Oxford NANOPORE Technologies

WORKFLOW: METAGENOMIC ASSEMBLY

Assembling microbial genomes from complex metagenomic samples using long nanopore reads

Covering a vast array of applications, metagenomic sequencing allows the rapid identification and analysis of culturable microorganisms, and, importantly, has made possible the analysis of those microbes which cannot be cultured.

Compared to short-read sequencing approaches, long nanopore reads can span repeat-rich sequences — characteristic of antimicrobial resistance (AMR) genes — and structural variants. This increases the resolution of microbial classification and simplifies genome assembly, thereby providing deeper insight into the composition and function of microbial communities. With sequencing data available in real time, it is also possible to rapidly identify the species, AMR and virulence factors in a mixed microbial community.

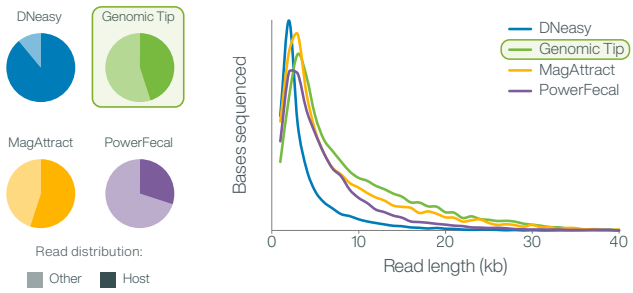
Here we present a simple workflow for assembling microbial genomes from metagenomic samples, using MinION™ Flow Cells on MinION or GridION™ sequencing devices or PromethION™ Flow Cells on the PromethION 2 or PromethION 2 Solo platforms.

EXTRACTION: obtaining high-quality DNA

Find our extraction protocols, as well as alternative size selection methods: **community.nanoporetech.com/docs/prepare**

Obtaining long fragments of high-quality genomic DNA is ideal for genome assembly. Selecting a suitable extraction method greatly depends on sample type; Oxford Nanopore provides a range of sample-specific extraction protocols, such as those soil, stool, and a variety of plant, human, and animal tissues.

For example, we recommend the **QIAGEN Genomic-tip 20/G** to extract DNA from stool, which we have found to produce the longest read lengths and a balanced distribution of reads between the host and microbiome. We recommend using **Agencourt AMPure XP beads** to size select for longer reads prior to sequencing.



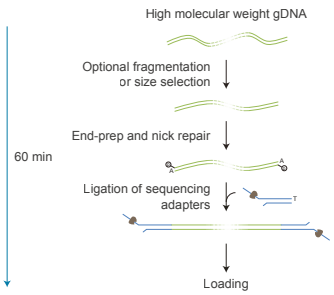
Find more sample prep guidance, such as options for host depletion, in our Metagenomic sequencing Getting started guide: nanoporetech.com/resource-centre/guide/metagenomic-sequencing

LIBRARY PREPARATION: selecting a kit

Find out more about library preparation kits, including rapid and low-input options: **nanoporetech.com/products/kits**

To prepare gDNA for sequencing and downstream metagenomic assembly, we recommend the **Ligation Sequencing Kit**, a PCR-free library prep approach, providing the greatest output and control over read lengths. This workflow is suitable for ≥100 ng input gDNA. If starting from lower quantities of DNA, low-input PCR-based options are available, such as the **Rapid PCR Barcoding Kit**. However, as PCR can introduce bias, we suggest avoiding PCR-based methods if possible.

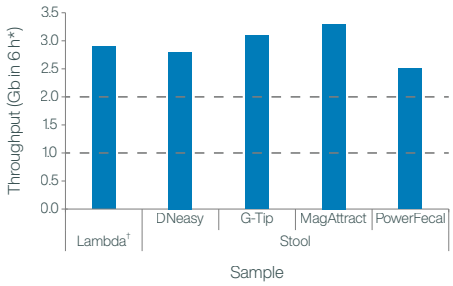
Ligation Sequencing Kit workflow



Sample multiplexing increases the cost efficiency of sequencing, but will also reduce depth of coverage per genome, so rarity of organisms, sample complexity, and presence of host DNA should be considered prior to choosing this approach. We recommend the **Native Barcoding 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: scaling throughput for your experimental goals

Learn more about the PromethION 2 devices:
nanoporetech.com/products/promethion-2



*Note that sequencing run time of a MinION Flow Cell is up to 72 h

†Lambda: This commercially available gDNA sample was run as a control

To assemble microbial genomes from metagenomic samples, we recommend sequencing to a depth of 30x of ≥10 kb reads, per genome. As a starting point, we recommend sequencing one metagenomic library per MinION Flow Cell. Based on a 30 Gb run output, and accounting for moderate host contamination, this should provide sufficient (i.e. ≥30x) depth to assemble ~3 Mb microbial genomes present at ≥1% prevalence. MinION Flow Cells are compatible with the portable MinION Mk1B and MinION Mk1C sequencing devices, ideal for sequencing anywhere, from the bench to the sample source. Alternatively, the benchtop GridION enables on-demand sequencing of up to five MinION Flow Cells; flow cell number can therefore be adjusted according to your experimental goals.

For higher-throughput sequencing of metagenomic libraries in multiplex, or to obtain higher depth of coverage of low-abundance microorganisms, we recommend the powerful and compact PromethION 2 and PromethION 2 Solo, with capacity for sequencing on up to two high-yield PromethION Flow Cells.

If the experimental aim is assembly-free identification, we recommend obtaining a depth of 20x (per organism) for microbial classification at species level.

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

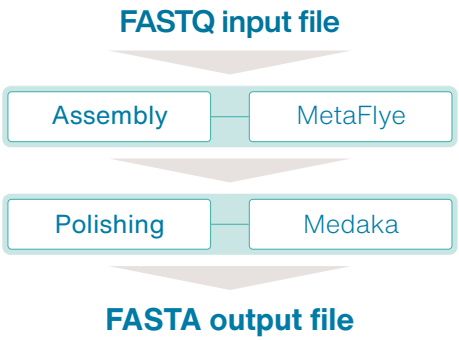
ANALYSIS: selecting an assembly tool

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

To assemble microbial genomes from metagenomic samples, we suggest using the third-party *de novo* assembly tool **MetaFlye**[1]. 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 additional polishing of the assembly with **Medaka**[2]. These tools can be found on GitHub.

Regarding analysis runtime, for 30 Gb of metagenomic sequence data, assembly with MetaFlye would require 18 hours, and polishing with Medaka a further 2–4 hours (based on compute with 8 CPUs); we recommend at least 100 GB RAM is available to run MetaFlye.

For assembly-free identification, the cloud-based EPI2ME™ platform offers a simple point-and-click workflow, 'What's in my pot?', with no command line experience needed. Alternatively, the workflow **wf-metagenomics** allows the flexibility of running through the EPI2ME™ Labs software or on your own laptop. Visit **https://labs.epi2me.io/** to find out more.



Find out more at: **nanoporetech.com/applications/metagenomics**

References:
1. Kolmogorov, M. et al. MetaFlye: scalable long-read metagenome assembly using repeat graphs. Nat Methods. 17: 1103-1110 (2020).
2. GitHub. Medaka. Available at: github.com/nanoporetech/medaka [accessed: 29 September 2022].
3. Dilthey, A. T. et al. Strain-level metagenomic assignment and compositional estimation for long reads with MetaMaps. Nat Commun. 10:3066 (2019).
4. Kim, D. et al. Centrifuge: rapid and sensitive classification of metagenomic sequences. Genome Res. 26:1721-1729 (2016).

WF_1137(EN)_V2_03Oct2022

Oxford NANOPORE Technologies

Twitter: @nanopore
www.nanoporetech.com