

SEP 29, 2023

OPEN ACCESS



DOI:
dx.doi.org/10.17504/protocols.io.5jyl8pe16g2w/v1

Document Citation: Alex Shaw, Joyce Akello, Catherine Troman, Aine.OToole, Erika Bujaki, c.ansley, Zoe Vance, rachel.colquhoun, Andrew Rambaut, Javier Martin, Nick Grassly 2023. Quality Control and Data Recording for DDNS. **protocols.io**
<https://dx.doi.org/10.17504/protocols.io.5jyl8pe16g2w/v1>

MANUSCRIPT CITATION:

License: This is an open access document distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

Created: Sep 26, 2023

🌐 Quality Control and Data Recording for DDNS

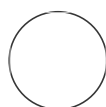
Alex Shaw¹, Joyce Akello¹, Catherine Troman¹, Aine.OToole², Erika

Bujaki³, c.ansley², Zoe Vance², rachel.colquhoun², Andrew Rambaut², Javier Martin³, Nick Grassly¹

¹Imperial College London; ²University of Edinburgh;

³Medicines and Healthcare products Regulatory Agency

Poliovirus Sequencing Consortium



Joyce Akello

ABSTRACT

This standard operating procedure indicates how to perform quality control checks for DDNS and provides guidance on best practice for ensuring relevant sequencing run data are recorded. The document is structured into quality control steps for setting up data files for DDNS, the RT-PCR and library preparation, and the sequencing run.

Last Modified: Sep 29,
2023

DOCUMENT integer ID:
88404

Quality Control and Data Recording for DDNS

Principle of the procedure

For DDNS results to be valid and suitable for reporting, sufficient sample metadata must be recorded and data integrity maintained throughout the planning stage of the experiment, during the experiment and after the experiment. Quality control checks have been included to ensure that the protocol has been performed correctly and that results are valid. Later comparison of the DDNS results with culture-based poliovirus detections can be facilitated by adding further metadata describing the timelines and results for processing of the same samples by cell-culture.

This SOP is accompanied by an example barcodes.csv file “`barcodes.csv`”. In addition to the essential data (barcode, sample), additional metadata should be added as it is available prior to the sequencing run. Do not use any special characters or spaces in this data or in the column headers.

When running PIRANHA, the `barcodes.csv` file is supplied to the software in addition to the location of the fastq files.

After the PIRANHA analysis is complete, the `barcodes.csv` file will be written into the “`detailed_run_report.csv`” and any additional metadata can be added to the “`detailed_run_report.csv`” file. This final report is the definitive document containing all the data for the sequencing run, and can be uploaded for storage and data rows shared when reporting detections.

And overview of the procedure is shown in Figure 1.

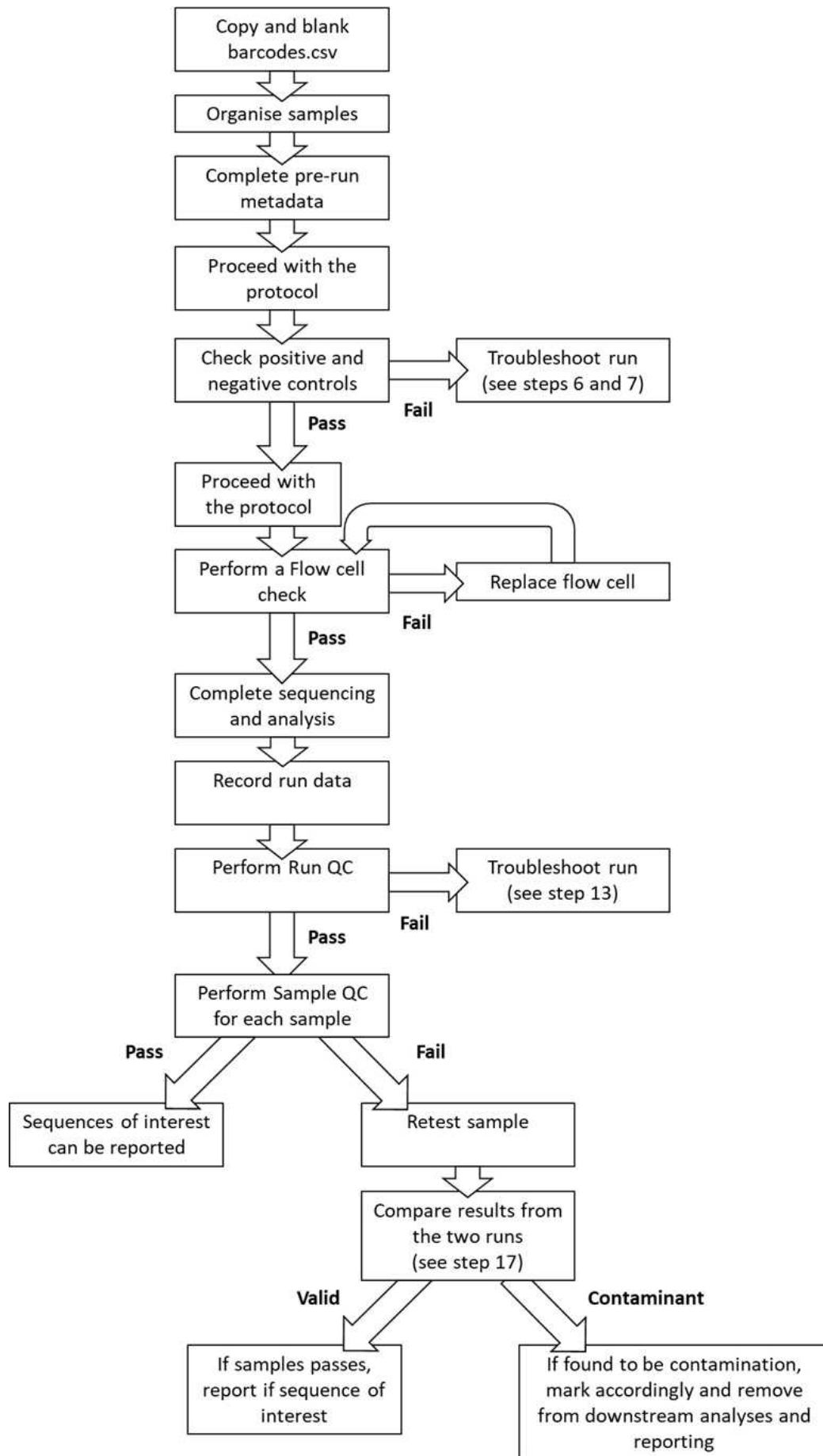


Figure 1. Overview of the data entry and quality control procedure for DDNS. “Steps” refer to the procedure steps below.

Personnel

All procedures can be performed by suitably trained members of staff

Additional documents

- DDNS protocol v2.0 ([DDNS protocol v \(protocols.io\)](#))
- barcodes.csv

And, once the DDNS sequencing run has been performed

- detailed_run_report.csv

Procedure

Planning a DDNS experiment

1. Make a copy of the barcodes.csv template
2. Organise the samples; pairs of samples (with the same EPID) can have consecutive barcodes but try not to group samples from the same geographic area together. This helps detect any potential cross-contamination because identical sequences are then unlikely to be detected in samples with consecutive barcodes that are adjacent to one another on the 96 well plate.
3. Insert controls into your sample list. A positive control (resuspended Coxsackievirus A20 provided by NIBSC) and negative control (water) should each be included on the first and last RNA extraction batches of the day.

For details on the use of the positive control virus, refer to the Guidance for DDNS Run Control SOP.

[protocol \(protocols.io\)](#)

4. Enter the data required for planning the run and any metadata available.
 - a. Complete the essential data “barcode” and “sample” in the first two columns
 - b. Mark any positive controls in the sample column as “positive” and negative controls as “negative”
 - c. Any additional metadata should be added to columns 3 to 23 (e.g. EPID, specimen ID, date variables etc).
 - d. If any included samples are repeats due to QC failure on previous experiments, mark “Yes” in column “QCCheck”.
 - e. If there has been a delay in the processing of the sample e.g. due to a lack of extraction kits or software updates preventing the run, note “Yes” in the column “DelaysInProcessingForDDNS” and enter the type of delay in

the column "DetailsOfDelays."

NB: Do not include any patient' names in any data column

RT-PCR and library preparation Quality Control steps

5. Enter details in columns 23 to 28 during sample amplification and library preparation. After the RT-PCR reaction (step 5

in the DDNS protocol), note any failed RT-PCR reactions (e.g. where the sample has evaporated) in "RT_PCR_comments". The RT-PCR for these samples will need to be repeated on a subsequent DDNS experiment

6. After the nested PCR reaction (step 6 in the DDNS protocol), note any failed reactions (e.g. where the sample has evaporated) in "PCR_comments". The RT-PCR and nested PCR for these samples will need to be repeated on a subsequent DDNS experiment.

7. Complete columns "PositiveControlCheck" and "NegativeControlCheck" after the running the nested VP1 positive and negative controls on a gel or tapestation (step 11 of the DDNS protocol).

a. All samples can be marked as "Pass" for the PositiveControlCheck if all positive controls extracted on the same day show a VP1 band on a gel or tapestation run (Figure 2). If a band is missing for any positive control, mark all samples extracted on that day as "Fail".

b. All samples can be marked as "Pass" for the NegativeControlCheck if all negative controls extracted on the same day show no VP1 band on a gel or tapestation run (Figure 2). If a band is present for any negative control, mark all samples extracted on that day as "Fail".

8. If the positive control check is failed, run the positive control panEV product(s) on a gel or tapestation.

a. If there is no panEV band (Figure 2), repeat the nested VP1 reaction for the control. If a band is visible, discard the VP1 amplicons and repeat the VP1 reactions for all samples.

b. If there is no band visible after repeating the nested VP1 reaction, repeat the RNA extractions after checking the RNA extraction kit is being used correctly and has not expired.

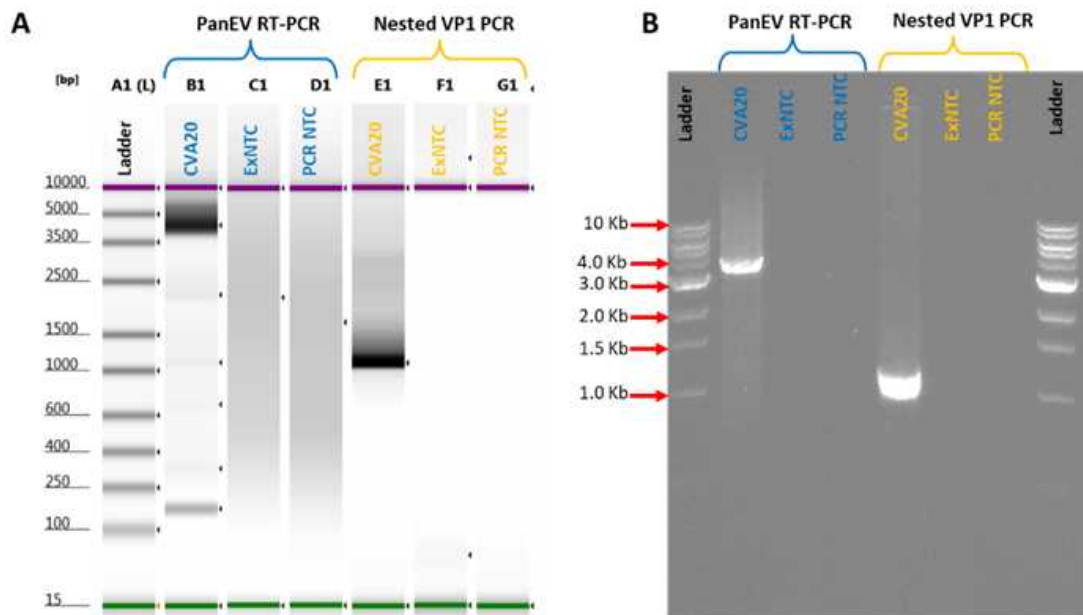


Figure 2. The Tapestation results for the DDNS assay controls from PanEV RT-PCR and nested VP1 PCR products (Panel A) and the gel electrophoresis results for the DDNS assay controls from PanEV RT-PCR and nested VP1 PCR products (Panel B). The positive extraction control labelled as CVA20 with a PanEV RT-PCR band at approx 4.2kb and nested VP1 PCR band at approx. 1.2kb. The negative extraction control as ExNTC with no band in both the PanEV RT-PCR and nested VP1 PCR. The PCR no template control as PCR NTC with no band in both the PanEV RT-PCR and nested VP1 PCR.

9. If the negative control check is failed, repeat the panEV reaction and the nested VP1 reaction.
 - a. If the negative control still shows a band on a gel or tapestation:
 - i. Thoroughly clean the PCR and RNA extraction workstations.
 - ii. Replace each of the nested VP1 reagents in turn whilst performing blank reactions to determine a contaminated reagent.
 - iii. Perform an additional Negative RNA extraction to confirm that that RNA extraction kit is not contaminated.
10. Once both Positive and Negative control checks are passed, proceed with the DDNS protocol.
11. Complete columns "RunNumber", "DateSeqRunLoaded", "FlowCellID", "FlowCellUses", "PoresAvailableAtFlowCellCheck" after performing a Flow Cell Check during the sequencing library preparation (prior to step 44 of the DDNS protocol). Note the intended run duration in "RunHoursDuration".
 - a. The FlowCell ID is written on the flow cell
 - b. Under "FlowCellUses" write the number of times the flow cell has been used (e.g. if it is a new flow cell write "1", if it has already been used for a single DDNS run write "2")
 - c. The number of pores available should be determined via a Flow Cell Check run in MinKNOW prior to library loading. If a flow cell has <700 pores do not use it for a 96 sample DDNS run.
 - d. For a run of 96 samples for routine testing, the flow cell should be run for 4 hours.

12. Run the PIRANHA analytical software. You need to provide the location of your barcodes.csv file with the added metadata and the location of your sequencing data (the demultiplexed fastq files). The sequencing results will be appended to your barcodes.csv file and the finished report written to the chosen output folder. When running PIRANHA for stool samples set the following options:

- a. Minimum read length: 1000
- b. Maximum read length: 1300
- c. Minimum depth: 50
- d. Minimum read percentage: 1

Sequencing Run Quality control checks

13. PIRANHA will have confirmed whether

- a. The positive control(s) yielded at least 500 sequences that have mapped to NPEVs or Cocksackievirus A20.
- b. The negative control(s) yielded less than 50 reads mapped to poliovirus or NPEVs

14. Confirm manually whether

- a. The sequencing run completed its full run duration (check the MinKNOW run report).
- b. The number of pores did not fall beneath 400 (or 25% of total pores) in the first hour of the sequencing run

(check

the MinKNOW run report, see Figure 3).

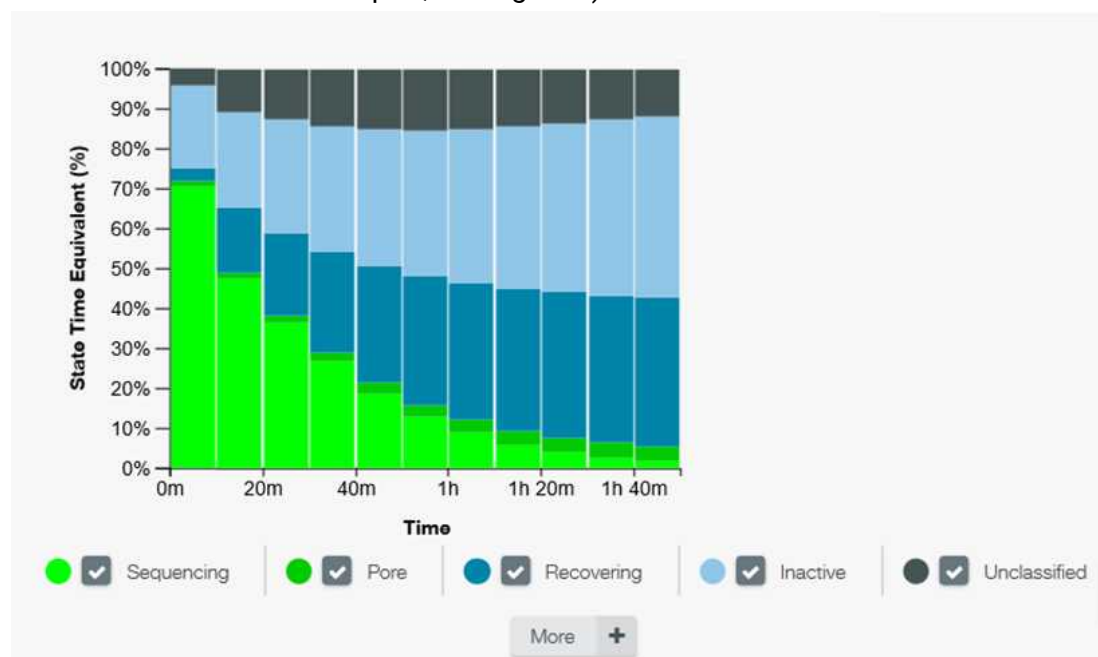


Figure 3. The MinKNOW run report showing a failed sequencing run where the percentage of available pores (sum of bright green “Sequencing” and darker green “Pore” bars) falls below 25 % of total pores within the first hour of the sequencing run.

Alter the RunQC field to “Fail” if either of these criteria are not met. Add an explanation of the failure in the “comments” field.

15. Depending on the reason for failure:

- a. Too few positive control reads:**
 - i. Repeat the library pooling and confirm the presence of your library after the clean-up steps using a Tapestation or a Qubit fluorometer.
 - ii. Check that you are ligating the correct adaptor (LA) and are using the short fragment buffer (SFB) during library preparation.
- b. Negative control has too many reads:**
 - i. Confirm that your earlier negative control check has passed QC checks.
 - ii. Rewash the flowcell with a DNase wash and repeat the library pooling and sequencing run.
- c. Run has not reached its full duration:**
 - i. Check in the system messages in MinKNOW to see if it gave a reason for stopping that you can resolve such as insufficient memory on the device.
 - ii. Restart the sequencing run. Check that there are still >700 pores available for sequencing (this will be reported when the run is restarted).
 - iii. If there are insufficient pores available, repeat the library pooling and sequencing library preparation and load into a different flow cell.
- d. Sudden reduction in pore numbers during the first hour of the run**
 - i. If the positive control has > 10,000 reads, continue with analysis (sufficient sequencing depth has likely been achieved)
 - ii. If the positive control has < 10,000 reads, repeat the library pooling and sequencing library preparation and load into a different flow cell.

Sample Quality control checks

16. If the run QC is passed, perform an alignment on the sequences generated by PIRANHA and generate a phylogenetic tree with the DDNS sequences from prior runs using Nextstrain or Geneious.

17. For samples with 50 or more reads, perform the sample quality control checks. The barcodes.csv template includes a field called "SampleQC". Samples can be marked as "Pass" if they have met the following criteria

- a.** Samples with the same EPID (i.e. from the same case) are no more than 2 nucleotides different from each other over VP1.
- b.** Samples adjacent to each other on the plate (on either axis) with different EPIDs differ by >1 nucleotide over VP1

18. Check for bleed through between runs if the flow cell has been reused; for samples with 50 to 1,000 reads, check the results of the prior sequencing run with that flow cell to see if there is a sample with the same barcode that yielded a highly similar sequence (<2 nucleotides different over VP1). If a sample matches a sequence from a previous run in this manner, mark it as a "Fail".

19. Any sample that does not meet these criteria should be marked as a "Fail" in column "SampleQC", an explanation added in column AL and the sample repeated on a later run (with column "QCCheck" flagged "Yes" in that run). These samples should not be grouped together upon retesting and should use different barcodes.

20. If on retesting:

a. Samples yield the same sequence as in the initial run, mark them on each run as “Pass”. If a pair of samples (same EPID) yield the the same sequences as the first run, but these differ by more than 2 nucleotides they can be marked as “Pass”.

b. Samples that no longer yield sequences, mark them on each run as negative “DDNSClassification” and a note of “Likely contamination” made in column “SampleQC” in the original run. These sequences can be removed from further analyses of the sequencing run and should not be submitted.

21. Any data that is not available before or during the run (e.g. Sanger sequencing results) can be added to the detailed_run_report.csv when available.

