



Environmental Analysis Program

Pomona College
Biogeochemistry Laboratory

SOP-L702

ddPCR Data Export and Analysis – Nitrogen Cycle Gene Abundance

Standard Operating Procedure

SOP Number	SOP-L702
Revision	0.1
Author(s)	Marc Los Huertos
Date	2026-06-26
Organization	EA Program, Pomona College

FOR INTERNAL USE ONLY

This document is the property of the Environmental Analysis Program at Pomona College.
Unauthorized distribution is not permitted.

Table of contents

1 Purpose	2
2 Scope	2
3 Responsibilities	2
4 Definitions	2
5 Safety and Hazards	3
6 Equipment and Materials	3
6.1 Equipment	3
6.2 Software	3
7 Procedure	4
7.1 Data Export from QuantaSoft	4
7.2 Quality Control Checks	4
7.3 Calculating Gene Copy Numbers	5
7.4 Data Organization and Merging	6
7.5 Statistical Analysis	6
7.5.1 Data Exploration and Visualization	6
7.5.2 Comparing Gene Abundance Across Treatments	7
7.5.3 Gene Ratios	7
7.5.4 Correlations with Environmental Variables	7
7.5.5 Reporting	7
8 Quality Control and Acceptance Criteria	8
9 Waste Disposal	8
10 References	8
11 Training Record	9
12 Appendix A: Example QC Filtering Script	9
13 Appendix B: Visualization Template	10

14 Appendix C: Sending Samples to External Laboratories

10

1. Purpose

This procedure describes the workflow for exporting raw data from the Bio-Rad QX200 ddPCR system, performing quality control checks, calculating gene copy numbers normalized to soil mass or DNA concentration, and conducting statistical analyses of nitrogen cycle gene abundance data. The SOP links the instrument output from [SOP-L701](#) to reproducible analysis scripts written in R (R Core Team, 2024).

2. Scope

- **Applies to:** Processing and analysis of ddPCR data (QuantaSoft or QuantaSoft Analysis Pro output) for nitrogen cycle functional genes (*amoA*, *nirK*, *nirS*, *nosZ*, *nifH*, *nrfA*) from environmental samples
- **Does not apply to:** Amplicon sequencing data, metagenomics, or qPCR data processed with different software
- **Location:** Instrument computer (QuantaSoft) and RStudio Server (see [SOP-L801 RStudio and GitHub](#))

3. Responsibilities

Role	Responsibility
Lab PI	Approve analysis pipeline and interpretation
Trained Personnel	Export data, run QC, perform analysis
New Students	Complete training before independent analysis

4. Definitions

- **Amplitude threshold:** The fluorescence intensity boundary set in DropletQuant (or equivalent software) that separates positive droplets (containing amplified target) from negative droplets; correct threshold placement is the single most critical step in ddPCR data analysis.
- **Copies per microliter (copies/ μ L):** The absolute concentration of target DNA in the final ddPCR reaction volume as calculated by Poisson statistics from the fraction of positive droplets; must be back-calculated through all dilution steps to report concentration in the original sample.
- **DropletQuant / QuantaSoft:** Bio-Rad software used to read the QX200 droplet reader output, set amplitude thresholds, and calculate target concentrations; raw .qlp files must be archived unmodified.
- **Duplex assay:** A ddPCR assay in which two targets (e.g., a gene of interest and a reference gene) are quantified simultaneously using two spectrally distinct fluorescent probes (FAM and HEX/VIC) in the same reaction; requires careful threshold setting for each channel.

- **Fractional abundance:** The proportion of positive droplets relative to total accepted droplets: $\lambda = -\ln(1 - p)$ where p is the positive fraction; used by Poisson correction to calculate copies/ μL .
- **No-template control (NTC):** A reaction containing all ddPCR components except sample DNA; must yield zero or near-zero positive droplets; persistent positive droplets in NTCs indicate reagent contamination.
- **Poisson correction:** The statistical adjustment applied by DropletQuant to account for the probability that some droplets contain more than one target molecule; assumes random partitioning and is essential for accurate absolute quantification.
- **Rain:** Droplets with intermediate fluorescence amplitude between the positive and negative clusters; caused by partial amplification, degraded DNA, or probe competition; must be assigned to positive or negative by threshold adjustment and noted in data records.
- **Rejected droplets:** Droplets excluded from analysis by the instrument because they fall outside acceptable size or fluorescence quality criteria; runs with fewer than 10,000 accepted droplets per well should be flagged and potentially repeated.

5. Safety and Hazards

No physical or chemical hazards are associated with this procedure. Data integrity and reproducibility are the primary concerns.

Data Integrity

- Never overwrite raw exported data files. Work only with copies.
- All analysis scripts must be version-controlled in the lab GitHub repository.
- Back up raw .q1p files from QuantaSoft to the lab shared drive immediately after each run.

6. Equipment and Materials

6.1. Equipment

Item	Location	Notes
QX200 instrument computer	Instrument room	Runs QuantaSoft or QuantaSoft Analysis Pro
RStudio Server	Remote (see SOP-L801)	For R-based analysis
USB drive or network share	Lab	For transferring .q1p and .csv files

6.2. Software

Software	Version	Purpose
QuantaSoft or QuantaSoft Analysis Pro	1.7+ or 1.0+	Threshold setting, data export
R	4.3+	Statistical analysis
RStudio	Current	IDE for R scripts
tidyverse R package	Current	Data wrangling and visualization
ggplot2 R package	Current	Plotting
<i>TBD – identify additional R packages for ddPCR analysis (e.g., ddpccr, twoddpccr)</i>		

7. Procedure

7.1. Data Export from QuantaSoft

1. Open the completed experiment file (.qlp) in QuantaSoft
2. Review the droplet plots for each well. Confirm that the positive/negative threshold was set appropriately during the run (see SOP-L701, Section 6.6)
3. If threshold adjustment is needed, use QuantaSoft Analysis Pro to re-apply thresholds manually. Document any threshold changes in the lab notebook.
4. Export data using **Options > Export CSV** to generate a comma-separated file containing well-level results
5. The exported CSV includes the following key columns:
 - **Well:** plate position (e.g., A01)
 - **Sample:** sample name as entered during plate setup
 - **Target:** gene target name
 - **Concentration:** copies per uL of the final ddPCR reaction (Poisson-corrected) (Hindson et al., 2011)
 - **Positives:** number of positive droplets
 - **Negatives:** number of negative droplets
 - **AcceptedDroplets:** total accepted droplets
 - **MeanAmplitudeOfPositives:** mean fluorescence of positive cluster
 - **MeanAmplitudeOfNegatives:** mean fluorescence of negative cluster
6. Save the CSV with the naming convention: YYYY-MM-DD_gene_project_plate.csv (e.g., 2025-05-14_amoA_wetland_plate1.csv)
7. Copy the .qlp file and the exported .csv to the lab shared drive under ddPCR_raw_data/

7.2. Quality Control Checks

Perform the following checks on the exported data before proceeding to analysis. These can be scripted in R for reproducibility.

1. **Droplet count filter:** Exclude any well with fewer than 10,000 accepted droplets. Wells below this threshold have unreliable Poisson statistics (Bio-Rad Laboratories, 2024).
2. **No-template controls (NTC):** NTC wells must show 0 or fewer than 3 positive droplets. If NTC wells show more than 3 positives, the entire plate or the affected target should be considered contaminated and rerun.
3. **Extraction blanks:** Extraction blanks (from SOP-L80) must show fewer positive droplets than the NTC acceptance threshold. Elevated counts in extraction blanks indicate contamination in the DNA extraction workflow.
4. **Positive controls:** Positive control gBlock wells should fall within 20% of the expected concentration based on the dilution prepared.
5. **Technical replicate agreement:** If technical replicates were run, calculate the coefficient of variation (CV) across replicates. A CV greater than 25% suggests pipetting error or sample heterogeneity.



Tip

Create a QC report template in R Markdown or Quarto that automatically flags wells failing any of the criteria above. This makes batch QC efficient and ensures nothing is missed.

7.3. Calculating Gene Copy Numbers

The QuantaSoft **Concentration** output is in copies per uL of the ddPCR reaction mixture. To express results per gram of dry soil (or per unit of extracted DNA), apply the following conversions:

$$\text{Copies per g dry soil} = \frac{C_{\text{ddPCR}} \times V_{\text{reaction}} \times DF \times V_{\text{elution}}}{V_{\text{template}} \times m_{\text{soil}} \times (1 - \theta_g)}$$

where:

- C_{ddPCR} = concentration from QuantaSoft (copies/uL)
- V_{reaction} = total ddPCR reaction volume (typically 20 uL)
- DF = dilution factor of the DNA template (e.g., 10 if diluted 1:10)
- V_{elution} = total DNA elution volume from extraction (uL; see SOP-L80)
- V_{template} = volume of DNA template added to the ddPCR reaction (uL)
- m_{soil} = mass of soil extracted (g)
- θ_g = gravimetric soil moisture (as a decimal; see [SOP-L400](#))

To express results per ng of extracted DNA instead:

$$\text{Copies per ng DNA} = \frac{C_{\text{ddPCR}} \times V_{\text{reaction}}}{V_{\text{template}} \times C_{\text{DNA}}}$$

where C_{DNA} is the Qubit DNA concentration (ng/uL) from SOP-L80.

```
# Example calculation: copies per g dry soil
c_ddpcr <- 52.3      # copies/uL (from QuantaSoft)
v_reaction <- 20     # uL
dilution_factor <- 10 # 1:10 dilution of template
v_elution <- 100    # uL (from SOP-L80)
v_template <- 2      # uL added to ddPCR reaction
soil_mass_g <- 0.25   # g soil extracted
moisture_decimal <- 0.35 # gravimetric moisture as decimal

copies_per_g <- (c_ddpcr * v_reaction * dilution_factor * v_elution) /
  (v_template * soil_mass_g * (1 - moisture_decimal))

cat(sprintf("Gene copies per g dry soil: %.0f\n", copies_per_g))
```

Gene copies per g dry soil: 3218462

```
cat(sprintf("Gene copies per g dry soil (log10): %.2f\n", log10(copies_per_g)))
```

Gene copies per g dry soil (log10): 6.51

7.4. Data Organization and Merging

1. Merge the QC-filtered ddPCR results with the sample metadata (site, treatment, depth, date, soil moisture, DNA concentration) using a shared sample ID column
2. Organize the merged data in a tidy (long) format with one row per sample-by-gene combination (Wickham, 2014)
3. Save the merged analysis-ready dataset as a .csv in the project GitHub repository under data/processed/
4. TBD – define the standard column naming convention for the lab's ddPCR datasets (e.g., *sample_id*, *gene*, *copies_per_g_dry*, *copies_per_ng_dna*, *site*, *treatment*, *depth_cm*, *date*)

7.5. Statistical Analysis

Note

The choice of statistical approach depends on the experimental design (e.g., paired vs. independent sites, factorial treatments, time series). The guidance below covers common analyses for nitrogen cycle gene abundance. Consult the PI before selecting a final analytical approach.

7.5.1. Data Exploration and Visualization

1. Plot gene copy numbers ($\log_{10}$ -transformed) by treatment, site, or depth using boxplots or dot plots

2. Assess normality of log-transformed data using Shapiro-Wilk tests and Q-Q plots
3. Check for outliers and investigate whether they represent true biological variation or analytical artifacts

7.5.2. Comparing Gene Abundance Across Treatments

- For two-group comparisons (e.g., amended vs. control): use Welch's t-test on $\log_{10}$ -transformed copies per g dry soil, or the non-parametric Wilcoxon rank-sum test if normality assumptions are not met (Frosteg aard et al., 1999)
- For multi-group comparisons: use one-way ANOVA on $\log_{10}$ -transformed data with Tukey HSD post hoc tests, or Kruskal-Wallis with Dunn's test for non-parametric data
- For factorial designs (e.g., treatment x depth): use two-way ANOVA or linear mixed-effects models with site as a random effect (Pinheiro & Bates, 2000)

7.5.3. Gene Ratios

Ratios of functional gene abundances can provide insight into the balance of nitrogen cycling processes (Hallin et al., 2009):

- *nirK/nirS*: relative contribution of the two nitrite reductase pathways to denitrification
- *nosZ/(nirK + nirS)*: completeness of denitrification (higher values suggest greater N_2O reduction capacity)
- *amoA* AOA/*amoA* AOB: relative dominance of archaeal vs. bacterial ammonia oxidation
- *nrfA/(nirK + nirS)*: relative importance of DNRA vs. denitrification

TBD – determine which gene ratios are most relevant for the lab's current research questions and establish standard visualization formats.

7.5.4. Correlations with Environmental Variables

- Correlate gene abundances with soil physicochemical data (pH, moisture, total C, total N, nitrate, ammonium) using Pearson or Spearman correlations
- *TBD – evaluate whether multivariate approaches (e.g., PCA, redundancy analysis) are appropriate given the number of samples and variables*
- *TBD – determine whether gene abundance data should be related to process rate measurements (e.g., N_2O flux, nitrification potential) and define the analytical framework for doing so*

7.5.5. Reporting

- Report gene abundance as copies per g dry soil (mean +/- standard error) and/or copies per ng DNA
- Present data on a $\log_{10}$ scale in figures; report back-transformed means and confidence intervals in tables where appropriate (Bolker et al., 2009)
- Include all QC metrics (droplet counts, NTC results, positive control recovery) as supplementary tables
- *TBD – create a standardized R Markdown/Quarto reporting template for ddPCR results that the lab can reuse across projects*

8. Quality Control and Acceptance Criteria

- Raw .q1p files must be archived on the lab shared drive before any analysis proceeds
- All R scripts must be committed to the lab GitHub repository with meaningful commit messages
- QC filtering must be applied before any statistical analysis; unfiltered data should never be reported
- Positive control recovery within 20% of expected value
- TBD – define minimum sample sizes required for each statistical test based on power analysis
- TBD – establish whether and how to handle values below the limit of detection (e.g., substitution with LOD/2, censored regression, or Tobit models) (Helsel, 2011)

9. Waste Disposal

No physical waste is generated by this procedure. Data management practices are described in Section 6.4.

10. References

- Bio-Rad Laboratories. (2024). *QX200 droplet digital PCR system instruction manual*. <https://www.bio-rad.com/en-us/product/qx200-droplet-digital-pcr-system>
- Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J.-S. S. (2009). Generalized linear mixed models: A practical guide for ecology and evolution. *Trends in Ecology and Evolution*, 24(3), 127–135. <https://doi.org/10.1016/j.tree.2008.10.008>
- Frosteg aard, Aa., Courtois, S., Ramisse, V., Clerc, S., Bernillon, D., Le Gall, F. ccois, Jeannin, P., Nesme, X., & Simonet, P. (1999). Quantification of bias related to the extraction of DNA directly from soils. *Applied and Environmental Microbiology*, 65(12), 5409–5420. <https://doi.org/10.1128/AEM.65.12.5409-5420.1999>
- Hallin, S., Jones, C. M., Schlöter, M., & Philippot, L. (2009). Relationship between N-cycling communities and ecosystem functioning in a 50-year-old fertilization experiment. *The ISME Journal*, 3(5), 597–605. <https://doi.org/10.1038/ismej.2008.128>
- Helsel, D. R. (2011). *Statistics for censored environmental data using Minitab and R* (2nd ed.). John Wiley; Sons. <https://doi.org/10.1002/9781118162729>
- Hindson, B. J., Ness, K. D., Masquelier, D. A., Belgrader, P., Heredia, N. J., Makarewicz, A. J., Bright, I. J., Lucero, M. Y., Hiddessen, A. L., Legler, T. C., Kitano, T. K., Hodel, M. R., Petersen, J. F., Wyatt, P. W., Steenblock, E. R., Shah, P. H., Bousse, L. J., Troup, C. B., Mellen, J. C., ... Colston, B. W. (2011). High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. *Analytical Chemistry*, 83(22), 8604–8610. <https://doi.org/10.1021/ac202028g>
- Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. Springer-Verlag. <https://doi.org/10.1007/b98882>
- R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>

Wickham, H. (2014). Tidy data. *Journal of Statistical Software*, 59(10), 1–23. <https://doi.org/10.18637/jss.v059.i10>

11. Training Record

Before performing this procedure independently, personnel must:

1. Read this SOP, [SOP-L701 Primer Design](#), and [SOP-L700 DNA Extraction](#)
2. Complete [SOP-L801 RStudio and GitHub](#) training
3. Demonstrate competence exporting data from QuantaSoft
4. Successfully run the QC and analysis pipeline on a practice dataset under supervision
5. Be signed off by the PI or lab manager

12. Appendix A: Example QC Filtering Script

```
library(tidyverse)

# Read exported QuantaSoft CSV
raw <- read_csv("data/raw/2025-05-14_amoA_wetland_plate1.csv")

# QC filtering
qc_filtered <- raw %>%
  filter(
    AcceptedDroplets >= 10000          # Minimum droplet threshold
  ) %>%
  mutate(
    qc_flag = case_when(
      Sample == "NTC" & Positives > 3 ~ "NTC_contamination",
      Sample == "Extraction_Blank" & Positives > 3 ~ "Extraction_contamination",
      TRUE ~ "Pass"
    )
  )

# Flag and report QC failures
qc_failures <- qc_filtered %>%
  filter(qc_flag != "Pass")

if (nrow(qc_failures) > 0) {
  warning("QC failures detected:")
  print(qc_failures %>% select(Well, Sample, Target, Positives, qc_flag))
}

# Retain only passing sample wells
analysis_ready <- qc_filtered %>%
```

```
filter(qc_flag == "Pass",
       !Sample %in% c("NTC", "Extraction_Blank", "Positive_Control"))
```

13. Appendix B: Visualization Template

```
library(tidyverse)

# Assumes analysis_ready has columns: sample_id, gene, copies_per_g_dry, treatment
analysis_ready %>%
  ggplot(aes(x = treatment, y = log10(copies_per_g_dry), fill = treatment)) +
  geom_boxplot(alpha = 0.7, outlier.shape = NA) +
  geom_jitter(width = 0.2, alpha = 0.5, size = 1.5) +
  facet_wrap(~ gene, scales = "free_y") +
  labs(
    x = "Treatment",
    y = expression(log[10]~"(gene copies g"-1~"dry soil)"),
    title = "Nitrogen Cycle Gene Abundance by Treatment"
  ) +
  theme_minimal() +
  theme(legend.position = "none")
```

14. Appendix C: Sending Samples to External Laboratories

TBD – if ddPCR analysis is outsourced to an external service provider (e.g., university core facility, commercial lab), document the following:

- Preferred service provider(s) and contact information
- Sample submission requirements (volume, concentration, format)
- Data format and file types returned by the provider
- Turnaround time expectations
- Cost per sample and funding source/account number
- How to reconcile externally generated data with in-house QC standards