend Flye as a robust assembly tool, and Canu as an alternative. Classification tools include MEGAN and Centrifuge.

Case studies

Real-time analysis of pre-term infant gut microbiota

Necrotising enterocolitis (NEC) is one of the most devastating gastrointestinal diseases of preterm and low-birthweight babies, and microbiota dysbiosis has been identified as the most likely cause[4]. Leggett *et al*. performed metagenomic nanopore sequencing of faecal samples from healthy infants and those with NEC or suspected sepsis, to identify the pathogenic organisms present and their AMR profiles. By coupling real-time sequencing on the MinION with their NanoOK RT analysis software, the researchers captured the species diversity of the preterm gut microbiome over time and in response to antibiotics and probiotic supplementation.

The team were able to identify pathogenic bacteria and their resistance profiles in just 1 hour of sequencing (5 hours from sample acquisition). By preparing libraries with the Rapid Sequencing Kit (which requires only 10 minutes of preparation time), they then demonstrated how they could further reduce their turnaround time by an hour.

Sequencing on either the MinION or GridION platforms using Flongle Flow Cells also increased the cost efficiency of their workflow.

Read more here

https://nanoporetech.com/resource-centre/rapid-minion-profiling-preterm-microbiota-and-antimicrobial-resistant-pathogens

Assembly-free nanopore sequencing analysis of viral metagenomes

To gain a greater understanding of the diversity and nature of ocean viruses, DeLong and colleagues investigated the genomes of ocean phages. 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 platform 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.

Read more here

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

Reference

4. Denning, N-L. and Prince, J. M. Neonatal intestinal dysbiosis in necrotizing enterocolitis. Molecular Medicine. 24:4 (2018).

Oxford **NANOPORE**
Technologies

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

www.nanoporetech.com