

# CIDR Metagenomics

---

*CIDR*

*None*

## Table of contents

---

1. Metagenomics network hub	3
2. Lab Resources	4
2.1 Panmetagenomics protocol	4
3. Bioinformatics	18
3.1 Clinical metagenomics bioinformatics	18
3.2 Setting up CIDR Metagenomics bioinformatics workflow	21
3.3 Running the metagenomics workflow	23
3.4 Bioinformatics - Running metagenomics	26
3.5 Defaults	26
3.6 Setting up CIDR Metagenomics bioinformatics workflow - alternative deployment	27
3.7 Example for control sample	28
3.8 Naming schema	28
4. Analysis	30
4.1 SOPs for clinical interpretation - coming soon	30

# 1. Metagenomics network hub

---

Note

**The tools and documentation described here and on the CIDR GitHub are not validated for diagnostic use and are for research and evaluation purposes only.**

**The Network Hub is a resource for users of the CIDR clinical metagenomics workflow. Here, you can find SOPs, technical and FAQ/troubleshooting information regarding the implementation of metagenomics in a clinical evaluation/research setting.**

## 1.0.1 Lab protocols

The lab protocol is a same-day DNA/RNA extraction, host-depletion and ONT library preparation workflow for delivery of preliminary sequencing results in < 6 hours.

## 1.0.2 Informatics workflow

The workflow covers the end-to-end processing of respiratory samples sequencing data, delivering a metageconomic report describing the microbial communities within them. The workflow leverages ONT Nanopore sequencing at its core to produce real-time sequencing data on the GridION platform. The informatics workflow runs locally alongside the sequencing experiment, producing reports as early as 30 minutes after commencing sequencing.

## 1.0.3 Reporting framework

This SOP is applied in a service evaluation setting to parse results from the informatics workflow for application in a clinical evaluation setting

## 2. Lab Resources

---

### 2.1 Panmetagenomics protocol

---

#### Note

Please reference use of this method in any presentation or publication as [Unified metagenomic method for rapid detection of bacteria, fungi and viruses in clinical samples | Research Square which is currently going through journal review process. Work conducted during development and evaluation of metagenomics protocols are published in Baldan R et al J Infect. 2021 83:167. Charalampous T et al Genome Medicine 2021 13:182 Charalampous T et al Am J Resp Crit Care Med. 2024 209:164-174.

#### Important

**Risk assessment for handling respiratory samples needs to be performed by each laboratory.**

**The tools and documentation described here and on the CIDR GitHub are not validated for diagnostic use and are for research and evaluation purposes only.**

#### 2.1.1 Preparation of quality controls:

- Positive control: NATtrol™ Respiratory Panel 2.1 (RP2.1) Controls (Zeptometrix):

1. Mix 300ul of control 1 and 300ul of control 2 in an Eppendorf. Label with date and lot number.
2. Vortex for 1 minute Centrifuge at 1,200g for 5min immediately prior to each use.
3. Aliquot the volume specified by the manual or automated extraction method and process in parallel with equivalent clinical sample volumes.
4. Store the mixed positive controls at 4°C, it can be used for two extractions.
5. Run a target PCR (Viasure respiratory panel III) anytime a new box is opened and recorded it on the trend analysis spreadsheet.

- Negative control: Respiratory Swab Matrix Negative Control (Vircell):

Add 500ul of Microbial DNA-Free Water to a vial and mix until completely reconstituted as per manufacturer's instructions. Vortex for 30 sec to dissolve and homogenise completely. Spike with 1ul of TMV.

**NOTE:** run the PC and NC using the same barcodes. We use barcode 11 for the PC and 12 for the NC.

- Internal control (IC): Tobacco Mosaic Virus (TMV) PC-0107 (DSMZ):

1. Reconstitute a new vial of TMV every Monday with 3mL of Microbial DNA-Free Water (Qiagen).
2. Vortex for 1 minute.
3. Aliquot 500ul of IC into 6 Eppendorf tubes, labelling each tube with the preparation date and lot number. Store at 4°C.
4. Use one aliquot per day to spike your samples and negative control, vortex for one minute before spiking
5. Do a target PCR on the aliquot used each time a new vial is reconstituted and record the Ct value on the trend analysis spreadsheet (see appendix 1). In our hands this is between 0.5ul and 5ul of TMV to give a final Ct value in the sample of between 5 and 6.
6. Discard any remaining volume from used aliquot each day and any remaining aliquots by Friday.

See Appendix 1 for the TMV targeted PCR.

- Spiking samples with the internal control (IC):

<!-- -->

- Spiked the sample with <u>1ul of TMV</u> is the following samples:

- Broncho-alveolar lavage (BAL),
- Non-directed broncho-alveolar lavage (NBL)
- Endotracheal tube (ETT) aspirates
- Sputum.

- Spike with <u>0.5ul of TMV</u> the following samples:

- Pleural fluids

- Nasal throat swabs

**<u>Sample preparation:</u>**

-Respiratory samples validated for this protocol are: BAL, NBL, ETT, NTS, sputum and pleural fluids.

-Samples processed should be fresh up to four days old storage in the fridge, older than that it could not be guarantee the RNA virus detection.

- Samples are processed in a different way depending on the sample type.

## **1. Sputum and ETT:**

Mucoid samples such as sputum or ETT should be mucolysed before starting the Human DNA depletion process as follows:

Add approximately 2 mL sputum into a numbered, sterile, plastic universal container or if 2 mL already in original container then use this:

1. Add equal volume (2mL) of mucolytic agent Mucolyse to the sample container.
2. Vortex in the safety cabinet for 30 seconds.
3. Leave until liquefaction at RT.
4. Gently agitate for a further 15 seconds before to proceed to the human DNA depletion

If after mucolysing the sample, this is still mucoid, pipette 500ul of the sample and mix with 500ul of PBS, vortex well and proceed with the human DNA depletion.

## **<uHuman DNA depletion:</u**

The human DNA depletion is three steps: centrifugation, bead-beating and endonuclease treatment. However, BAL, NBL and ETT aspirates are processed in a different way depending on the appearance of the sample. This is due to the different composition of the sample.

1. <uFor turbid and mucoid samples</u: the human DNA depletion is performed as usual, centrifugation followed by bead-beating and endonuclease treatment.
2. <uTransparent samples with some mucoid debris:</u the centrifugation should be spiked and the only bead-beating and endonuclease treatment should be performed.
3. <uFor clear samples with watery appearance:</u the human DNA depletion should be avoided and the sample should just be extracted.

<!-- --

## 1) &lt;uCentrifugation:&lt;/u&gt;

- Vortex the sample
- Centrifuge 1,200g for 10 min
- Aliquot 500ul of the supernatant into an Eppendorf tube
- Vortex the TMV aliquot
- Spike 1ul of TMV to the sample and 0.5ul if processing a pleural fluid.
- Vortex the sample with the TMV
- Proceed with bead beating host-depletion step (2.)

## 2) &lt;uBead beating and microbial extraction&lt;/u&gt;:

- Add 500ul of the supernatant post centrifugation or 500ul of whole sample to the Matrix Lysing D tube (MP biomedicals<sup>TM</sup>) after vortexing for 1 minute.
- Place the matrix tube into the TissueLyser LT (Qiagen). Bead-beat at 50 osc for 3min. Note: sometimes the sample will become foamy after bead-beating, this is ok.
- Transfer 200ul of the bead-beaten sample to a new Eppendorf tube, placing the tip at the bottom of the tube to take-up the liquid instead of the foam.
- Add 10ul of HL-SAN enzyme (ArticZymes Technology).
- Mix for 10' at 37°C at 1000rpm in the Eppendorf ThermoMixer (Eppendorf<sup>TM</sup>).

<u**NOTE:**</u> The HL-SAN is stopped by the Magnapure extraction when the instrument adds proteinase k. However, not all the extraction instruments/kits add proteinase k. If the extraction instrument/kit does not have this step, add the proteinase K manually as followed:

-Add 10ul of proteinase K to 200ul of the samples previously treated with HL-SAN.

-Vortex for 10 seconds.

-Incubate at 56°C for 10 min at 1000rpm.

- Transfer 200ul to Magnapure 24 (Roche<sup>TM</sup>) for extraction or proceed using alternative extraction methods validated in your own laboratory.
- Use total nucleic acid extraction kit 1.1 with pre-set BAL sample parameters and 50ul elution volume. Fast pathogen 200 1.1 is used for processing <8 samples and Pathogen 200 3.2 for ≥8 samples. Settings for other extraction robots will need to be established in user laboratories.

### **Microbial nucleic acid processing and PCR amplification**

#### 1. <uRT-dsDNA:</u

- Transfer 16ul of the 50ul extract (from Magnapure) to a PCR tube and add 4ul LunaScript® RT SuperMix. Mix by flicking the tube and briefly spinning down in a small bench top Eppendorf centrifuge
- <u</u Place the tube in the thermocycler and incubate at 25°C for 2 minutes, 55°C for 10 minutes, 95°C for 1 minute
- Prepare the master mix for the double strand synthesis using the Sequenase Version 2.0 DNA Polymerase

| dsDNA synthesis           | Volume x _____ samples |
|---------------------------|------------------------|
| 5× Sequenase Buffer       | 2ul                    |
| Microbial DNA-Free Water  | 7.7ul                  |
| Sequenase Dilution Buffer | 0.9ul                  |
| Enzyme                    | 0.6ul                  |
| Total volume              |                        |

- Add 11.2 ul of the master mix to the 20 ul of product post RT step. Mix by flicking the tube and briefly spin down.
- Place the tube in the thermocycler. Incubation at 37°C for 8 minutes.
- Prepare stock of 70% (700ul absolute ethanol+ 300ul of Microbial DNA-free water for 1mL) ethanol in Microbial DNA-Free Water.
- Add 45ul of resuspended AMPure XP Beads (AXP) into a new Eppendorf tube for each sample.

- Add all of the product post dsDNA synthesis (31.2ul) into the Eppendorf tube containing the beads, mix well by flicking the tube and incubate at room temperature for 5 minutes.
- Briefly spin down the samples and pellet the beads on a magnet until the eluate is clear and colourless. Keep the tubes on the magnet and discard the supernatant.
- Keep the tubes on the magnet and wash the beads with 200 µl of freshly prepared 70% ethanol without disturbing the pellet. Remove the ethanol using a pipette and discard.
- Repeat the previous step.
- Briefly spin down and place the tubes back on the magnet. Pipette off any residual ethanol with the 10p pipette. Allow to dry for 30 seconds, but do not dry the pellet to the point of cracking.
- Remove the tubes from the magnetic rack and resuspend the pellet in 12ul the Microbial DNA free water. Spin down and incubate for 2 minutes at RT.
- Pellet the beads on a magnet until the eluate is clear and colourless.
- Remove and retain 10 µl of eluate for each sample into clean 1.5 ml Eppendorf DNA LoBind tubes, individually.

**NOTE:** the sample can be store at -80°C after the RT or dsDNA step

1. <u>Library preparation Rapid PCR Barcoding Kit (SQK-RPB004) Oxford Nanopore:</u>

- Add 3 ul of the previous product (dsDNA) to a new PCR tube.
- Add 1 ul of FRM (from ONT) into the previous Eppendorf, mix by flicking the tube and briefly spin down.
- Place the tube in the thermocycler and incubate at 30° for 1 minute and then 80°C for 1 minute.
- Prepare the master mix for the PCR, adding 20ul of water and 25ul of LongAmp Taq 2X Master Mix per sample and 1ul of the barcode primers to 4 ul of the product obtained after the FRM step.
- Mix by flicking the tube and briefly spin down.

- Place in the thermocycler. Incubate @ 95<sup>0</sup> 3', (95<sup>0</sup> 15'', 56<sup>0</sup> 15'', 65<sup>0</sup> 4') x30, 65<sup>0</sup> 4'

1. <uPost PCR:</u

- Perform Qubit™ 1X dsDNA High Sensitivity (HS) on the PCR products as the manufacturer's instructions.

| Sample number | Barcode | Qbit concentration (ng/uL) | Volume (uL) to be added to the pool |
|---------------|---------|----------------------------|-------------------------------------|
|               |         |                            |                                     |
|               |         |                            |                                     |
|               |         |                            |                                     |
|               |         |                            |                                     |

<u</u Pool samples as follows:

Based on the qubit readings, prioritize samples with a qubit reading below 1 ng/µL. For these samples, normalize the qubit concentration of your PCR products to 100 ng/µL.

n=sample

e.g. 100/(n)0.6= 166.67 add the total volume of PCR product.

<u</u Clean up 1:1: Add the same volume of resuspended AMPure XP Beads (AXP) as the pool volume to the pool sample

- Mix well by flicking the tube, incubate for 5 minutes in the Hula mixer (rotor mixer) at room temperature
- Briefly spin down the samples and pellet the beads on a magnet until the eluate is clear and colourless. Keep the tubes on the magnet and discard the supernatant.
- Keep the tubes on the magnet and wash the beads with 200 µl of freshly prepared 70% ethanol without disturbing the pellet. Remove the ethanol using a pipette and discard.
- Repeat the previous step.
- Briefly spin down and place the tubes back on the magnet. Pipette off any residual ethanol with the 10p pipette. Allow to dry for 30 seconds, but do not dry the pellet to the point of cracking.
- Remove the tubes from the magnetic rack and resuspend the pellet in 15 µl Microbial DNA-Free Trish buffer. Spin down and incubate for 2 minutes at RT.
- Pellet the beads on a magnet until the eluate is clear and colourless.

- Remove and retain 10 µl of eluate into clean 1.5 ml Eppendorf DNA LoBind tubes, individually.
  - Add RAP 1ul to 10ul of template, mix gently by flicking the tube, spin down and incubate at RT for 5 minutes.
  - Add 25.5ul of LB
  - Add 34 of SQB
  - Add 4.5ul water
  - Mix gently pipetting up and down
1. <uFlowcell loading:</u
- Thaw FLB and FLT at RT if not thawed from before
  - Add 30ul of FLT (vortex before by pipetting) to a FLB (vortex before by pipetting).
  - Mix by vortex.
  - Open the GridION device lid and slide the flow cell under the clip. Press down firmly on the flow cell to ensure correct thermal and electrical contact.
  - Slide the flow cell priming port cover clockwise to open the priming port.
  - After opening the priming port, check for a small air bubble under the cover. Draw back a small volume to remove any bubbles: 1.
  - Set a P1000 pipette to 200 µl 2. Insert the tip into the priming port 3.
  - Turn the wheel until the dial shows 220-230 ul, to draw back 20-30 ul, or until you can see a small volume of buffer entering the pipette tip Note: Visually check that there is continuous buffer from the priming port across the sensor array.
  - Load 800 µl of the priming mix into the flow cell via the priming port, avoiding the introduction of air bubbles. Wait for five minutes.

- Gently lift the SpotON sample port cover to make the SpotON sample port accessible.
- Load 200 µl of the priming mix into the flow cell priming port (not the SpotON sample port), avoiding the introduction of air bubbles.
- Mix the prepared library gently by pipetting up and down just prior to loading.
- Add 75 µl of the prepared library to the flow cell via the SpotON sample port in a dropwise fashion. Ensure each drop flows into the port before adding the next.
- Gently replace the SpotON sample port cover, making sure the bung enters the SpotON port, close the priming port and replace the MinION or GridION device lid.

1. <uGridION settings:</u

1. Select the position of your flow cell on MinNOW.
2. Type your experiment name and sample ID.
3. Continue to kit selection and select Rapid PCR barcoding kit SQK -RPB004.
4. Continue to run options and change to 24hr.
5. Continue to analysis and edit options and select barcode both ends and mid read barcode filtering.
6. Continue to output and unselect fast 5 and edit FastQ to 100 reads per file.

1. <uBioinformatic pipeline and reporting</u

Curated bioinformatic pipeline will generate automatic reports at 30 minutes (for organism identification) and 2 hours (for AMR determinants). 16-24 hours reports can be generated manually for SNP typing and genomic alignment.

Bioinformatic pipelines used for this SOP is based on the published pipeline (Charalampous T et al **Am J Resp Crit Care Med.** 2024 209:164-174). Further details available on request.

## Appendix 1.

## qPCR for the TMV

The RT step is performed using the LunaScript master mix explained previously in the section 3.

### -Master mix:

| <b>**Reagent **</b>   | <b>**Volumen **</b> | X__samples |
|-----------------------|---------------------|------------|
| Fast Syber-green      | 10ul                |            |
| Primer Forward (10uM) | 1ul                 |            |
| Primer Reverse (10uM) | 1ul                 |            |
| DNAse-free water      | 3ul                 |            |
| Template (cDNA)       | 5ul                 |            |
| <b>Total</b>          |                     |            |

1 cycle at 95°C for 20 s, 40 cycles at 95°C for 1 s, 40 cycles at 60°C for 20 s, 1 cycle at 95°C for 15 s, 1 cycle at 60°C for 1 min, and 1 cycle at 95°C for 15 s

Ellis MD, Hoak JM, Ellis BW, Brown JA, Sit TL, Wilkinson CA, Reed TD, Welbaum GE. Quantitative Real-Time PCR Analysis of Individual Flue-Cured Tobacco Seeds and Seedlings Reveals Seed Transmission of Tobacco Mosaic Virus. *Phytopathology*. 2020 Jan;110(1):194-205. doi: 10.1094/PHYTO-06-19-0201-FI. Epub 2019 Nov 18. PMID: 31502520.

## **Appendix 2. Sample processing sheet.**

| <b>Reagent</b>                                     | <b>Company</b>                 | <b>Lot</b> | <b>Expiry date/received date</b> | <b>Extraction day</b> |
|--|--------------------------------|------------|----------------------------------|-----------------------|
| Lysis Matrix D 2mL Tube                            | MP Biomedicals                 |            |                                  |                       |
| HL-SAN enzyme                                      | ArcticZymes Technolo           |            |                                  |                       |
| LunaScript® RT SuperMix Kit (E3010)                | New England Biolabs            |            |                                  |                       |
| Sequenase Version 2.0 DNA Polymerase               | ThermoFisher Scientific        |            |                                  |                       |
| LongAmp® Taq 2X Master Mix                         | New England Biolabs            |            |                                  |                       |
| Rapid barcoding Kit SQK-RPB004                     | Oxford Nanopore<br>Thechnology |            |                                  |                       |
| AMPure XP Beads                                    | BeckmanCoulter                 |            |                                  |                       |
| Qubit™ 1X dsDNA High Sensitivity (HS)<br>(Q33230)  | ThermoFisher Scientific        |            |                                  |                       |
| NATtrol™ Respiratory Panel 2.1 (RP2.1)<br>Controls | Zeptometrix                    |            |                                  |                       |
| Viasure respiratory panel III (PCR for PC)         | Pro-labs                       |            |                                  |                       |
| Tobacco mosaic virus (PC-0107)                     | DSMZ                           |            |                                  |                       |

| Reagent  | Company                 | Lot | Expiry date/received date | Extraction day |
|--|-------------------------|-----|---------------------------|----------------|
| TMV primers:<br><br>TMV_fwd_aps<br><br>-Catalog number: 10336022 | Invitrogen              |     |                           |                |
| GGATATGTCTAAGTCTGTTGC<br><br>10629186 Nucleotides (25 nmole)     |                         |     |                           |                |
| 11732013 Desalted<br>574910 W7149<br>(E12)                       |                         |     |                           |                |
| TMV_fwd_aps<br><br>574910 W7149 (F01)                            |                         |     |                           |                |
| 3 TMV_rvr_aps  |                         |     |                           |                |
| TMV_rvr_aps<br><br>-Catalog number: 10336022                     |                         |     |                           |                |
| CAGACAACTCGGGTGCG<br><br>10629186 Nucleotides (25 nmole)         |                         |     |                           |                |
| 11732013 Desalted  |                         |     |                           |                |
| Fast SYBR™ Green Master Mix                                      | Applied Biosystem       |     |                           |                |
| Microbial DNA-Free Water (ID: 338132)<br>Qiagen                  | ThermoFisher Scientific |     |                           |                |
| RESPIRATORY SWAB MATRIX<br>NEGATIVE CONTROL                      | Vircell                 |     |                           |                |
| Ethanol  |                         |     |                           |                |
| Elution buffer   | ONT                     |     |                           |                |

| Equipment                     | Company                 |
|-------------------------------|-------------------------|
| TissueLyser LT                | Qiagen                  |
| Thermocycler for the lib prep |                         |
| Real-time PCR for TMV and PC  | Quantstudio 7           |
| Eppendorf ThermoMixer         | Eppendorf               |
| Qubit 4 Fluorometer           | ThermoFisher Scientific |

## 3. Bioinformatics

---

### 3.1 Clinical metagenomics bioinformatics

---

#### 3.1.1 Overview

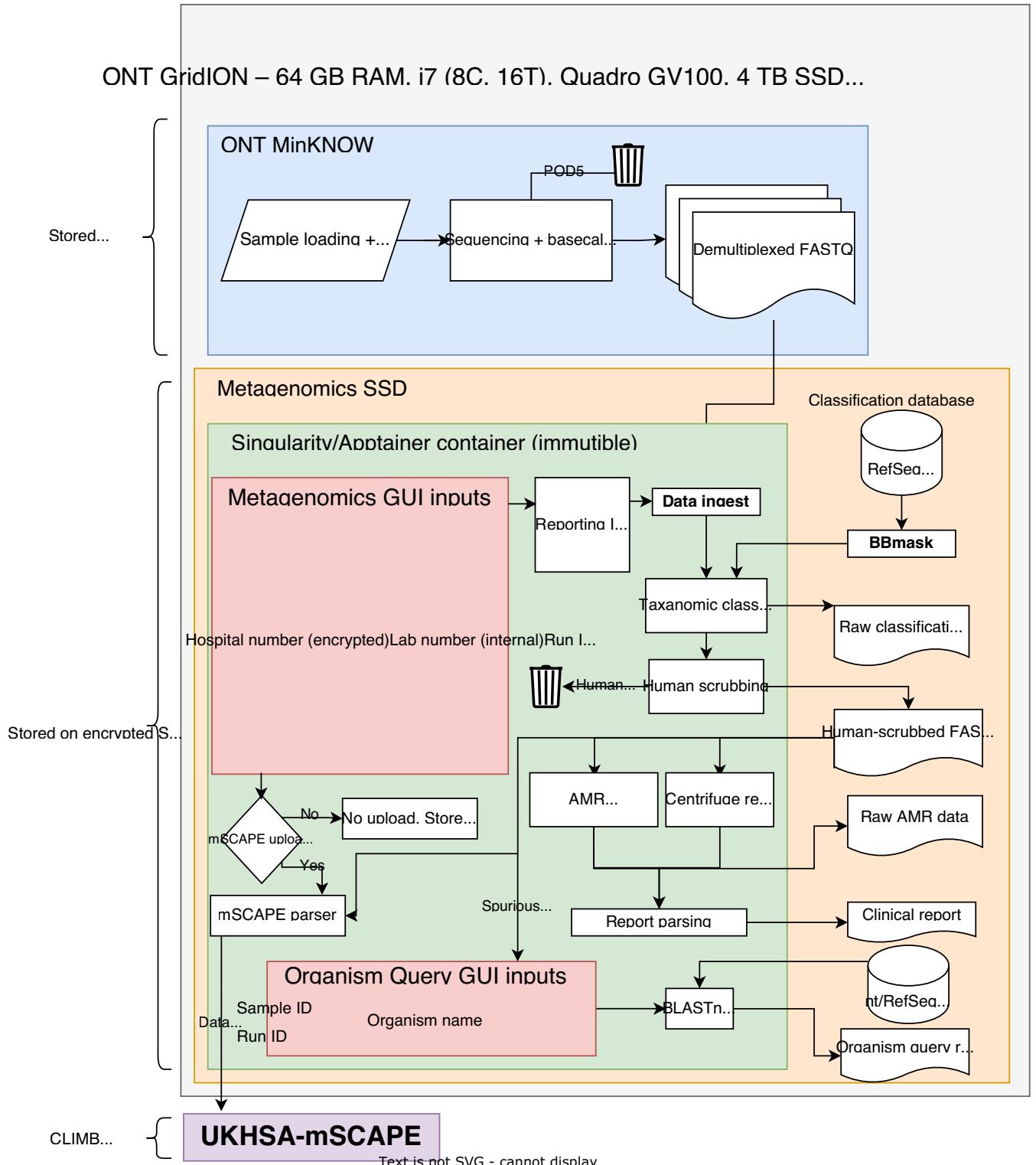
The principal output of the CIDR Metagenomics workflow is a PDF report listing organisms with detectable nucleic acids (RNA/DNA) and some additional information on AMR associated sequence data. The solution packages two applications - CIDR Metagenomics Workflow and [Organism Query](#) alongside a few scripts to help manage and analyse outputs. The Metagenomics Workflow runs ontop of MinNOW, analysing sequencing data in real time producing easily digested report. [Organism Query](#) can be used to scrutinise classifications contained within a report. It leverages the full NCBI nt and RefSeq databases producing a report similar to NCBI BLAST in ~15 minutes. The Organism Query report is designed to provide the user with appropriate information to scrutinise a significant taxonomic classification.

#### mSCAPE

Users can opt in to [mSCAPE](#) on an per-experiment basis for an automatic upload of sequencing data to UKHSA mSCAPE.

#### Technical facets

After loading a metagenomic library on to an ONT sequencing device and launching the sequencing experiment in ONT MinNOW the pipeline is initialised by the user through the Metagenomics Launcher graphical user interface (GUI). The software periodically ingests base called FASTQ data from the GridION [/data/](#) directory at set intervals - 0.5, 1, 2, 16 and 24 hours. At each interval, the pipeline performs human scrubbing, taxonomic classification, AMR detection and MLST which is then consolidated in to a PDF reports which are saved in the [/media/grid/metagenomics/results/](#) directory. The diagram below illustrates further details of how the pipeline works:



### Taxonomic classification

At its core, the pipeline leverages a taxonomic classification tool called [Centrifuge](#). There are numerous alternatives, the most common being Kraken. We chose centrifuge namely because of its smaller memory footprint and existing deployment in ONT's WIMP. Each read in the raw data is aligned against an index, and is

assigned a confidence score and a taxonomy. Our index is an optimised database of curated eukaryotic, prokaryotic, and viral reference sequences formed primarily from [NCBI RefSeq](#) and [FDA-ARGOS](#) databases. This is provided on the SSD sent to each site.

The index was built from the following [sequences](#) - this map file contains the accession number for each sequence in the database and the accompanying taxa ID. This file is required to assemble the Centrifuge index alongside the sequences in FASTA format. Before building the index, the sequences were masked using [BBmask](#). This tool is applied primarily to prevent false-positive matches in highly-conserved or low-complexity regions of genomes.

### 3.1.2 Related code snippets

---

#### Masking a FASTA database using BBmask

```
bbmask.sh in=unmasked.fasta out=masked.fasta entropy=0.7 -Xmx80g maskrepeats=t
```

#### Building a centrifuge index -

```
# --bmax needs tuning based on available memory. centrifuge-build -p 10 --conversion-table accession2taxid.map --taxonomy-tree .
```

## 3.2 Setting up CIDR Metagenomics bioinformatics workflow

---

### 3.2.1 Overview

Each Network site will receive an ONT GridION sequencing platform and an external SSD containing the software and databases required for analysing metagenomic datasets. The software has been designed such that it will be easy for anybody to set up and use. Follow the instructions below to install the bioinformatics workflow.

### 3.2.2 Install instructions

1. Insert the USB SSD in to one of the blue USB ports at the rear of the GridION. Try to place the disk away from the warm exhaust as this may lead to overheating.
2. After logging in to the GridION Ubuntu operating system, modify the file browser setting by following the video below. This is to enable the running of scripts without using the terminal.
  1. Using the file browser, on the taskbar on the left side of the screen, navigate to the **metagenomics** disk, which can be found in the navigation pane inside the file browser.

#### Info

As a security feature, the removable SSD has been encrypted. Enter the encryption key provided to you and confirm that you'd like the key remembered.

1. Navigate to the `metagenomics` disk in the file explorer and double-click `launch_installer.sh`, selecting to 'Run in terminal'. When prompted to do so, type the password for the GridION device (not the encryption key). See below for a video guide.

#### Info

As you type, no lettering or symbols will appear. This is normal. If you mistype, press enter and try again.

There may be additional outputs in your terminal window compared to the video.

1. Some icons should appear on the desktop linking to each app. You will need right click on the icons and select `Allow launching` before continuing with the next stage.

Success!

We have now installed the CIDR metagenomics workflow. The next step will be to run through a control dataset to test the workflow has run sucessfully.

### 3.2.3 Install validation

---

Included with the software is a small dataset based on the Zymo community standard. In this step we will validate the function of the workflow with this dataset and generate a report.

1. Double click the Metagenomics Launcher icon on the desktop.
2. Fill out the fields, as indicated in the video below. More information on how to fill the fields and run the launcher can be found in the [Starting the metagenomics workflow](#) section.
1. Wait ~10 minutes for the workflow to complete. Open up the PDF report which can be found in the `reports` folder on the metagenomics disk in a folder corresponding to the name of the sample provided in the launcher eg. `gstt_control_1`. See video below for further information.
1. Inspect the `/metagenomics/reports/validation_sample/` PDF report at the **0.5 hr timepoint**, it should match the CIDR validation report provided [here](#).

Success!

We have now tested the CIDR metagenomics workflow. The next step will be to run a sequencing experiment, running the workflow in real-time.

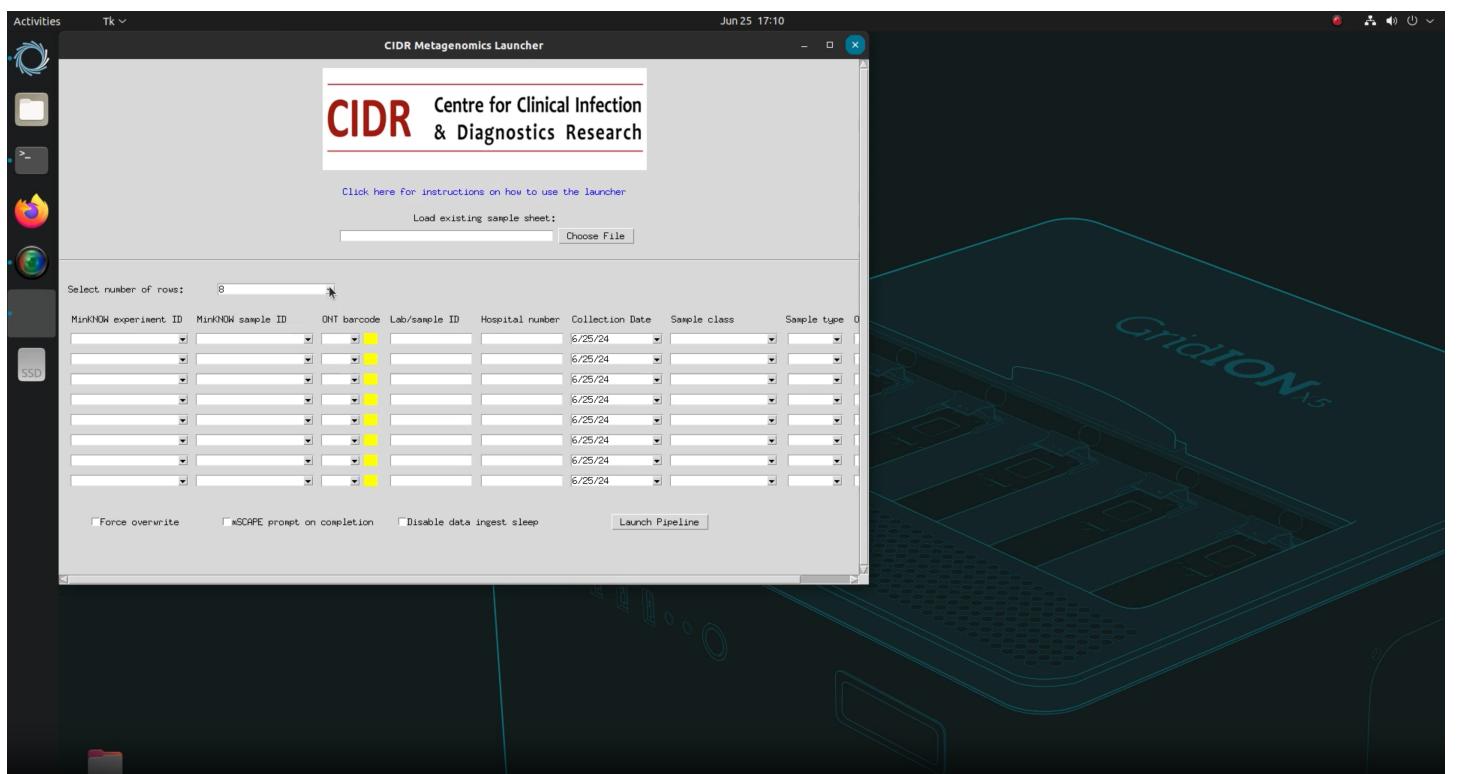
## 3.3 Running the metagenomics workflow

Before starting

1. The CIDR metagenomics workflow must be started during a sequencing experiment or after a sequencing experiment has completed. The pipeline must not be activated before a sequencing experiment has started in MinNOW and has **started producing reads** (See MinNOW setup - Lab Protocol).
2. Ensure the SSD is inserted in to one of the rear USB 3.1 ports, has been mounted and the encryption key has been entered successfully. Test the disk has been mounted by navigating to it in the Ubuntu file explorer.

### Starting a run

1. Double click the **Metagenomics Launcher** icon on the GridION desktop, the CIDR Metagenomics Launcher should appear alongside a terminal window.



## Known issues

The `'geocryptfs error not found...'` error can be ignored as it is not essential to the workflow.

1. Select the number of samples to be analysed from the dropdown.
2. You can choose to initiate the launcher using **one** of the below methods:
  - Fill out the fields on the form for each sample to be analysed.
  - Loading a pre-existing TSV - [see example](#).

## Field descriptions:

| Field                        | Description  |
|------------------------------|--|
| <b>MinKNOW experiment ID</b> | The exact name matching the experiment name on MinKNOW entered by the user when initiating a sequencing run. This is populated automatically from the <code>/data</code> directory.              |
| <b>KinKNOW sample ID</b>     | The exact name matching the Sample name on MinKNOW entered by the user when initiating a sequencing run. This is populated automatically from the <code>/data/{experiment_id}/</code> directory. |
| <b>ONT barcode</b>           | The ONT library index/barcode used. Green colour indicates the barcode directory has been validated.   |
| <b>Lab/Sample ID</b>         | The unique lab accession number for the sample. This data is encrypted before transmission.  |
| <b>Hospital number</b>       | Hospital number corresponding to the sample. Can be anonymised. This is encrypted before transmission.   |
| <b>Collection date</b>       | Collection data of the sample.   |
| <b>Sample Class</b>          | The class of sample loaded.  |
| <b>Sample type</b>           | The methodology used to collect the sample.  |
| <b>Operator</b>              | Arbitrary identifier of the user operating the sequencer.  |

## Note

- Option 1 will generate a sample sheet stored in the `metagenomics/sample_sheets` directory. This can be reused if a repeat run is required - or quick edits need to be made to a set of samples without having to fill out the fields again.
- For full information on data encryption protocol visit the [mSCAPE uploader](#) page.
- Test

1. With the metadata form filled, select the run parameter check boxes.

| Parameter                  | Description  |
|----------------------------|--|
| <b>Force overwrite</b>     | The exact name matching the experiment name on MinNOW entered by the user when initiating a sequencing run. This is populated automatically from the <code>/data</code> directory. |
| <b>mSCAPE prompt</b>       | After the sequencing and analysis run has completed, open the mSCAPE uploader for user input. No data is uploaded without par-sample expressed authorisation.                      |
| <b>Disable data ingest</b> | <b>Not for real-time analysis!</b> Analyse all data immediately - do not wait for it to be generated by the sequencer.   |
| <b>sleep</b>               |  |

## Known issues

You should wait to launch the pipeline after the sequencer has reported producing reads in MinNOW, the workflow will display errors in red if no reads have been found.

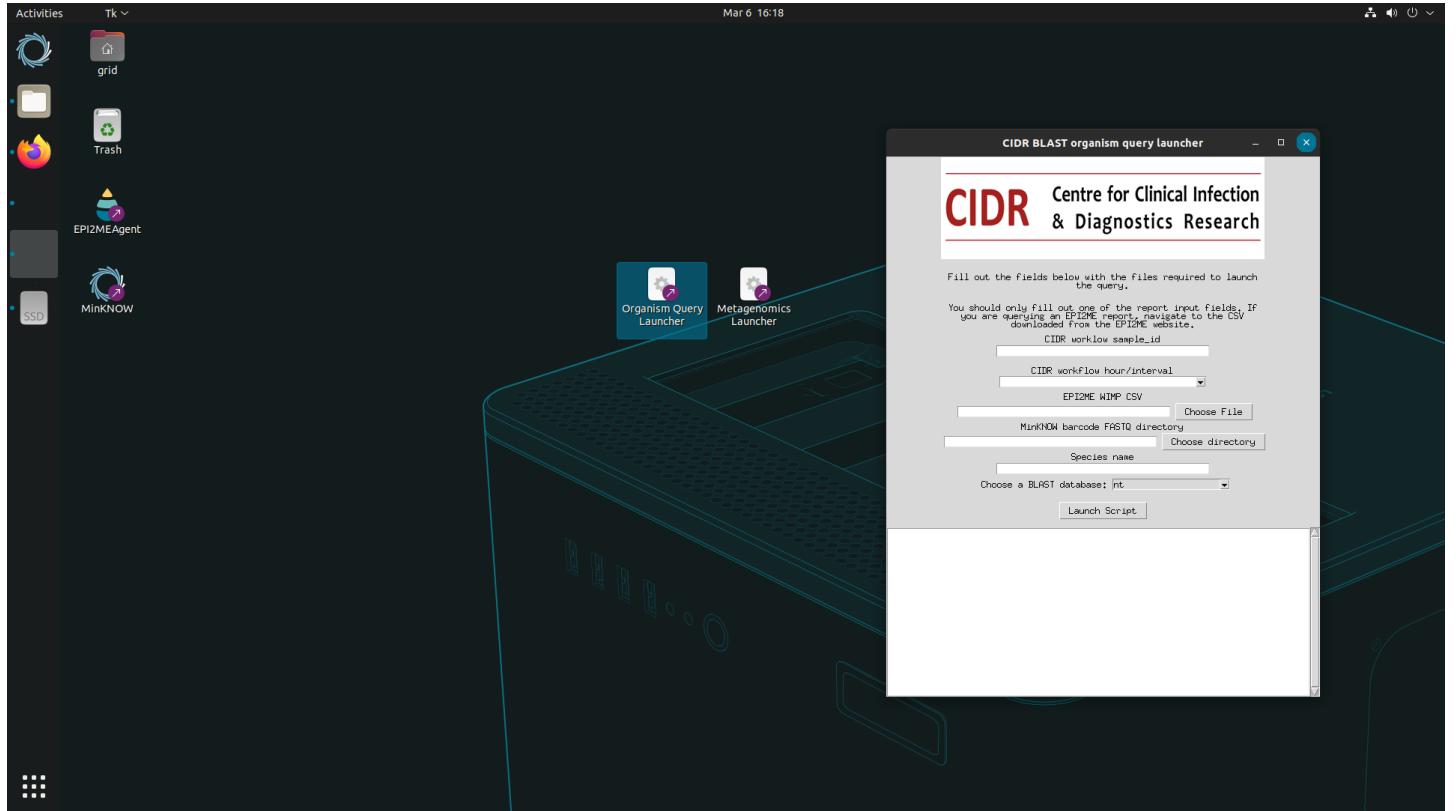
You can stop the analysis or close the Launcher window at any point by typing `CTRL+C` while the terminal window is active and closing the launcher window.

1. Click on `Launch pipeline` and click `OK` to start analysis.
2. After a minute, the terminal window accompanying the workflow launcher should start displaying log outputs from the workflow. See below for an example.
1. ~40 minutes after launching the sequencing experiment alongside the metagenomics workflow, the first reports will be available in `/media/grid/metagenomics/reports/{sample_name}/{timepoint}`. See below for a guide on how to access this.

## Success!

We have now run the CIDR metagenomics workflow. The workflow will run for ~24 hours.

## 3.4 Bioinformatics - Running metagenomics



## 3.5 Defaults

```
singularity exec --bind /tmp/.X11-unix:/tmp/.X11-unix --env DISPLAY=$DISPLAY organism_query_v1.3.sif /bin/bash -c 'cd /organism_query/ ;source /opt/conda/etc/profile.d/conda.sh && conda activate organism_query && python launcher_defaults.py'
```

## 3.6 Setting up CIDR Metagenomics bioinformatics workflow - alternative deployment

### Note

The tools and documentation described here and on the CIDR GitHub are not validated for diagnostic use and are for research and evaluation purposes only.

### 3.6.1 Overview

For collaborators outside of the Network, an alternative configuration can be provided. This will bypass the GUI allowing users to provide a `sample_sheet.csv` through a CLI. Organism query will not be available to headless users as this tools is heavily reliant on GUI I/O.

### 3.6.2 Install instructions

1. Decompress CIDR\_metagenomics\_vX.X.tar.gz:

```
tar -xvzf CIDR_metagenomics_vX.X.tar.gz
```

1. Install conda/mamba.
2. Build the appropriate environment for running the CIDR metagenomics containers.

```
wget https://raw.githubusercontent.com/GSTT-CIDR/metagenomics_container/main/conda/apptainer.yml conda env create -f apptainer.yml
```

1. Allocate a directory for MinKNOW data outputs. This will be mounted to the `/data` directory in the container in a later step.

## Note

The directory structure of data for ingest must be maintained as in standard MinNOW outputs eg.

...

## **3.7 Example for control sample**

---

[minknow\_outputs\_directory]/GSTT\_control\_sample\_01/  
GSTT\_control\_sample\_01/20240424\_1408\_X4\_FAY88387\_d3868a4f/fastq\_pass/barcode11

## **3.8 Naming schema**

---

[minknow\_outputs\_directory]/[experiemnt]/[sample\_id]/[\*]/fastq\_pass/barcodeXX

...

### **3.8.1 Install validation**

---

1. Navigate to the root of the `CIDR_metagenomics_vX.X` directory.
2. Move `CIDR_metagenomics_vX.X/GSTT_control_sample_XX` to the allocated directory for MinNOW data outputs (from Install instructions: Step 4).
3. activate the apptainer conda environment: `conda activate apptainer`
4. Initiate the run for analysing the control dataset:

```
apptainer exec --bind .:/mnt --bind ./data:/data ./containers/cidr_metagenomics_v3.6.sif bash -c 'cd /workflow ; source /opt/con
```

1. When the workflow has completed, inspect the `CIDR_metagenomics_vX.X/reports/CIDR_control_1` PDF report, it should match the CIDR validation report provided [here](#).

## Info

Variables to change in step 3

**--bind ./:/mnt** - Binding the workflow root directory to the container /mnt.

**--bind ./data:/data** - binding the allocated directory for MinKNOW data outputs to /data.

**./containers/cidr\_metagenomics\_v3.6.sif** - launching the metagenomics container.

**for t in 0.5 1 2 16 24** - time-points for analysis.

**--cores 20** - number of samples to be processed simultaneously - not the same as threads.

**samples=/mnt/sample\_sheets/CIDR\_control\_1.csv** - the mounted path for the sample sheet - remember this is the relative mounted path, so **/mnt/sample\_sheets** corresponds to

**CIDR\_metagenomics\_vX.X/sample\_sheets** on the host machine.

## 3.8.2 Implementation

1. Build a **sample sheet** copying the structure of the example in **CIDR\_metagenomics\_vX.X/sample\_sheets**. Importantly, 'Experiment', 'SampleID' and 'Barcode' must be correct and correspond to the

**[minknow\_outputs\_directory]/[experiemnt]/[sample\_id]/[\*]/fastq\_pass/barcodeXX** scheme.

2. activate the apptainer conda environment: **conda activate apptainer**

3. Run the container, changing the flags explained in the validation step:

```
apptainer exec --bind ./:/mnt --bind ./data:/data ./containers/cidr_metagenomics_v3.6.sif bash -c 'cd /workflow ; source /opt/cor
```

1. PDF outputs should be found in **CIDR\_metagenomics\_vX.X/reports/** corresponding to each LabID in the **sample sheet** loaded.

## 4. Analysis

---

### 4.1 SOPs for clinical interpretation - coming soon

---

Note

The tools and documentation described here and on the CIDR GitHub are not validated for diagnostic use and are for research and evaluation purposes only.