Assembling microbial genomes using long nanopore sequencing reads

In order to understand the true diversity and functions of microorganisms, producing fully annotated, complete genomes is essential. However, 90% of bacterial genomes are predicted to be incomplete[1].

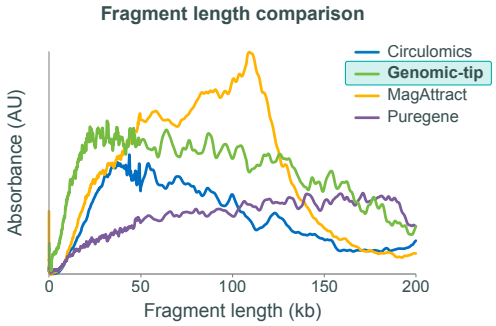
Compared to short-read sequencing data, long nanopore sequencing reads simplify genome assembly, with the ability to span repeat-rich sequences (characteristic of antimicrobial resistance genes) and structural variants. Nanopore sequencing also shows a lack of bias in GC-rich regions, in contrast to other sequencing platforms[2], further supporting the assembly of complete genomes.

Here we present a simple workflow for bacterial genome assembly from a single-organism culture, using MinION™ Flow Cells on MinION or GridION™ sequencing devices.

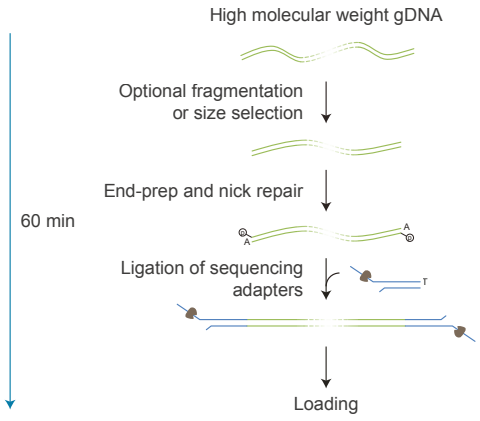
EXTRACTION:
obtaining high molecular weight DNA

To extract DNA from a bacterial isolate, we recommend the **QIAGEN Genomic-tip 500/G**, which we have found to produce the longest read lengths and highest yield. Performing size selection on the extracted DNA further increases read length N50; size selection options include **Agencourt AMPure XP beads** and the **Circulomics Short Read Eliminator Kit**.

Find more extraction protocols, such as those for stool, soil, and culture, as well as alternative size selection methods: **community.nanoporetech.com/extraction_methods**



LIBRARY PREPARATION:
selecting a kit



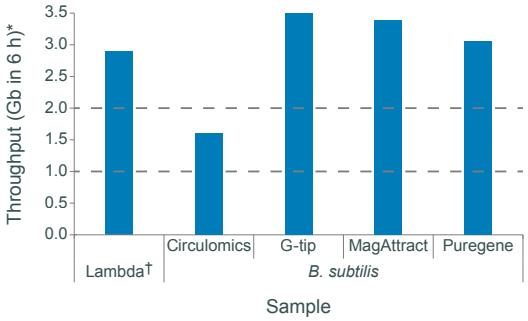
To prepare gDNA for sequencing, we recommend the **Ligation Sequencing Kit**, providing the greatest throughput and control over read lengths.

Find out more about library preparation kits, including rapid, 10-minute options: **nanoporetech.com/products/kits**

Multiplexing options are available to increase the cost efficiency of your sequencing. We recommend the **Native Barcoding Expansion** kits, which are PCR-free, and enable up to 96 samples to be sequenced in multiplex on one flow cell. Alternatively, **PCR Barcoding Expansion** kits are available for PCR-based multiplexing of up to 96 samples.

SEQUENCING: achieving ultimate flexibility with MinION Flow Cells

Find out more about MinION:
nanoporetech.com/products/minion



*Please note, sequencing run time of a MinION Flow Cell is up to 72 h.
†Lambda: This commercially available gDNA sample was run as a control

We recommend sequencing bacterial genomes on MinION Flow Cells. These flow cells can be used individually on the portable MinION Mk1B and Mk1C sequencing platforms; alternatively, the benchtop GridION enables on-demand sequencing of up to five MinION Flow Cells. You can therefore adjust flow cell number according to the degree of multiplexing, number of samples, and/or your experimental goals.

For assembly, we recommend basecalling in high accuracy mode, and sequencing to a minimum depth of 30x of ≥10 kb reads, per genome. We suggest multiplexing 12–24 samples per flow cell, scaled up or down according to your sample (e.g. expected genome length; fragment length distribution) and experimental aims. However, if variant calling is also desired, increasing sequencing depth is advisable. We have found that Q score (accuracy) continues to improve up to 100x.

Find out more and compare nanopore sequencing platforms: nanoporetech.com/products/comparison

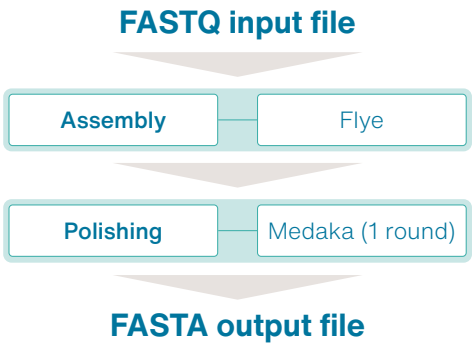
ANALYSIS: selecting an assembly tool

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

To perform microbial genome assembly, we suggest using the third-party *de novo* assembly tool Flye[3]. This analysis package represents a complete pipeline, taking raw nanopore reads as input, and producing polished contigs as output. We also recommend one round of polishing with Medaka[4]. These tools can be found on GitHub.

Microbial genomes up to ~10 Mb can be assembled with Flye using a standard laptop (~16 GB memory).

This complete analysis pipeline is also available in EPI2ME Labs, which provides best practice workflows and interactive tutorials to support the analysis of your nanopore sequencing data and develop your bioinformatics skills. Find out more at **labs.epi2me.io.**



Find out more at: **nanoporetech.com/investigating/genome-assembly**

Oxford NANOPORE Technologies

Twitter: @nanopore
www.nanoporetech.com

References:
1. Land, M. *et al.* Insights from 20 years of bacterial genome sequencing. *Funct. Integr. Genomics.* 15(2): 141–161 (2015).
2. Browne, P. D. *et al.* GC bias affects genomic and metagenomic reconstructions, underrepresenting GC-poor organisms. *GigaScience*, 9(2): giaa008 (2020).
3. Kolmogorov, M. *et al.* Assembly of long, error-prone reads using repeat graphs. *Nat. Biotech.* 37: 540-546 (2019).
4. Oxford Nanopore Technologies. Medaka. Software. Available at: https://nanoporetech.github.io/medaka [Accessed: 01 August 2021]

WF_1067(EN)_V2_01Aug2021.