Oxford NANOPORE Technologies

Whole-genome sequencing: small genomes

GETTING STARTED

nanoporetech.com

Introduction

Microorganisms are the most abundant life forms on Earth, with an estimated 1 trillion species[1]. They are fundamental to the maintenance of all ecosystems and play crucial roles in health, disease, and many environmental and industrial processes.

Full characterisation of microbial genomes is advancing numerous fields of research. In health, sequencing of the human microbiome is shedding light on its role in an ever-growing list of diseases, whilst sequencing of the pathogens behind outbreaks of diseases such as influenza and COVID-19 improving our understanding of their origins and transmission. In agriculture, timely detection of animal and plant pathogens is of great importance to animal welfare and food production.

Despite their abundance and significance, only 0.1% of the 10 million microbial species so far catalogued have been successfully cultured in the laboratory[1], and even with the significant advances to small genome sequencing enabled by short-read sequencing, around 90% of bacterial genomes remain incomplete[2].

Long nanopore sequencing reads provide considerable advantages in small genome sequencing and assembly **(Table 1)**: fewer, longer fragments, with greater overlap, enable easier genome assembly — much like building a jigsaw **(Figure 1)**.

Reads exceeding 4 Mb have been generated with nanopore technology[3], meaning entire viral, and even bacterial, genomes can be obtained in a single read, completely bypassing the need for assembly.

Table 1. Advantages of nanopore technology for whole-genome sequencing

Easier assembly
Long sequencing reads mean fewer fragments to assemble

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

Phasing
Long reads enhance unambiguous allele phasing

Real-time monitoring
Reads can be basecalled and analysed as sequencing progresses; runs can be stopped once a coverage target is reached

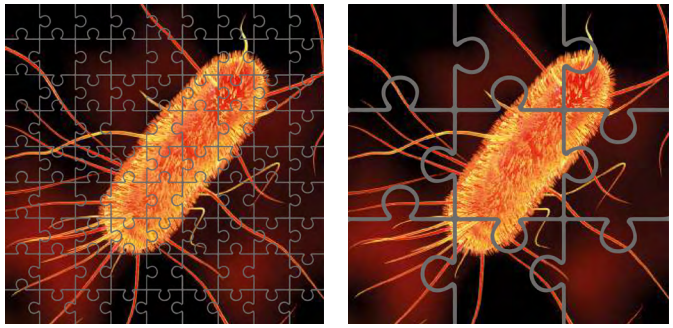
Rapid time to result
Real-time analysis enables the characterisation of disease samples or in-field outbreak surveillance and prevention with a quick turnaround time

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

Absence of GC bias
GC bias is virtually absent in native nanopore sequencing data, meaning greater uniformity of coverage compared to assemblies generated from short-read data

Figure 1. Long nanopore reads simplify genome assembly

Like a jigsaw puzzle with large pieces, long DNA reads are much easier to assemble than short DNA reads. The *Escherichia coli* genome comprises 4.6 million bases, which would equate to 92,000 fragments of 50 bp or just 9 fragments of 500,000 bp in length.



50 bp reads
92,000 “pieces”

500,000 bp reads
9 “pieces”

1. Locey, K.J. and Lennon, J.T. Scaling laws predict global microbial diversity. Proc Natl Acad Sci USA. 113(21):5970-5 (2016).
2. Land, M. et al. Insights from 20 years of bacterial genome sequencing. Funct Integr Genomics. 15(2): 41–161 (2015).
3. Oxford Nanopore Technologies (2021). Ultra-Long DNA Sequencing Kit. Available at: https://store.nanoporetech.com/productDetail/?id=ultra-long-dna-sequencing-kit [Accessed: 20 July 2021]

Furthermore, the longer the reads, the lower the depth of coverage required to achieve a complete, single-contig assembly **(Figure 2)**. Long reads can also span structural variants, which greatly enhances their resolution; structural variants are associated with important functions in microorganisms, such as the development of antimicrobial resistance and virulence factors. Similarly, long reads are better able to span repeats, simplifying the resolution of their location and structure. Repeats are characteristic of plasmids, for example, which can be sequenced in single, long nanopore reads.

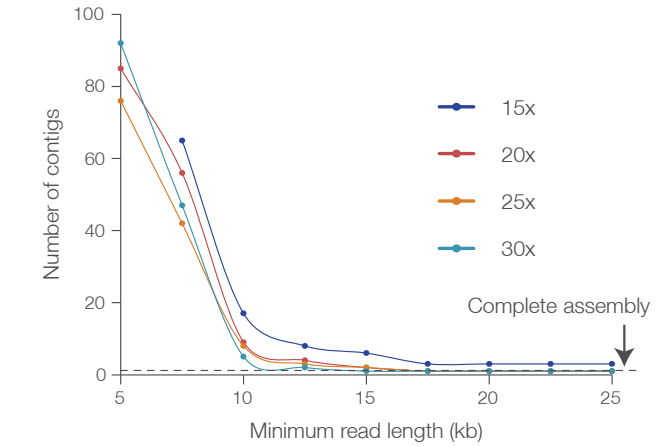


Figure 2. Longer reads facilitate assembly completion from lower-coverage data

RNA viral genome assembly

Many viral diseases, including Ebola, Zika, influenza, and COVID-19, are caused by RNA viruses. Oxford Nanopore provides options for both direct RNA sequencing, in which native RNA can be sequenced with modifications preserved, and cDNA sequencing, with PCR-based and PCR-free methods available. This guide focuses on whole-genome DNA sequencing; for more information on direct RNA and cDNA sequencing for RNA genome assembly, please visit **nanoporetech.com/applications/rna-sequencing**

Nanopore sequencing devices

Nanopore sequencing devices are scalable to your needs, offering sequencing solutions for genomes of every size. The MinION™ outputs 10s of Gb of data per run, and with its portable design it can be used anywhere, from the lab bench to a field setting. The Flongle™ is a 'flow cell dongle': an adapter which converts MinION and GridION devices for use with small, cost-effective Flongle Flow Cells. With a 1-2 Gb output, Flongle Flow Cells are perfect for sequencing small genomes when the full capacity of a MinION Flow Cell is not needed.

Figure 3 provides a guide to the size of genomes that can be sequenced to high depth of coverage on each device — from viruses and bacteria on the cost-competitive, single-use Flongle, up to the larger animal and plant genomes on the ultra-high-throughput PromethION™. However, these devices are also highly flexible: a larger genome can be split and sequenced as needed across several MinION Flow Cells if required, whilst barcoding allows many samples to be sequenced in multiplex on one flow cell, making the most of its sequencing capacity. The modular structure of the GridION and PromethION enables the use of flow cells independently or concurrently, providing efficient, on-demand sequencing. With streamlined workflows and real-time analysis, data can be rapidly obtained and even analysed as soon as sequencing starts.

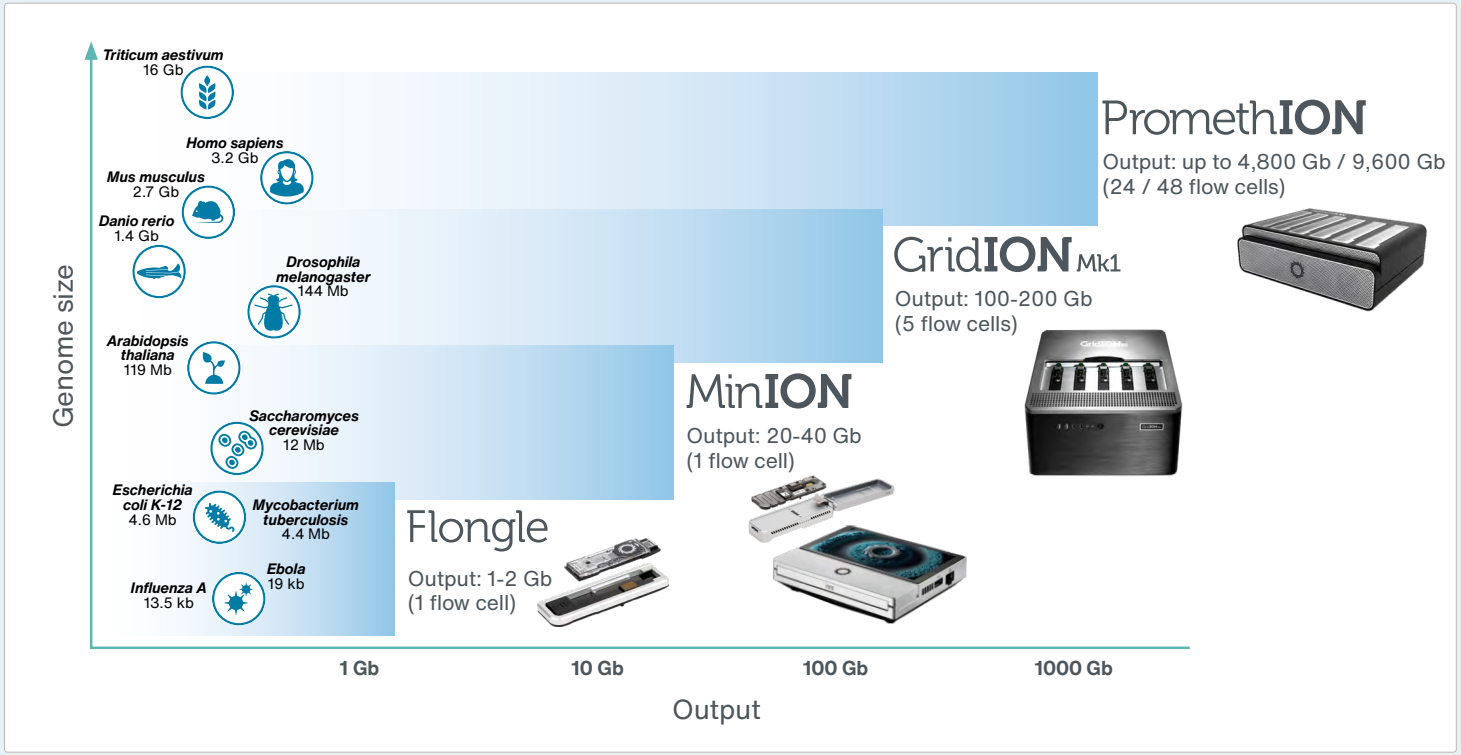
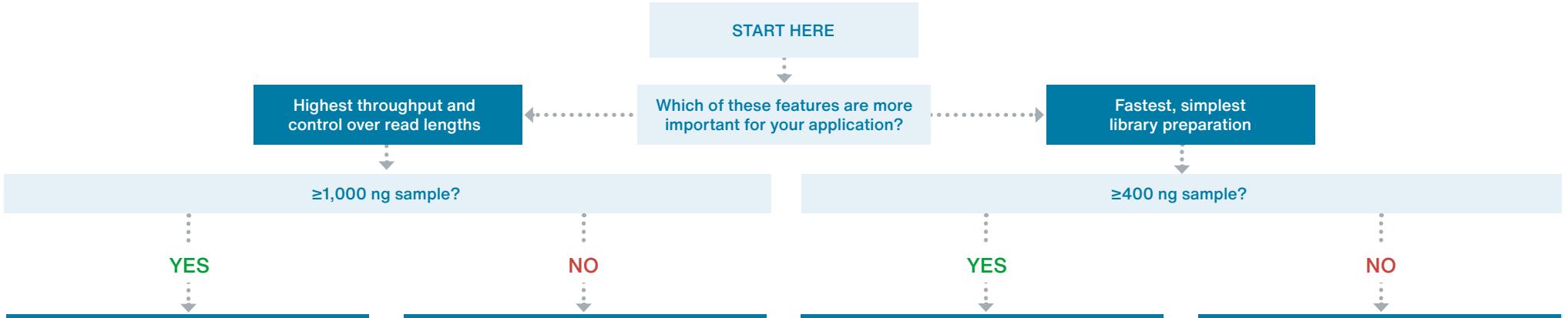


Figure 3. Genome size versus device and output

Which kit do I choose?



Ligation Sequencing Kit	
Preparation time	60 minutes
Input recommendation	<1,000 ng dsDNA
Fragmentation	Optional, recommended for inputs of 100–500 ng
Read length	Equal to fragment length
Typical throughput	● ● ●
BUY NOW: 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	● ● ○
BUY NOW: 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	● ● ○
BUY NOW: 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	~2 kb
Typical throughput	● ● ○
BUY NOW: 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 30°C for 1 month or 2–8°C for 3 months (unopened), for ultimate portability.

From sample to answer

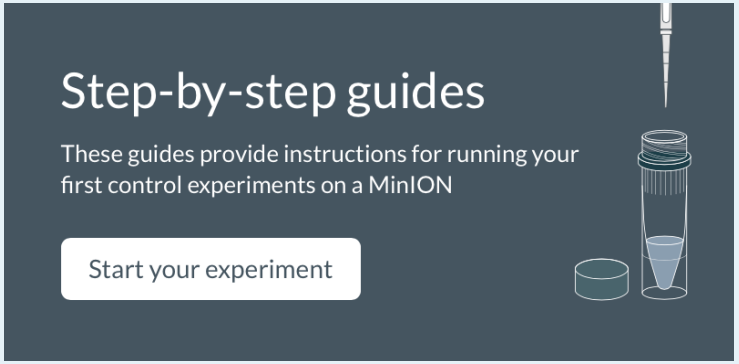
QUESTION

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 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

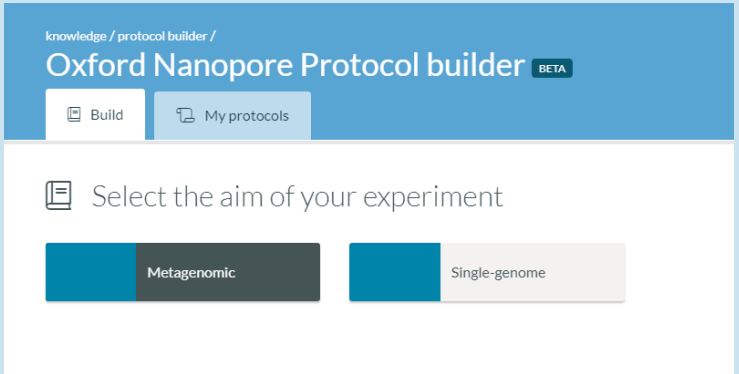


PLANNING

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 whole-genome sequencing protocol:
community.nanoporetech.com/knowledge/protocol_builder



From sample to answer

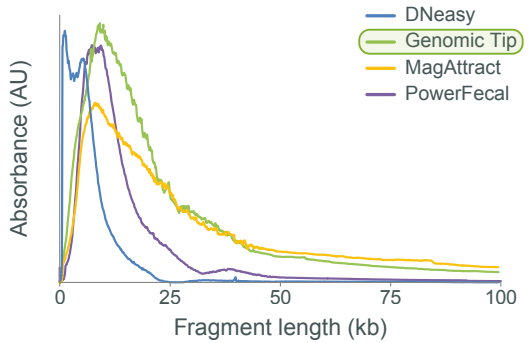
EXTRACTION

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 bacteria, 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; if performing genome assembly, we would recommend size selecting for long reads 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?

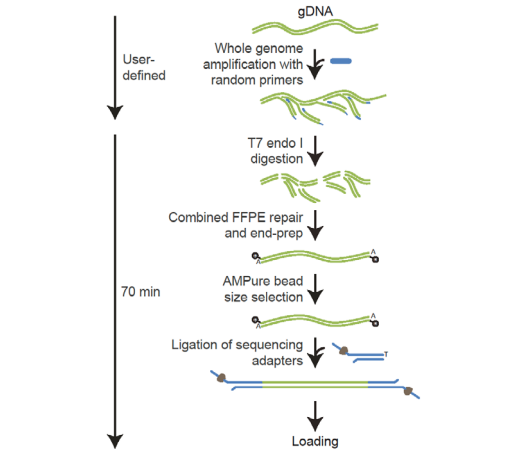
Our kit selection workflow (see page 5) displays sample input recommendations, with both PCR and PCR-free options depending on how much starting material is available. When starting with very low quantities of DNA, whole-genome amplification (WGA) can be performed to generate sufficient sample for sequencing:

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.

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

Premium whole genome amplification workflow



From sample to answer

LIBRARY PREP

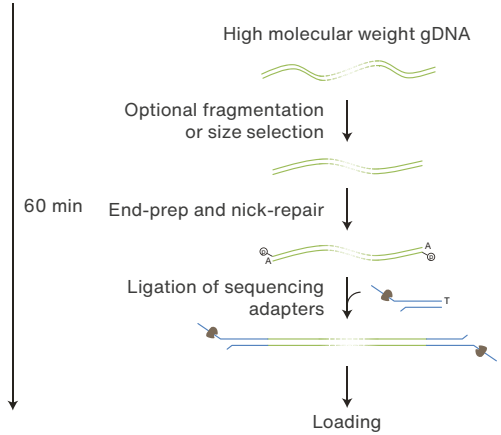
How do rapid kits differ from ligation-based kits?

Ligation: the Ligation Sequencing Kit is optimised for high throughput and gives users control over read length. Fragmentation is optional and long-range PCR is possible. In these protocols, adapters are added using a ligase.

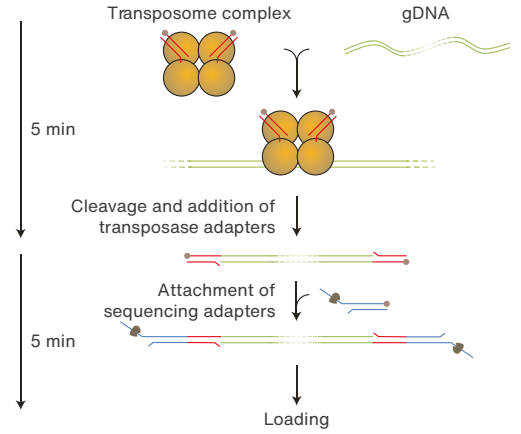
Rapid: our rapid sequencing kits are designed to enable the fastest, simplest library prep possible. A transposome complex cuts the DNA and adds transposase adapters in a single step. Rapid sequencing adapters are then attached without the need for a ligase. Due to the cuts introduced by the transposome complex, read lengths are a random distribution, though due to the minimal pipetting involved, ultra-long reads may also be generated.

Browse our selection of library prep kits :
store.nanoporetech.com/sample-prep.html

Ligation Sequencing Kit workflow



Rapid Sequencing kits workflow



From sample to answer

LIBRARY PREP

Can I multiplex my samples?

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, 24, or 96 samples)
PCR:	PCR Barcoding Kit (12 samples); Ligation Sequencing Kit + PCR Barcoding Expansion Packs (12 or 96 samples)

Rapid:

PCR-free:	Rapid Barcoding Kit (12 or 96 samples)
PCR:	Rapid PCR Barcoding Kit (12 samples)

View Knowledge Exchanges on nanopore sequencing, from sample prep to analysis:
community.nanoporetech.com/knowledge_exchange

SEQUENCING

Can I re-use my flow cell?

Yes — the Flow Cell Wash Kit can be used to remove >99.9% of a sequenced library, leaving the flow cell ready either to sequence a fresh library or to be prepared for storage until it is needed again. This is especially useful when sufficient depth of coverage of a genome is obtained early on in a sequencing run: the run can be stopped when a coverage target is reached, then washed and used again for a new sample. By barcoding each library, any residual carried-over reads can then be filtered out in analysis.

View the Flow Cell Wash Kit:
store.nanoporetech.com/productDetail/?id=flow-cell-wash-kit-r9

From sample to answer

DATA ANALYSIS

How can I analyse my data?

A range of solutions is available for whole-genome assembly using nanopore sequencing data, suitable for all levels of bioinformatics expertise.

We recommend the third-party command-line tool Flye as a robust and reliable workflow for small genome assembly. This tool can be found on GitHub and is also packaged into a complete small genome assembly analysis pipeline available in EPI2ME Labs. The EPI2ME Labs resource provides best practice workflows and interactive tutorials to support the analysis of nanopore sequencing data and the development of bioinformatics skills. Find out more at **labs.epi2me.io**.

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

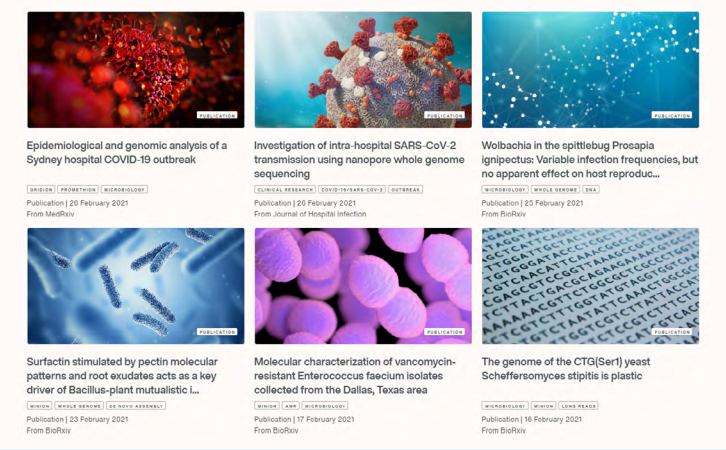
DISCUSSION

Where can I discuss whole-genome sequencing with other Oxford Nanopore users?

The Nanopore Community is a thriving online hub, helping users to get started, share their work and experiences, and collaborate. Members of the Nanopore Community are continually developing new protocols and tools for extraction, library prep, and analysis, for an increasingly diverse range of applications. You can also find the latest publications featuring whole-genome nanopore sequencing in the 'Resources' section of our website.

Join the Community discussion and ask the experts here:
community.nanoporetech.com

Find out how nanopore technology is being used for whole-genome sequencing:
nanoporetech.com/applications/whole-genome-sequencing



Case studies

Case study 1: Rapid identification of pathogens, resistance genes, and plasmids in blood cultures

Bloodstream infections (BSI) and associated sepsis represent significant causes of mortality worldwide; the incidence of BSI has increased over the past decade, partly owing to the increased prevalence of antimicrobial resistance (AMR)[4]. Moreover, current methods for detecting BSIs based on the phenotypic determination of antibiotic susceptibility have turnaround times of up to three days. With the increasing burden of AMR, rapid detection of pathogens, and antimicrobial susceptibility, has become paramount in securing the most appropriate antibiotic therapy for BSIs. Furthermore, determining the exact location of an AMR gene is vital for a full risk assessment of AMR transmission. Locating AMR genes is more readily achieved using long sequencing reads than short sequencing reads due to the superior mappability of reads[5]. Real-time, .whole-genome sequencing, with a provision of long reads, could represent a means of enabling the rapid and accurate identification of both the pathogens and the AMR genes involved.

To that end, in a proof-of-concept study, Arne Taxt (Oslo University Hospital) used real-time nanopore sequencing of bacterial DNA extracted from positive blood cultures, with an aim of identifying pathogens, plasmids, and AMR-encoding genes within hours. Taxt and the team observed that *'identification of pathogens was possible after 10 minutes of sequencing'*, and that the detection of predefined AMR-encoding genes and plasmids could be achieved within hours[6]. They suggested that such an approach could identify pathogens and specific AMR-encoding genes within four hours[6], after a blood culture is initially flagged as positive. This approach therefore demonstrates the potential to rapidly detect clinically-relevant resistance genes — which could contribute to early, optimised intervention, as well as reduce the use of broad-spectrum therapy.

Read the publication:
nanoporetech.com/resource-centre/rapid-identification-pathogens-antibiotic-resistance-genes-and-plasmids-blood

Case study 2: *De novo* assembly of a plant pathogen genome using long nanopore reads

The genus *Phytophthora*, meaning 'plant destroyer' in Greek, is the genus name aptly given to the most destructive plant pathogen of vegetable crops, *Phytophthora capsici*. This soil-borne oomycete causes blight and fruit rot in many commercial crops, such as bell peppers and beans[7], and is responsible for sizeable economic losses across the globe. Therefore, methods for better controlling *P. capsici* remain pressing, and the development of such methods requires an improved understanding of the molecular mechanisms underlying the pathogenicity of *P. capsici*. While several iterations of the *P. capsici* genome sequence exist[8], obtaining a high-quality reference genome has remained problematic[9]. For example, many virulence genes are embedded within repetitive regions of the genome[7] which are challenging to characterise using traditional, short-read sequencing technologies.

Shi *et al*. used long nanopore reads to *de novo* assemble the genome of *P. capsici*[9]. The genome was sequenced to 116x depth of coverage and *de novo* assembled using Canu, with subsequent assembly polishing, producing a final assembly of 100.5 Mb — approximately 5.6% longer than the most recent *P. capsici* assembly. The number of contigs was also reduced by 45%, and the scaffold N50 was improved by 3.2-fold. Moreover, preliminary gene synteny analysis revealed a more continuous genomic sequence than previous *P. capsici* assemblies, most likely due to the improved capability of long nanopore reads to assemble highly repetitive regions. This improved assembly will provide an invaluable resource for understanding the disease mechanisms and adaptability of *P. capsici*, and ultimately for developing effective management strategies[9].

Read the publication:
nanoporetech.com/resource-centre/improved-whole-genome-sequence-phytophthora-capsici-generated-long-read-sequencing

4. Goto, M. and Al-Hasan, M. Overall burden of bloodstream infection and nosocomial bloodstream infection in North America and Europe. Clinical Microbiology and Infection. 19(6): 501-509 (2013).
5. Berbers, B. et al. Combining short and long-read sequencing to characterize antimicrobial resistance genes on plasmids applied to an unauthorized genetically modified Bacillus. Sci Rep. 10:4310 (2020).
6. Taxt, A., Avershina, E., Frye, S., Naseer, U., and Ahmad, R. Rapid identification of pathogens, antibiotic resistance genes and plasmids in blood cultures by nanopore sequencing. Scientific Reports. 10:7622 (2020).
7. Cui, C., Herlihy, J., Bombarely, A., McDowell, J., and Haak, D. Draft assembly of Phytophthora capsici from long-read sequencing uncovers complexity. MPMI. 32(12):1559-1563 (2019).
8. Wang, Y., Tyler, B., and Wang, Y. Defense and counterdefense during plant-pathogenic oomycete infection. Annual Review of Microbiology. 73(1):667-696 (2019).
9. Shi, J. et al. Improved whole genome sequence of Phytophthora capsici generated by long-read sequencing. MPMI. DOI: https://doi.org/10.1094/MPMI-12-20-0356-A (2021).

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

www.nanoporetech.com

GS_1027(EN)_V2_20Aug2021