

A protocol for sequencing and analysing  
16S-ITS-23S amplicons using Oxford Nanopore's  
Native Barcoding kit to profile prokaryotes on  
species-level in a mixed community [Under  
development]

Christian Krohn, PhD, RMIT University

2024-09-10



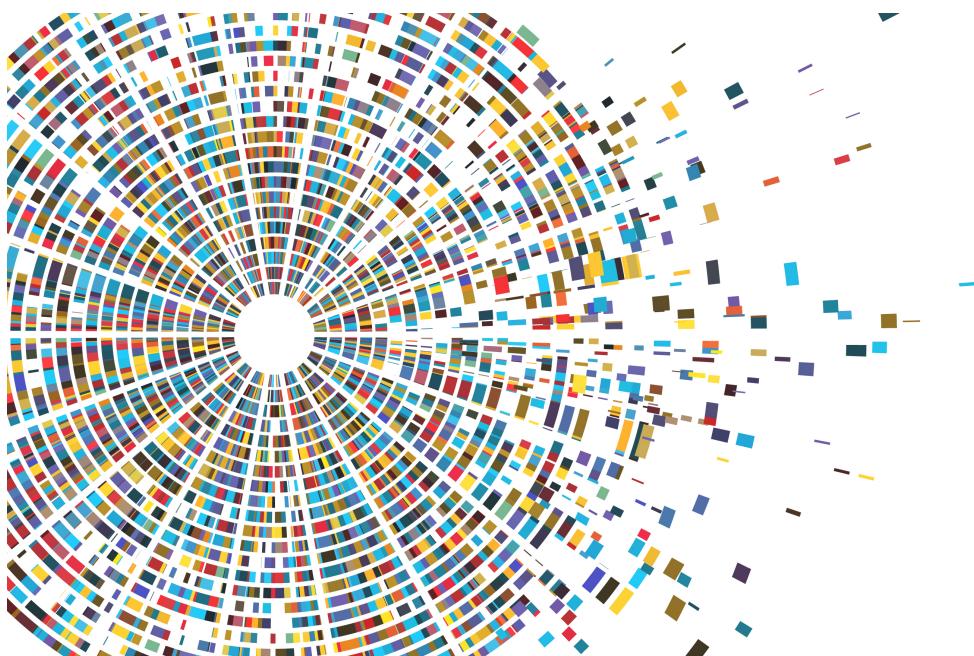
# Contents

<b>1</b>	<b>About this GitBook</b>	<b>5</b>
<b>2</b>	<b>Prerequisites, equipment and consumables</b>	<b>9</b>
2.1	Prerequisites . . . . .	9
2.2	DNA extraction kit . . . . .	9
2.3	Other consumables . . . . .	10
2.4	Protocols . . . . .	11
2.5	Equipment . . . . .	11
2.6	Computational resources and database . . . . .	12
<b>3</b>	<b>Protocol</b>	<b>13</b>
3.1	DNA extraction . . . . .	13
3.2	Amplification of 16S-ITS-23 operon . . . . .	14
3.3	PCR product clean-up . . . . .	16
3.4	Library preparation . . . . .	17
3.5	Priming and loading of flow cell - run start . . . . .	21
3.6	Flow cell wash . . . . .	24
3.7	Basecalling . . . . .	25
3.8	Read classification . . . . .	25
3.9	Results . . . . .	28



# Chapter 1

## About this GitBook



This GitBook is under development (2024).

This GitBook provides a step-by-step protocol for anyone who want to get started with sequencing multiple samples of high quality, full-length microbial amplicons (16S-ITS-23S) with an Oxford Nanopore instrument, and then process and visualise the data using various packages in R. The aim is to profile a community of bacteria and archaea on **species-level**.

Read and use this protocol either directly or to develop your own.

Note that using the 16S-ITS-23S primer pair we use in this workflow does not result in a true representation of microbial abundances. Coverage seems scattered with some key bacteria and archaea underrepresented. Also note that the read classification used (`wf-metagenome` workflow from EPI2ME, with kmer-based read predictions) will provide classified reads of Bacteria and Archaea but, depending on the read quality, will result in some false positives. And again, abundances will not be a true representation as no clustering, dereplication or ASV generation is involved. Results will not be comparable to other studies.

However, this method will provide read classification to species-level, which may be useful in cases when you are not interested in the true representation of relative abundances. For example, if you are after temporal variation, or differences between your samples in your experiment.

This protocol reflects the experiences we had in the lab and may help to get you started on your own journey of discovery. One of the biggest hurdles is the effort required to learn new technologies, hence this protocol aims at lowering your ‘activation’ energy by providing some guidance for each step from DNA extraction, to amplification, library preparation, sequencing, data processing and finally some basic visualisations in R.

This protocol uses the DNeasy Powersoil Pro Kit (Qiagen, Hilden, Germany) to extract DNA from wastewater sludge, as well as the Native Barcoding Kit (SQK-NBD114.96) with a R10.4.1 flowcell (FLO-MIN114) to sequence long (~4-4.4 kb) amplicons from environmental DNA. The primer used was developed by (Martijn et al., 2019). Initial costs to purchase all consumables will be ~AUD\$7,700.

Check out our workflows for short-read 16S sequencing at <https://chrismitbiz.gitub.io/ABlab-workflows> if you want to get started with Miseq 16S sequencing instead.

If you are interested in sequencing the living biomass via PMA treatment in combination with short-read 16S sequencing, check out our recent paper Dead in the water – Role of relic DNA and primer choice for targeted sequencing surveys of anaerobic sewage sludge intended for biological monitoring (Krohn et al., 2024)

### Get in touch

We work at the Andy Ball lab, RMIT University, Bundoora, Melbourne and are part of the Industry-led Biosolids Training Centre. Email me or comment on the discussion section of the GitHub repository for this GitBook. You will need to get a GitHub account to join the discussion. Its free.

More about me and my PhD research can be found here: <https://clean-dirt-digests.netlify.app>.

Follow me on X (hardly use that anymore though) or LinkedIn.



Bio:

Dr Chris Krohn is an early career researcher whose interests could be summed up with: “Environmental sequencing, microbial ecology, chlorinated pollutants, organic matter, wastewater, anaerobic digestion, and how everything connects”.

In 2021 I joined the ARC Biosolids Training Centre at RMIT ([www.transformingbiosolids.org.au](http://www.transformingbiosolids.org.au)), where we help water utilities to improve circular resource management by getting more renewable biogas and carbon/fertiliser values out of our municipal biosolids (essentially our poo). In project 1C of the Centre I develop metagenomic (DNA-based) methods to monitor the microbiome of anaerobic digestion, an important microbial treatment process for wastewater. I believe DNA-based diagnoses of wastewater sludges will help the water/biosolids sector improve resource recoveries and risk management.

Before that, after a career in one of the most fast-cycled and short-sighted manufacturing industries that took me from Germany to Vietnam and Hong Kong/Shenzhen, I decided to hit the switch and start thinking long-term and circular. Ten back-to-uni years later, in 2021 I finished a PhD in Soil Science at La Trobe Uni where I sequenced soil DNA and explored if and how soil biology was involved in the degradation of extremely persistent legacy pesticides that contaminate agricultural surface soils since several decades.

This work is licensed under a Creative Commons Attribution 4.0 International License.



# **Chapter 2**

## **Prerequisites, equipment and consumables**

### **2.1 Prerequisites**

- Experience with basic molecular laboratory techniques and DNA quality control (use of laminar flow hood, pipetting micro volumes, handling and preserving small volume reagents, sterilisation of consumables, doing gel electrophoresis)
- Experience with loading Nanopore flowcells
  - It is vital that loading of the flowcell is practiced. The key is to avoid any air entering the flow cell as that will completely destroy the pores. Use an old flowcell if available and run through the process with water to see where the buffers flow. Afterwards we recommend doing a Control run with Lambda DNA using the Control Expansion Kit).
- Confident with R, R Studio and package environment managers such as Conda and Docker. Knowledge in shell scripting and Linux syntax.

### **2.2 DNA extraction kit**

We commonly extract DNA from soils or from wastewater sludges using the DNeasy Powersoil Pro Kit (Qiagen) for both, soil and sludge. It resulted in high quality DNA, suitable for library preparation and sequencing. Extraction with this kit includes a bead beating step to release DNA from cells out of difficult matrices such as soils and sludge (biofilms) or similar. The sheared DNA will be fragmented with a N50 of ~7 kb in some cases (Jensen et al., 2024). Shearing of DNA may be minimised by reducing/optimising bead-beating duration and speed. Other commonly used and commercially available kits for DNA

## 10 CHAPTER 2. PREREQUISITES, EQUIPMENT AND CONSUMABLES

extraction are available (Jensen et al., 2024; Gand et al., 2023).

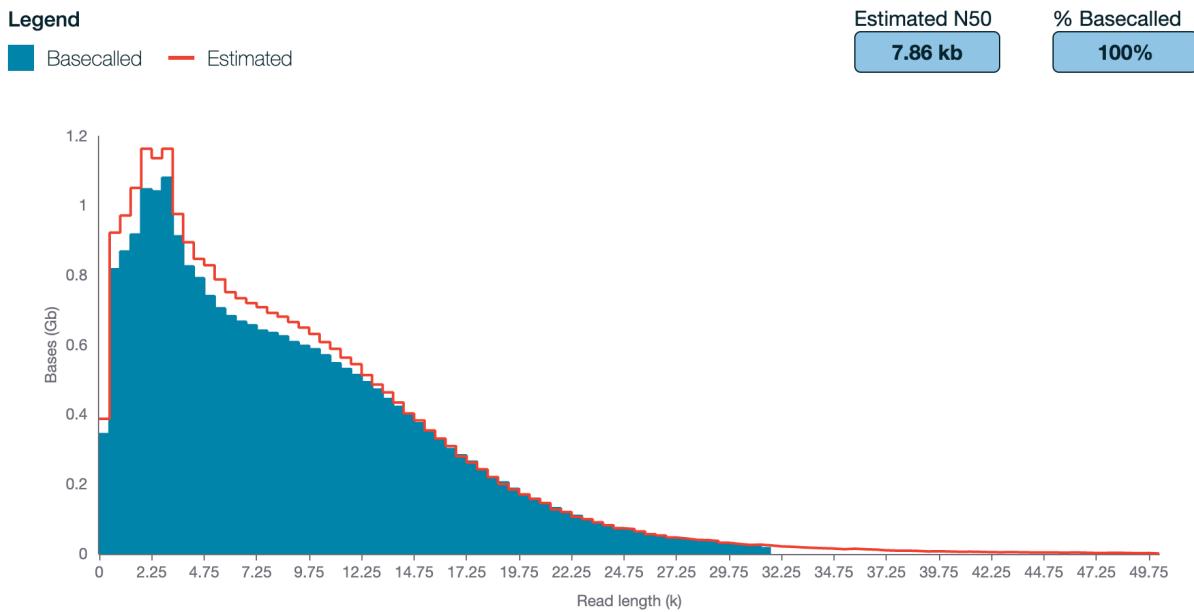


Figure 2.1: “Example of read size distribution of sludge DNA extracted with the DNeasy Powersoil Pro Kit in our lab”

### 2.3 Other consumables

- Native Barcoding Kit (Oxford Nanopore, SQK-NBD114.96)
- R10.4.1 flowcell (Oxford Nanopore, FLO-MIN114)
- NEBNext Quick Ligation Module (E6056, New England Biolab)
- NEBNext Ultra II End repair/dA-tailing Module (E7546S, New England Biolab)
- Blunt/TA Ligase Master Mix (M0367, New England Biolab)
- 10 mM dNTPs (N0447S, New England Biolab)
- Q5 Hot Start High-Fidelity DNA Polymerase (M0493, New England Biolab)
- DNeasy Powersoil Pro Kit (#47016, Qiagen)
- Qubit 1X dsDNA BR or HR Assay Kit (Q33266 or Q33231, Thermo Fisher)
- Eppendorf LoBind tubes (Eppendorf)
- twin.tec® PCR plate 96 LoBind, semi skirted (0030129504, Eppendorf)
- JetSeq Clean Magnetic beads (MER-BIO-68031, Millenium Science) - to clean up PCR products

- 10  $\mu$ M Forward Primer A519F (CAGCMGCCGCGGTAA) (Martijn et al., 2019)
- 10  $\mu$ M Reverse Primer U2428R (CCRAMCTGTCTCACGACG) (Martijn et al., 2019)
- Nuclease-free water
- 10 mM Tris-HCl pH 8.0 with 50 mM NaCl (UltraPure<sup>TM</sup> 1M Tris-HCl, pH 8.0 #15568025 & NaCl (5 M), RNase-free #AM9760G, Thermo Fisher)
- Bovine Serum Albumin (BSA) (50 mg/ml) (AM2616, InvitrogenTM UltraPure)
- HyperLadder 1kb (BIO-33025, Bioline, Millenium Science)
- 80% ethanol, freshly prepared in nuclease-free water
- Flow Cell Wash Kit (e.g. EXP-WSH004, Oxford Nanopore)

## 2.4 Protocols

- A protocol for PCR amplification and clean-up (included in this Gitbook).
- **Ligation sequencing amplicons - Native Barcoding Kit 96 V14** available on nanoporetech.com.
- Prepare spreadsheets to normalise DNA into 96-well plates to avoid pipetting errors.
- **Flow Cell Wash Kit (EXP-WSH004 or EXP-WSH004-XL)-minion.pdf** available on nanopore.com.

## 2.5 Equipment

- Nanopore sequencer. This protocol is based on the MinION Mk1C but is not limited to this model.
- A Linux computer or server instance for basecalling nanopore .pod5 files at super high accuracy (SUP) (unless you own the Promethion P2i which has sufficient GPU power to keep up with SUP basecalling). More info below.
- Nanodrop spectrophotometer to check DNA quality.
- Qubit fluorometer (Thermo Fisher).
- PCR thermal cycler.
- Magnets for 96-well plates and 1.5 ml tubes.
- Gel electrophoresis equipment.
- Vortex. We use the Vortex-Genie 2, including the 24 x 2 ml tube adapter for bead beating.
- Hula mixer (#15920D, Thermo Fisher) or similar overhead mixers.
- Eppendorf LoBind tubes (Eppendorf)

## 2.6 Computational resources and database

- For basecalling the raw nanopore data (pod5 files) with dorado - a Linux instance or computer is required with 2 TB storage, 64 GB DDR4 RAM and at least one Nvidia A100 GPU (e.g. GeForce RTX 4090). Note that the MinION Mk1C or the GridION can also basecall the data (much easier). But in the case of the MinION only at High Accuracy (HAC) and not at Super High Accuracy (SUP). But even at HAC it takes days to finish. Note that SUP achieves ~Q20-Q25 while HAC only ~Q18.
- 16S-ITS-23S Database. To be confirmed.

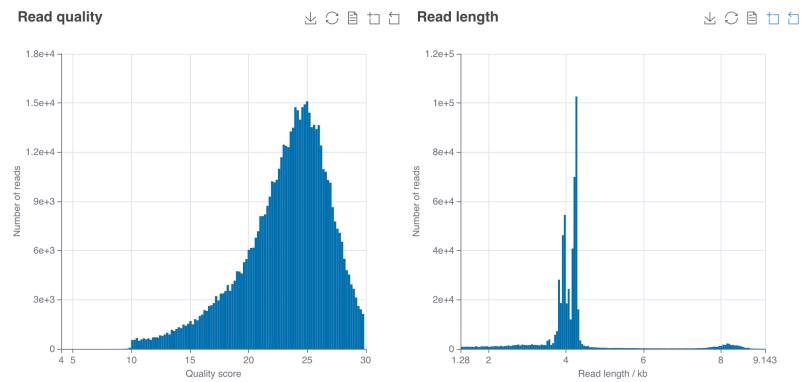


Figure 2.2: “Example of quality distribution of 4kb amplicon reads basecalled at super high accuracy with dorado.”

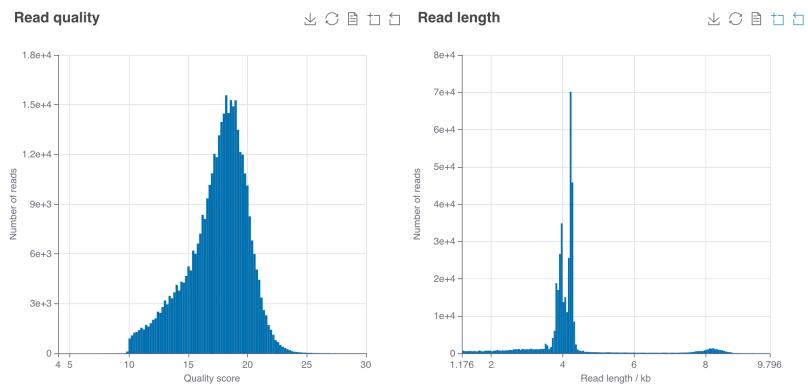


Figure 2.3: “Example of quality distribution of 4kb amplicon reads basecalled at high accuracy with dorado.”

# **Chapter 3**

## **Protocol**

### **3.1 DNA extraction**

#### **3.1.1 Checklist**

- Extraction kit - DNeasy Powersoil Pro, Qiagen, Hilden, Germany. See Section 2 for more details.
- Alternatively the high throughput version of the kit used with the QIAcube DNeasy 96 PowerSoil Pro QIAcube HT Kit.
- Vortex with 24 x 1.5 mL tube adapter (e.g. Vortex Genie 2 + adapter). Alternatively, a PowerLyzer Homogenizer.
- NanoDrop spectrophotometer to assess DNA quality.
- Qubit fluorometer to accurately measure DNA concentrations.

#### **3.1.2 Steps**

- Follow the extraction kit's protocol with 15 mins bead beating using a Vortex-Genie and 24-tube adapter. Reduce this to 10 mins if less than 24 samples are extracted or if shearing of DNA should be minimised.
- Measure DNA quality using DNA extract (1 µL) using a Nanodrop spectrophotometer
- Measure DNA concentration using a Qubit fluorometer.
- Into a 96-well plate, normalise extracted DNA to required PCR concentrations. For example, if 10 ng template is required for PCR, normalise DNA to 5 ng/µL.
- Store DNA at 4 °C until library preparation - no more than 1 week.
- Store DNA at -20 °C if sequencing is more than 1 week away.

## 3.2 Amplification of 16S-ITS-23 operon

### 3.2.1 Checklist

- 10 mM dNTPs (N0447S, New England Biolab)
- Q5 Hot Start High-Fidelity DNA Polymerase (M0493, New England Bio-lab)
- 10 µM Forward Primer A519F (CAGCMGCCGCGGTAA) (Martijn et al., 2019)
- 10 µM Reverse Primer U2428R (CCRAMCTGTCTCACGACG) (Martijn et al., 2019)
- JetSeq Clean Magnetic beads - or equivalent (MER-BIO-68031, Millenium Science) - to clean up PCR products
- twin.tec® PCR plate 96 LoBind, semi skirted (0030129504, Eppendorf)
- Nuclease-free water
- 10 mM Tris-HCl pH 8.0 with 50 mM NaCl (UltraPure™ 1M Tris-HCl, pH 8.0 #15568025 & NaCl (5 M), RNase-free #AM9760G, Thermo Fisher)
- HyperLadder 1kb (BIO-33025, Bioline, Millenium Science)
- 80% ethanol, freshly prepared in nuclease-free water
- Qubit 1X dsDNA BR Assay Kit (Q33266, Thermo Fisher)

### Notes

- Do not vortex tubes during library preparation to prevent DNA fragmentation. Fragmentation of amplicons may lead to incomplete reads.
- The primer amplifies the whole rrn operon.

Benefits of targeting the whole rrn operon:

- Superior species-level resolution and accuracy (Cuscó et al., 2019; Srinivas et al., 2024).
- Covers Bacteria and Archaea.

Risks of targeting the whole rrn operon:

- Not as representative of true abundances as full-length 16S amplicons.
- Species with unlinked 16S and 23S rrn DNA will be missed with this approach (for example < 9% of rRNA genes in wastewater sludge (Brewer et al., 2020)).

### 3.2.2 PCR

Time required: ~4 hrs incl. 2hrs, 40mins PCR.

- Prepare a PCR mastermix for the required number of 50-µl reactions (Table 3.1). It may be necessary to combine two 50 µl reactions for each sample to produce sufficient amplicon mass for a final concentration of 200 fmol as input for the Native Barcoding Kit from Oxford Nanopore.
- Add 3 µl of eDNA (5ng /µl) into a 96-well plate (e.g. Eppendorf twin.tec® PCR plate 96 LoBind, semi-skirted) using a multichannel pipette.

Table 3.1: List of components for each reaction (each tube) for PCR. See <https://www.neb.com/en-au/protocols/2012/08/30/pcr-using-q5-hot-start-high-fidelity-dna-polymerase-m0493> for details.

Component	50 µl reaction	Final concentration
5X Q5 Reaction Buffer - M0493S NEB	10 µL	1X
10 mM dNTPs N0447S	1 µL	200 µM
10 µM Forward Primer	2.5 µL	0.5 µM
10 µM Reverse Primer	2.5 µL	0.5 µM
Template DNA - 15 ng	3 µL (5 ng/µL)	< 1,000 ng
Q5 Hot Start High-Fidelity DNA Polymerase -M0493S NEB	1.0 µL	0.04 U/µL
5X Q5 High GC Enhancer M0493S NEB	10 µL	(1X)
Nuclease-Free Water	to 50 µL	

- Add 47 µl of mastermix using a multichannel pipette and carefully pipette up and down 10 x
- Run thermocycler (Table 3.2).
- Verify amplification length via 1% agarose gel electrophoresis @ 100V for 30 min including a 1 kb ladder.
- Store at 4 °C overnight if needed.

**Important:**

- Use hot start polymerase for ease of use, with high fidelity/accuracy and one that is suitable for long amplicons. For example, Q5 High- Fidelity DNA Polymerase kit (New England Biolabs) with GC enhancer (Martijn et al., 2019).
- 200 fmol is required per sample for the Native Barcoding Kit from ONP. Based on 4.25 kb (4-4.5kb), the final DNA concentration after cleaning up PCR products should be no less than 48 ng/µL (at 11.5 µl input volume) - giving 552 ng of 4.25kb amplicons.
- It may require two PCR reactions to achieve the required DNA amount (200 fmol); e.g. pool 2 x 50µl PCR products, clean combined and elute in 32.5 µl Tris. Check if 25 cycles (instead of 30) are enough to get sufficient yield. A lower cycle number lowers the risk of errors.

Table 3.2: Thermocycler conditions (Martijn et al 2019).

Cycle conditions
1 cycle:
30 s - Initial Denaturation 98 degree C
30 cycles:
10 s - 98 degree C
30 s - 64 degree C
210 s - 72 degree C
1 cycle:
10 min - Final Extension 72 degree C

### 3.3 PCR product clean-up

Clean with 0.6X JetSeq Clean Magnetic beads (or equivalent) and wash twice with 80% ethanol. This follows a similar protocol to the Illumina 16S-metagenomics library prep guide in case you are familiar with that.

Time required: ~ 1 hr per 24 samples.

#### 3.3.1 Checklist

- Magnetic rack for 96-well plate (e.g. #AM10027 Thermo Fisher)
- 80% ethanol, freshly prepared in nuclease-free water
- Nuclease-free water
- 10 mM Tris-HCl pH 8.0 with 50 mM NaCl (UltraPure™ 1M Tris-HCl, pH 8.0 #15568025 & NaCl (5 M), RNase-free #AM9760G, Thermo Fisher)
- JetSeq Clean Magnetic beads (MER-BIO-68031, Millenium Science) or equivalent - to clean up PCR products
- Qubit 1X dsDNA BR or HR Assay Kit (Q33266 or Q33231, Thermo Fisher)
- HyperLadder 1kb (BIO-33025, Bioline, Millenium Science)
- A worksheet to enter concentrations and calculate volumes for normalising DNA to 200 fmol

#### Note

Use wide bore tips for adding mag beads and subsequent (careful) mixing by pipetting to minimise DNA fragmentation.

#### 3.3.2 Steps

- Pool replicate reactions into one well - e.g. 2 x 50 µl = 100 µl in one well.
- Using a multi-channel pipette, add 0.6X beads to PCR products in the same wells (e.g. 60 µl to 100 µl pooled reactions), pipette 10 x up and

down and/or a gentle vortex after sealing the plate, careful not to spill (Vortex Genie speed 2).

- 10 min incubation at RT
- Place plate on a magnetic rack for 5 mins and remove supernatant.
- Wash twice with 200  $\mu$ l 80% ethanol in same 96-well plate, discard ethanol and let evaporate for ~ 5 mins or until completely removed.
- Take plate off magnet and elute pellet to a volume that provides appropriate DNA concentrations
  - e.g. add 32.5  $\mu$ l of elution buffer (10 mM Tris 50mM NaCl). The pellet will be stuck higher than the buffer surface in the well.
  - Hence, pipetting up and down 20–30 x may be necessary to carefully wash the pellet of the tube wall, using a wide bore tip.
  - And/or centrifuge plate at 200 rcf for 1 min.
- Off-magnet, incubate plate for 15 minutes at 37°C (e.g. in a water bath with only the wells submerged in water). Every 2 minutes, agitate the sample by gently vortexing the plate for 10 seconds (after sealing plate) to encourage DNA elution - and spin plate down.
- Collect 32  $\mu$ l of cleaned DNA after ~ 5 mins on magnet using wide bore tips.
- Quantify and record amplicon concentrations using Qubit HS or BS chemistry and record values. Based on 4 kb amplicons, the final DNA concentration after cleaning up PCR products should be no less than 45 ng/ $\mu$ L to have sufficient material for downstream library preparation (200 fmol, > 520 ng of 4 kb amplicons).
- Verify amplification length via 0.5-1% agarose gel electrophoresis @ 100V for 30 min including a 1 kb ladder.
- For each sample calculate volumes required of cleaned DNA and H<sub>2</sub>O to get 200 fmol in 11.5  $\mu$ l - for library preparation with the Native Barcoding Kit from Oxford Nanopore. Use V1 = C2\*V2 / C1, where V1 = x, C1 = DNA concentrations (ng/ $\mu$ l), C2 = 45.22 ng/ $\mu$ l, V2 = 11.5  $\mu$ l). H<sub>2</sub>O volume = 11.5  $\mu$ l - V1. These volumes will be used in the next step.

### 3.4 Library preparation

DNA Library preparation, including barcoding and pooling of samples (multiplexing). Here the DNA input are >4 kb amplicons of the 16S-ITS-23S rRNA

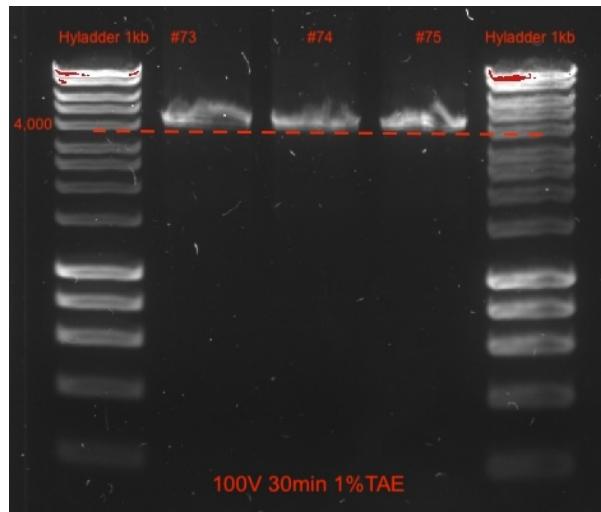


Figure 3.1: Gel confirming amplification of ~4 kb amplicons

Libray input calcs for Native Barcoding Kit V14	
What is the mean DNA fragment size?	4,000 bp
	650 g/mol
	2,600,000 g/mol
	0.0002 required nanomol
	0.20 required picomol
Input concentration	<b>200</b> required femtomol
DNA mass needed	520 ng
Volume needed (V2)	<b>11.5</b> $\mu$ l
DNA concentration needed (C2)	<b>45.22</b> ng/ $\mu$ l

Figure 3.2: Target concentration of library to normalise to 200 fmol based on a 4 kb amplicon .

Sample ID	C1 (ng/ $\mu$ l)	V1 ( $\mu$ l)	C2 (ng/ $\mu$ l)	H2O ( $\mu$ l)
#073	107.00	4.86	45.2	6.64
#074	92.70	5.61	45.2	5.89
#075	134.00	3.88	45.2	7.62
#076	109.00	4.77	45.2	6.73
#077	110.00	4.73	45.2	6.77
#078	111.00	4.68	45.2	6.82
#117	102.00	5.10	45.2	6.40

Figure 3.3: Example of a worksheet for calculating volumes ( $\mu$ l) of PCR products and H<sub>2</sub>O to get 200 fmol in 11.5  $\mu$ l, required as input for each sample with the Native Barcoding Kit. C1 = DNA concentration (ng/ $\mu$ l), C2 (library concentration) = 45.22 ng/ $\mu$ l, V2 (target volume) = 11.5  $\mu$ l.

operon from environmental DNA. The Long Fragment Buffer (LFB) is used (not the Short Fragment Buffer)

Time required: ~ 6 hours from start to finished library ready for loading.

#### Notes

- Follow the protocol [Ligation sequencing amplicons - Native Barcoding Kit 96 V14 \(SQK-NBD114.96\)-minion.pdf](#) available on nanoporetech.com. For details about the barcodes check here: [https://community.nanoporetech.com/technical\\_documents/chemistry-technical-document/v/chtd\\_500\\_v1\\_revaq\\_07jul2016/barcode-sequences](https://community.nanoporetech.com/technical_documents/chemistry-technical-document/v/chtd_500_v1_revaq_07jul2016/barcode-sequences). For final library calculations we assume that around 300 bps are added to the amplicons through barcodes, flanking sequences and adapters. If you know more - let us know on the Github discussion section.
- Update the MinION packages and MinKNOW software.
- Prepare a sample sheet for upload to instrument if desired (not essential but saves a little time when setting up the sequencing in MinKNOW - AND you can use it for demultiplexing with dorado later). More info on how to upload sample sheets [here](#).
- Amplicon product yields from PCR can vary widely depending on environmental source.
- To achieve consistent, normalised concentrations (200 fmol) across all samples, add appropriate input volume (calculated with  $C1 \cdot V1 = C2 \cdot V2$ ) and fill to 11.5  $\mu$ l with H<sub>2</sub>O to the first step of the Native Barcoding kit protocol.
- If you plan to sequence after library preparations, insert a flow cell now

and do a **flow cell check** and record the number of pores. This will take around 10 minutes. Afterwards store it at 4 °C until ready to load. We got 1,500-1,750 pores with new flowcells (R10.4.1) that were stored for two weeks at 4 degrees.

### 3.4.1 Checklist

- Native Barcoding Kit (Oxford Nanopore, SQK-NBD114.96)
  - End-prep
    - \* DNA Control Sample (DCS)
    - \* Elution Buffer (EB)
  - Native barcode ligation
    - \* AMPure XP Beads (AXP)
    - \* EDTA
    - \* Native Barcodes (NB01-96)
  - Sequencing adapter ligation
    - \* Native Adapter (NA)
    - \* Elution Buffer (EB)
    - \* AMPure XP Beads (AXP)
    - \* Long Fragment Buffer (LFB)
- NEBNext Quick Ligation Module (E6056, New England Biolab)
- NEBNext Ultra II End repair/dA-tailing Module (E7546S, New England Biolab)
- Blunt/TA Ligase Master Mix (M0367, New England Biolab)
- 80% ethanol, freshly prepared in nuclease-free water
- Nuclease-free water
- Qubit 1X dsDNA BR or HR Assay Kit (Q33266 or Q33231, Thermo Fisher)
- Eppendorf LoBind tubes (Eppendorf)
- Magnet for 1.5 ml tubes
- Hula mixer (#15920D, Thermo Fisher) or similar overhead mixers

### 3.4.2 Steps

**Follow the Nanopore protocol. There are four parts:**

1. End-prepping of 11.5 µL DNA (200 fmol).  
Pipetting into the 96-well plate may take time as there are two rounds of pipetting, one for the DNA amplicon volumes and one for H<sub>2</sub>O volumes to get 200 fmol amplicon DNA in 11.5 µl, as well as the remaining reagents for end-prepping the DNA. It is not possible to use multichannel pipettes for this step.
2. Native barcode ligation using 0.75 µL of end-prepped DNA using one of 96 barcodes from the kit for each sample, plus pooling of all barcoded samples into a LoBind tube (1.5 mL).

3. Bead clean-up of pooled end-prepped DNA library and elution of 35  $\mu\text{L}$  volume; quantification with Qubit 1X dsDNA HS Assay Kit. We used all of the available end-prepped volume (e.g. 33  $\mu\text{l}$  after Qubit quantification) for the next step, i.o. only 30  $\mu\text{l}$  as per protocol.

SAFE STOP: Store at 4 °C overnight if needed.

4. Adapter ligation. Once this step is completed, you get a final volume of 15  $\mu\text{l}$  of clean end-prepped/barcoded/adapter ligated DNA.

- We have stored the library at 4 °C overnight and sucessfully sequenced it the next day (although it is preferred to sequence right after adapter ligation).
- The concentration we achieved was ~20-25 ng/ $\mu\text{l}$ , which means there was sufficient DNA for two runs. After quantify DNA concentrations with 1  $\mu\text{l}$  DNA there are 14  $\mu\text{l}$  left.
- Nanopore's recommended loading concentrations are 35–50 fmol at 12  $\mu\text{L}$  volume. For example, at a concentration of 25.3 ng/ $\mu\text{l}$  (4,300 kb amplicon) a total of 5.50  $\mu\text{l}$  of DNA library (plus 6.50  $\mu\text{l}$  H<sub>2</sub>O) are sufficient. The remaining 6.50  $\mu\text{l}$  DNA library can be stored in fridge overnight if required.

## 3.5 Priming and loading of flow cell - run start

Time required: ~10-15 mins for priming and loading. 24 hours for sequencing.

### 3.5.1 Checklist

- The same protocol: [Ligation sequencing amplicons - Native Barcoding Kit 96 V14 \(SQK-NBD114.96\)-minion.pdf](#) available on nanoporetech.com
- Bovine Serum Albumin (BSA) (50 mg/ml) (AM2616, InvitrogenTM UltraPure)
- Native Barcoding Kit (Oxford Nanopore, SQK-NBD114.96)
  - Flow Cell Flush (FCF)
  - Flow Cell Tether (FCT)
  - Sequencing Buffer (SB)
  - Library Beads (LIB)
- R10.4.1 flowcell (Oxford Nanopore, FLO-MIN114)

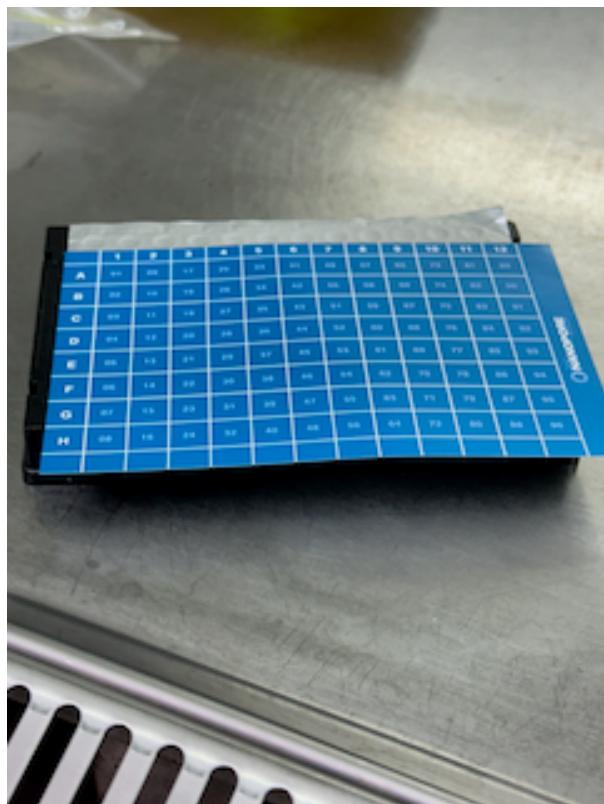


Figure 3.4: The Native Barcode Kit (SQK-NBD114.96) comes with 96 individual barcodes

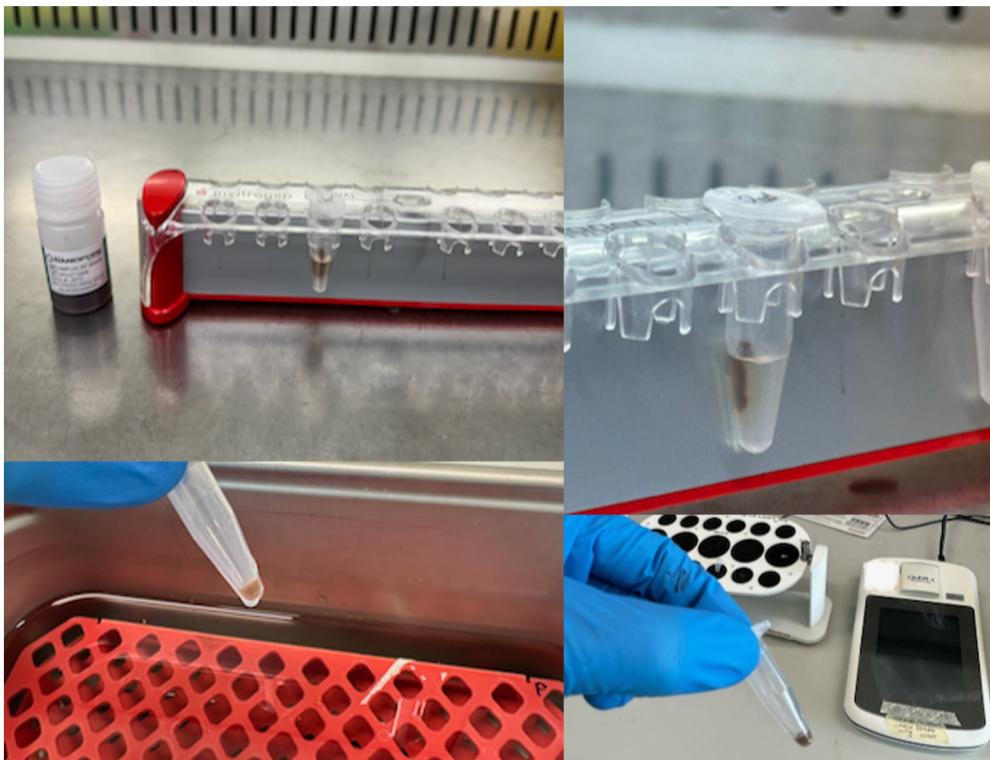


Figure 3.5: Bead clean-up of pooled end-prepped and adapter-ligated DNA library

FINAL LIBRARY FOR LOADING	
What is the mean DNA fragment size?	4,300 bp
	650 g/mol
	2,795,000 g/mol
	0.00005 required nanomol
	0.05 required picomol
Input concentration	50 required femtomol
DNA mass needed	140 ng
Volume needed (V2)	12.0 $\mu$ l
DNA concentration needed (C2)	11.65 ng/ $\mu$ l
Actual concentration of amplicon library (C1)	25.3 ng/ $\mu$ l
V1	5.52 $\mu$ l

Figure 3.6: Calculation to get library volume for loading (In this example we used 5.52  $\mu$ l of library, plus 6.48  $\mu$ l H<sub>2</sub>O for a final library volume of 12  $\mu$ l at 50 fmol concentration)

### 3.5.2 Steps

**Follow the Nanopore protocol. There are five steps:**

1. Preparation and loading of sequencing buffer and wait for 5 minutes.
2. Preparation of library. Nanopore's protocol recommends loading 35–50 fmol at 12 µL volume.
3. Loading of remaining sequencing buffer
4. Loading of library
5. Setting up the run for 24 hours and press start! More details will follow about recommended hours and expected yield based on 24 samples.

## 3.6 Flow cell wash

Time required: ~ 1.5 hours.

Wash the flow cell once the run has finished. Note that after the wash you can either load a new library directly afterwards or add storage buffer to load a new library later (always store flow at 4 °C).

However, it is only possible to check the number of pores (and estimate potential yield) after adding storage buffer. The washing buffer is not suitable for a running a flow cell check. So loading a new library directly is only sensible if the user knows how many active pores to expect. Although, the instrument will do a pore scan at the start, it will be too late once you loaded the new library.

As a gauge, we ran a flowcell with 1,500 pores for 24 hours (loaded with 24 bar-coded samples of 16S-ITS-23S amplicons). At the end 558 active pores remained before the wash.

Note that washing the flowcell does not interfere with any active basecalling processes on the Mk1C, which may take days to finish if High Accuracy was chosen.

### 3.6.1 Checklist

- Flow Cell Wash Kit (e.g. EXP-WSH004, Oxford Nanopore)
- Protocol [Flow Cell Wash Kit \(EXP-WSH004 or EXP-WSH004-XL\)-minion.pdf](#) available on nanopore.com

### 3.6.2 Steps

**Follow the Nanopore protocol. There are three steps:**

1. Preparation of 400 µl washing buffer
2. Load 200 µl of washing buffer, wait 5 minutes, then add the remaining 200 µl and wait 1 hour.
3. If you plan to store the flowcell for later use - add storage buffer. Otherwise load new library.

## 3.7 Basecalling

Time required: ~ several days depending on computational resources.

If you have a PromethION integrated (P2i) you can run Super High Accuracy (SUP) basecalling on the device in almost real time (it has an Nvidia GPU). In that case there is no need to copy POD5 files from the instrument to another computer for SUP basecalling on a separate GPU device.

The below assumes that the **SQK-NBD114-96** kit was used for preparing the amplicon library.

### 3.7.1 Checklist

- dorado installed on an instance or device with either Apple silicon (M1/M2) or Nvidia GPUs (the latest Nvidia RTX 4090 performs best it seems).
- POD5 files from Nanopore sequencer (either copy on USB or connect do instrument folders through `ssh`)
- samplesheet.csv (More details about the sample sheet in the library prep section).

### 3.7.2 Steps

1. First let the Nanopore sequencer finish basecalling with the Fast model (unless you have a Promethion Pi, which will provide you with SUP-basecalled bam/fastq files, i.e. the following steps are not required in this case).
2. Copy POD5 files from sequencer to a GPU-powered computer or instance, into a folder called `pod5/`. In our case that was 85GB of data (But expect a bit more; our MinION had a power outage just before all pod5 finished generating... :( ).
3. Run `dorado basecaller sup pod5/ > calls.bam`.
4. Demultiplex the resulting `.bam` file using `dorado demux --kit-name SQK-NBD114-96 --sample-sheet samplesheet.csv --output-dir 16SITS23S_samples/ --emit-fastq call`.

At the end you should have one `.fastq` file for each sample, ready for further quality trimming and read classification.

## 3.8 Read classification

The following method (using the `wf-metagenome` workflow from EPI2ME, with kmer-based read predictions) will provide classified reads of Bacteria and Archaea but, depending on the read quality, will result in some false positives.

And abundances will not be a true representation of abundances as no clustering, dereplication or ASV generation is involved. Results will not be comparable to other studies.

However, it will provide read classification to species-level to some degree of certainty, which may be useful in cases when you are not interested in the true representation of relative abundances. For example, if you are after temporal variation or differences between your samples in your experiment.

As rrna operon sequencing is relatively new, the classification methods and databases are still evolving. If you want to explore other methods, I recommend starting with the following papers: (Curry et al., 2022), (Rodríguez-Pérez et al., 2021) and (Walsh et al., 2024).

### 3.8.1 Checklist

- EPI2ME installed (we use the desktop version). It requires Docker and Java so if this is your first time installing/using EPI2ME it may take you some time to get started.
- Fastq files in a folder
- Chopper installed.

### 3.8.2 Steps

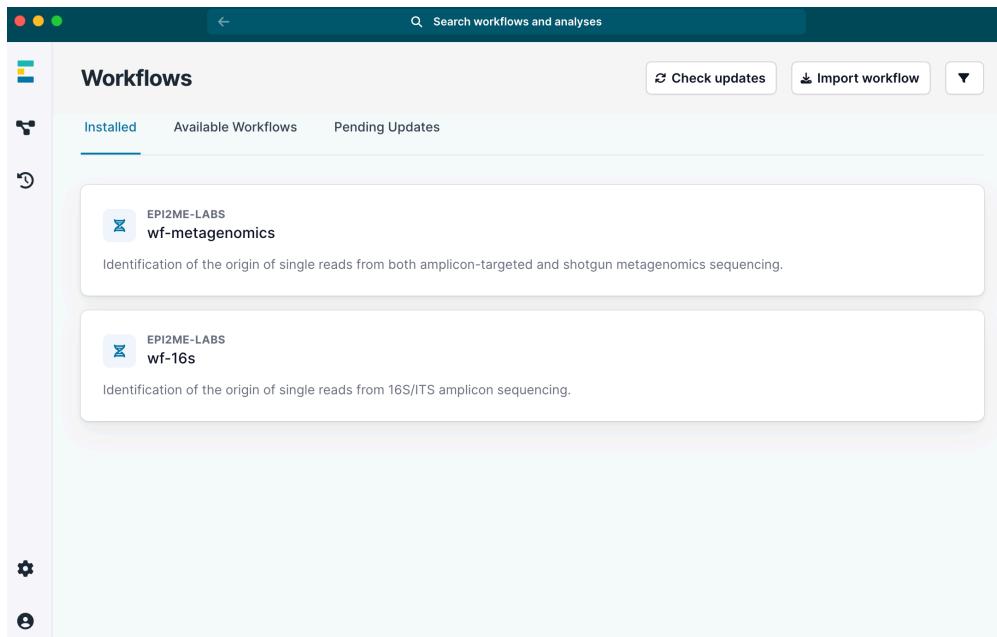
1. Trim fastq files. In this case we used the tool Chopper to remove any reads below Q20 and filter by length.

```
#!/bin/bash
for file in *.fastq; do
    # Construct the output filename based on input filename
    output_file="${file%.fastq}_filtered.fastq"

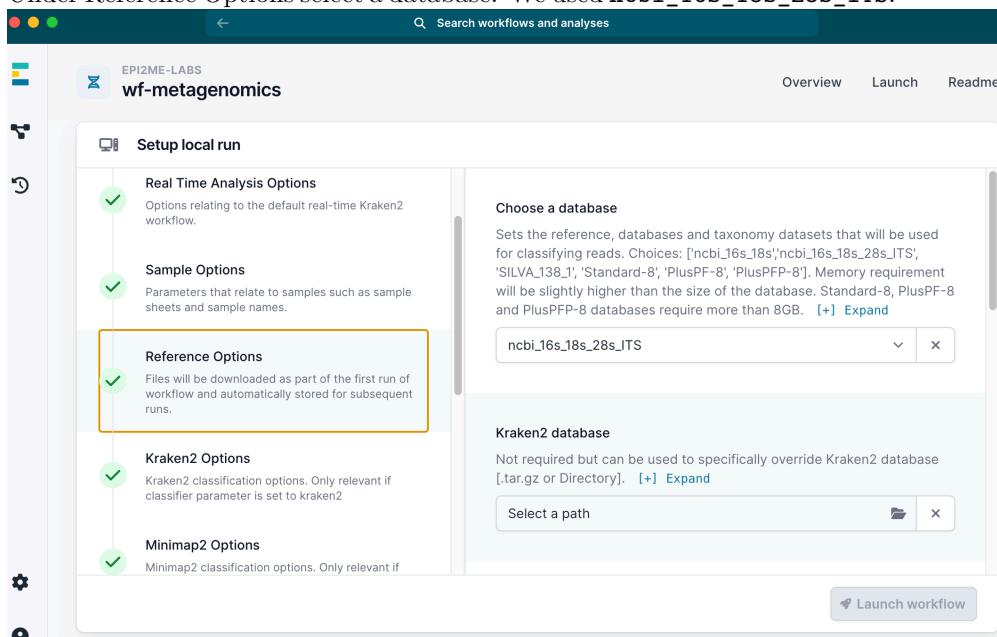
    # Run the chopper command on the current file
    chopper -q 20 --minlength 3000 --maxlength 5000 -i "$file" > ./Test/$output_file

    echo "Filtered reads from $file saved to $output_file"
done
```

2. Within EPI2ME, download, install and select the **wf-metagenome workflow**.



3. Click Run this workflow, then Run on your computer.
4. Select the path to your fastq files.
5. Select classification method. Either kraken2 or minimap2
6. Under Reference Options select a database. We used `ncbi_16s_18s_28s_ITS`.



7. Under Sample Options select you sample\_sheet.csv.

**Sample Options**  
Parameters that relate to samples such as sample sheets and sample names.

**Reference Options**  
Files will be downloaded as part of the first run of workflow and automatically stored for subsequent runs.

**Kraken2 Options**  
Kraken2 classification options. Only relevant if classifier parameter is set to kraken2.

**Minimap2 Options**  
Minimap2 classification options. Only relevant if classifier parameter is set to minimap2.

**Sample sheet**  
A CSV file used to map barcodes to sample aliases. The sample sheet can be provided when the input data is a directory containing sub-directories with FASTQ files. Disabled in the real time pipeline. [\[+\] Expand](#)

Select a path

**Sample name**  
A single sample name for non-multiplexed data. Permissible if passing a single .fastq(.gz) file or directory of .fastq(.gz) files. Disabled in the real time pipeline.

Enter a value

**Launch workflow**

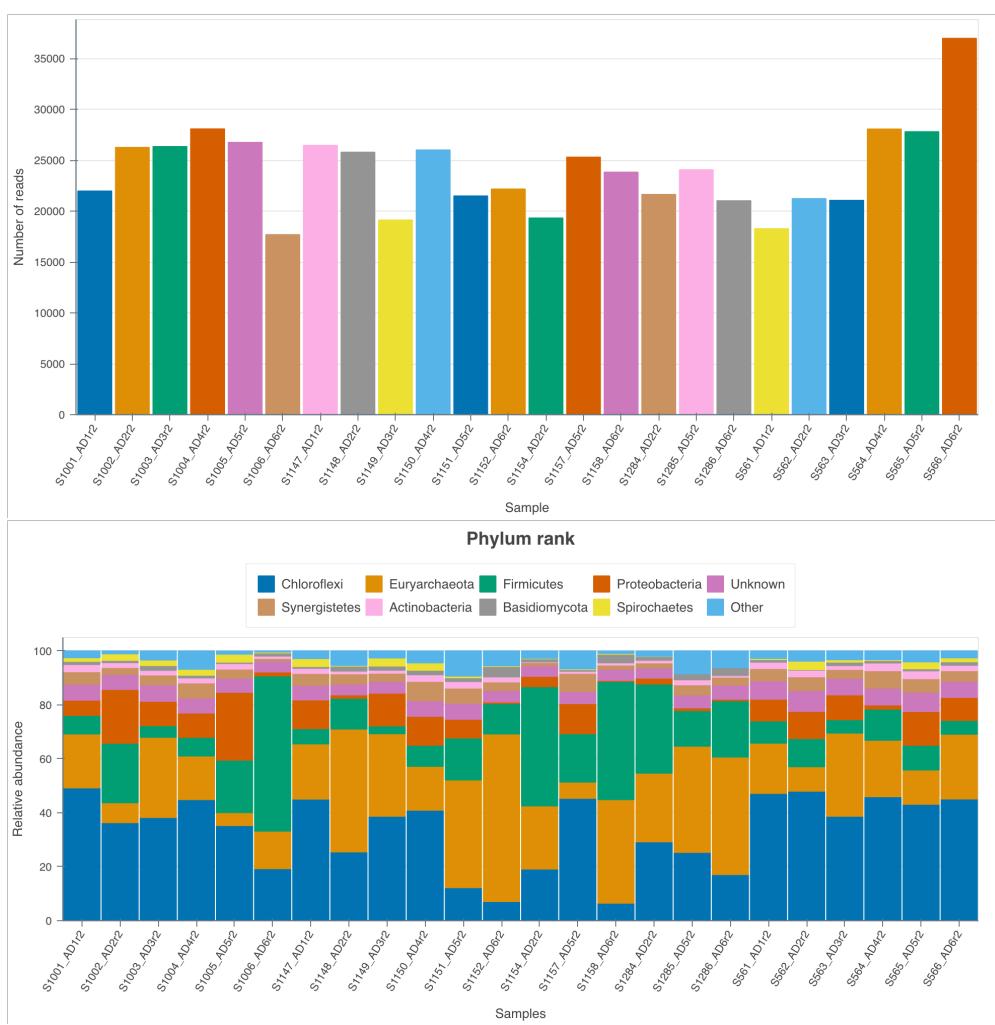
position_id	flow_cell_product_code	kit	sample_id	experiment_id	barcode	alias
MC-113910	FLO-MIN114	SQK-NBD114-96	Sludge	2024_16SITS23S-3	barcode01	S561_AD1r2
MC-113910	FLO-MIN114	SQK-NBD114-96	Sludge	2024_16SITS23S-3	barcode02	S562_AD2r2
MC-113910	FLO-MIN114	SQK-NBD114-96	Sludge	2024_16SITS23S-3	barcode03	S563_AD3r2
MC-113910	FLO-MIN114	SQK-NBD114-96	Sludge	2024_16SITS23S-3	barcode04	S564_AD4r2
MC-113910	FLO-MIN114	SQK-NBD114-96	Sludge	2024_16SITS23S-3	barcode05	S565_AD5r2
MC-113910	FLO-MIN114	SQK-NBD114-96	Sludge	2024_16SITS23S-3	barcode06	S566_AD6r2
MC-113910	FLO-MIN114	SQK-NBD114-96	Sludge	2024_16SITS23S-3	barcode07	S1001_AD1r2
MC-113910	FLO-MIN114	SQK-NBD114-96	Sludge	2024_16SITS23S-3	barcode08	S1002_AD2r2
MC-113910	FLO-MIN114	SQK-NBD114-96	Sludge	2024_16SITS23S-3	barcode09	S1003_AD3r2
MC-113910	FLO-MIN114	SQK-NBD114-96	Sludge	2024_16SITS23S-3	barcode10	S1004_AD4r2
MC-113910	FLO-MIN114	SQK-NBD114-96	Sludge	2024_16SITS23S-3	barcode11	S1005_AD5r2
MC-113910	FLO-MIN114	SQK-NBD114-96	Sludge	2024_16SITS23S-3	barcode12	S1006_AD6r2
MC-113910	FLO-MIN114	SQK-NBD114-96	Sludge	2024_16SITS23S-3	barcode13	S1147_AD1r2
MC-113910	FLO-MIN114	SQK-NBD114-96	Sludge	2024_16SITS23S-3	barcode14	S1148_AD2r2
MC-113910	FLO-MIN114	SQK-NBD114-96	Sludge	2024_16SITS23S-3	barcode15	S1149_AD3r2
MC-113910	FLO-MIN114	SQK-NBD114-96	Sludge	2024_16SITS23S-3	barcode16	S1150_AD4r2
MC-113910	FLO-MIN114	SQK-NBD114-96	Sludge	2024_16SITS23S-3	barcode17	S1151_AD5r2
MC-113910	FLO-MIN114	SQK-NBD114-96	Sludge	2024_16SITS23S-3	barcode18	S1152_AD6r2
MC-113910	FLO-MIN114	SQK-NBD114-96	Sludge	2024_16SITS23S-3	barcode19	S1154_AD2r2
MC-113910	FLO-MIN114	SQK-NBD114-96	Sludge	2024_16SITS23S-3	barcode20	S1157_AD5r2
MC-113910	FLO-MIN114	SQK-NBD114-96	Sludge	2024_16SITS23S-3	barcode21	S1158_AD6r2
MC-113910	FLO-MIN114	SQK-NBD114-96	Sludge	2024_16SITS23S-3	barcode22	S1284_AD2r2
MC-113910	FLO-MIN114	SQK-NBD114-96	Sludge	2024_16SITS23S-3	barcode23	S1285_AD5r2
MC-113910	FLO-MIN114	SQK-NBD114-96	Sludge	2024_16SITS23S-3	barcode24	S1286_AD6r2

8. Then launch the workflow.
9. Once completed you will receive a `report.html` with some results, plus a `.tsv` table with abundances to species-level for further analysis in your favorite platform.

## 3.9 Results

In our case we got ~ 25,000 reads per sample containing 2,137 species. Upon first inspection it is clear that read depth is too low and that longer sequencing times are required than 24 hours for 24 samples. We aim for around 100,000 reads per sample. On a MinION flowcell I would aim to sequence for the full 72

hours next time.





tax	A	B	C	D	E	F	G
Bacteria;Bacteria_none;Chloroflexi;Anaerolineae;Anaerolineales;Anaerolineaceae;Levilinea;Levilinea saccharolytica	S1001,AD1r;S1002,AD2r;S1003,AD3r;S1004,AD4r;S1005,AD5r;S1006,AD6r;	4079	4053	3740	8687	4017	1307
Bacteria;Bacteria_none;Chloroflexi;Anaerolineae;Anaerolineales;Anaerolineaceae;Leptolinea;Leptolinea tardivitalis		4625	4782	4616	3208	4740	1728
Archaea;Archaea_none;Euryarchaeota;Methanobacteria;Methanobacteriales;Methanobacteriaceae;Methanobacterium;Methanobacterium beijingense	3776	706	5009	2889	882	397	
Archaea;Archaea_none;Euryarchaeota;Methanobacteria;Methanobacteriales;Methanobacteriaceae;Methanobacterium;Methanobacterium congolense	21	19	29	5	5	9	
Unclassified;Unknown;Unknown;Unknown;Unknown;Unknown		1291	1435	1595	1545	1371	726
Bacteria;Bacteria_none;Firmicutes;Clostridia;Eubacteriales;Peptostreptococcaceae;Romboutsia;Romboutsia timonensis		234	2888	204	443	2653	5624
Bacteria;Bacteria_none;Proteobacteria;Alphaproteobacteria;Rhodospirillales;Acetobacteraceae;Acidomonas;Acidomonas methanolia		272	4517	105	2132	5905	111
Archaea;Archaea_none;Euryarchaeota;Methanobacteria;Methanobacteriales;Methanobacteriaceae;Methanobrevibacter;Methanobrevibacter acididurii		48	1085	48	12	306	1990
Bacteria;Bacteria_none;Proteobacteria;Gammaproteobacteria;Pasteurellales;Pasteurellaceae;Rodentibacter;Rodentibacter rarus		758	8	1965	28	7	27
Archaea;Archaea_none;Euryarchaeota;Methanobacteria;Methanobacteriales;Methanobacteriaceae;Methanobrevibacter;Methanobrevibacter boviskori		6	6	2	3	5	3
Bacteria;Bacteria_none;Synergistetes;Synergistetes;Incertae_sedis;Synergistetes;Incertae_sedis;Thermosynergaceae;Thermosynergites;Thermosyn	926	150	856	943	172	15	
Bacteria;Bacteria_none;Chloroflexi;Anaerolineae;Anaerolineales;Anaerolineaceae;Bellilinea;Bellilinea caldiftitiae		1291	39	952	77	58	21
Bacteria;Bacteria_none;Synergistetes;Synergistetes;Synergistates;Synergistaceae;Acetomicrobium;Acetomicrobium mobile		18	412	9	463	571	96
Archaea;Archaea_none;Euryarchaeota;Methanobacteria;Methanobacteriales;Methanobacteriaceae;Methanobacterium;Methanobacterium kanagiense		328	13	2265	39	9	2
Bacteria;Bacteria_none;Spirochaetes;Spirochaetales;Spirochaetaceae;Rectinemataceae;Rectinema;Rectinema cohabitans		294	625	532	611	787	56
Bacteria;Bacteria_none;Firmicutes;Clostridia;Eubacteriales;Oscillospiraceae;Oscillospiraceae;Incertae_sedis;Clostridium;cellulosi		0	4	1	3	12	34
Bacteria;Bacteria_none;Coprothermobacterota;Coprothermobacteria;Coprothermobacteriales;Coprothermobacteraceae;Coprothermabacter;Coproth		2	4	2	31	15	3
Eukarya;Fungi;Basidiomycota;Agaricomycetes;Agaricales;Stephanosporaceae;Stephanospora;Stephanospora kanuka		101	187	131	145	122	187
Bacteria;Bacteria_none;Chloroflexi;Anaerolineae;Anaerolineales;Anaerolineaceae;Leptolinea;Leptolinea submarina		244	239	260	136	247	67
Bacteria;Bacteria_none;Thermotogae;Thermotogae;Thermotogales;Fervidobacteriaceae;Fervidobacterium;Fervidobacterium riparium		15	351	14	50	322	686
Bacteria;Bacteria_none;Firmicutes;Clostridia;Eubacteriales;Clostridaceae;Clostridium;Clostridium sudense		66	20	78	1252	46	3
Bacteria;Bacteria_none;Chloroflexi;Anaerolineae;Anaerolineales;Anaerolineaceae;Ornatilinea;Ornatilinea apprima		28	357	18	32	315	706
Bacteria;Bacteria_none;Firmicutes;Clostridia;Eubacteriales;Oscillospiraceae;Acetivibrio;Acetivibrio saccincola		350	7	123	5	3	0

Figure 3.7: Results

# Bibliography

- Brewer, T. E., Albertsen, M., Edwards, A., Kirkegaard, R. H., Rocha, E. P., and Fierer, N. (2020). Unlinked rRNA genes are widespread among bacteria and archaea. *ISME Journal*, 14(2):597–608.
- Curry, K. D., Wang, Q., Nute, M. G., Tyshaieva, A., Reeves, E., Soriano, S., Wu, Q., Graeber, E., Finzer, P., Mendling, W., Savidge, T., Villapol, S., Dilthey, A., and Treangen, T. J. (2022). Emu: species-level microbial community profiling of full-length 16S rRNA Oxford Nanopore sequencing data. *Nature Methods*, 19(7):845–853.
- Cuscó, A., Catozzi, C., Viñes, J., Sanchez, A., and Francino, O. (2019). Microbiota profiling with long amplicons using Nanopore sequencing: Full-length 16S rRNA gene and the 16S-ITS-23S of the rrn operon. *F1000Research*, 7.
- Gand, M., Bloemen, B., Vanneste, K., Roosens, N. H., and De Keersmaecker, S. C. (2023). Comparison of 6 DNA extraction methods for isolation of high yield of high molecular weight DNA suitable for shotgun metagenomics Nanopore sequencing to detect bacteria. *BMC Genomics*, 24(1):438.
- Jensen, T. B. N., Dall, S. M., Knutsson, S., Karst, S. M., and Albertsen, M. (2024). High-throughput DNA extraction and cost-effective miniaturized metagenome and amplicon library preparation of soil samples for DNA sequencing. *PLoS ONE*, 19(4 April):e0301446.
- Krohn, C., Jansriphilub, K., Dias, D. A., Rees, C. A., van den Akker, B., Boer, J. C., Plebanski, M., Surapaneni, A., O’Carroll, D., Stuetz, R., Batstone, D. J., and Ball, A. S. (2024). Dead in the water – Role of relic DNA and primer choice for targeted sequencing surveys of anaerobic sewage sludge intended for biological monitoring. *Water Research*, 253:121354.
- Martijn, J., Lind, A. E., Schön, M. E., Spiertz, I., Juzokaite, L., Bunikis, I., Pettersson, O. V., and Ettema, T. J. (2019). Confident phylogenetic identification of uncultured prokaryotes through long read amplicon sequencing of the 16S-ITS-23S rRNA operon. *Environmental Microbiology*, 21(7):2485–2498.
- Rodríguez-Pérez, H., Ciuffreda, L., and Flores, C. (2021). NanoCLUST: a

- species-level analysis of 16S rRNA nanopore sequencing data. *Bioinformatics*, 37(11):1600–1601.
- Srinivas, M., Walsh, C. J., Crispie, F., O’Sullivan, O., Cotter, P. D., Sinderen, D., and Kenny, J. G. (2024). Evaluating the Efficiency of 16S-ITS-23S operon Sequencing: A Comparison of Primer Pairs, Sequencing Platforms, and Taxonomic Classifiers.
- Walsh, C. J., Srinivas, M., Stinear, T. P., van Sinderen, D., Cotter, P. D., and Kenny, J. G. (2024). GROND: a quality-checked and publicly available database of full-length 16S-ITS-23S rRNA operon sequences. *Microbial genomics*, 10(6).