

Article

Challenges using droplet digital PCR for environmental samples

Vasilis Kokkoris^{1,2,3*}, Eric Vukicevich¹, Andrew Richards¹, Corrina Thomsen¹, Miranda M. Hart¹¹Department of Biology, University of British Columbia, Okanagan campus, 3333 University Way, Kelowna, BC V1V 1V7, Canada² Department of Biology, University of Ottawa, Marie-Curie Private, Ottawa, ON K1N 9A7, Canada³ Agriculture and Agri-Food Canada, Ottawa Research and Development Centre, Ottawa, ON K1A 0C6, Canada*Corresponding author. E-mail address: vkokkori@uottawa.ca

Abstract: Droplet digital polymerase chain reaction (ddPCR) is a method used to detect and quantify nucleic acids even when present in exceptionally low numbers. While it has proven to be valuable for clinical studies, it has failed to be widely adopted for environmental studies. Due to the complexity of the chemical and biological composition of environmental samples, protocols tailored to clinical studies are not appropriate, and results are difficult to interpret. We used environmental DNA samples originating from field studies to determine a protocol for environmental samples. Samples included field soils which had been inoculated with the soil fungus *Rhizoglyphus irregularis* (**environmental positive control**), field soils that had not been inoculated and the targeted fungus was not naturally present (**environmental negative control**), and root samples from both field categories. To control for the effect of soil inhibitors, we also included DNA samples of an organismal control extracted from pure fungal spores (**organismal positive control**). Finally, we included a **no-template control** consisting only of the PCR reaction reagents and nuclease free water instead of template DNA. Using original data, we examined which factors contribute to poor resolution in root and soil samples and propose best practises to ensure accuracy and repeatability. Furthermore, we evaluated manual and automatic threshold determination methods and we propose a novel protocol based on multiple controls that is more appropriate for environmental samples.

Keywords: ddPCR, droplet digital PCR, environmental samples, environmental DNA, soil microbiology, threshold determination.

1. Introduction

Droplet digital polymerase chain reaction (ddPCR) allows for the quantification of nucleic acid sequences by partitioning single samples into thousands of nanolitre droplets (~20,000), each containing either single copy or a few copies of the target, or no initial target sequences [1]. Partitioning of PCR samples creates a sensitive amplification protocol that is less likely to be affected by inhibitors and non-target DNA [2,3]. As a result, it is becoming the approach of choice for assays looking to detect low copy targets, for example, deleterious alleles in early stage cancer patients[4,5], introduced genes in genetically modified organisms [6,7] and as a proxy for the abundance of pathogens in environmental samples [8,9]. An advantage of ddPCR is that amplified droplets that can contain only a single copy of the target sequence increasing the possibility to quantify rare sequence variants [1] compared with approaches such as quantitative PCR (qPCR) (Fig. 1).

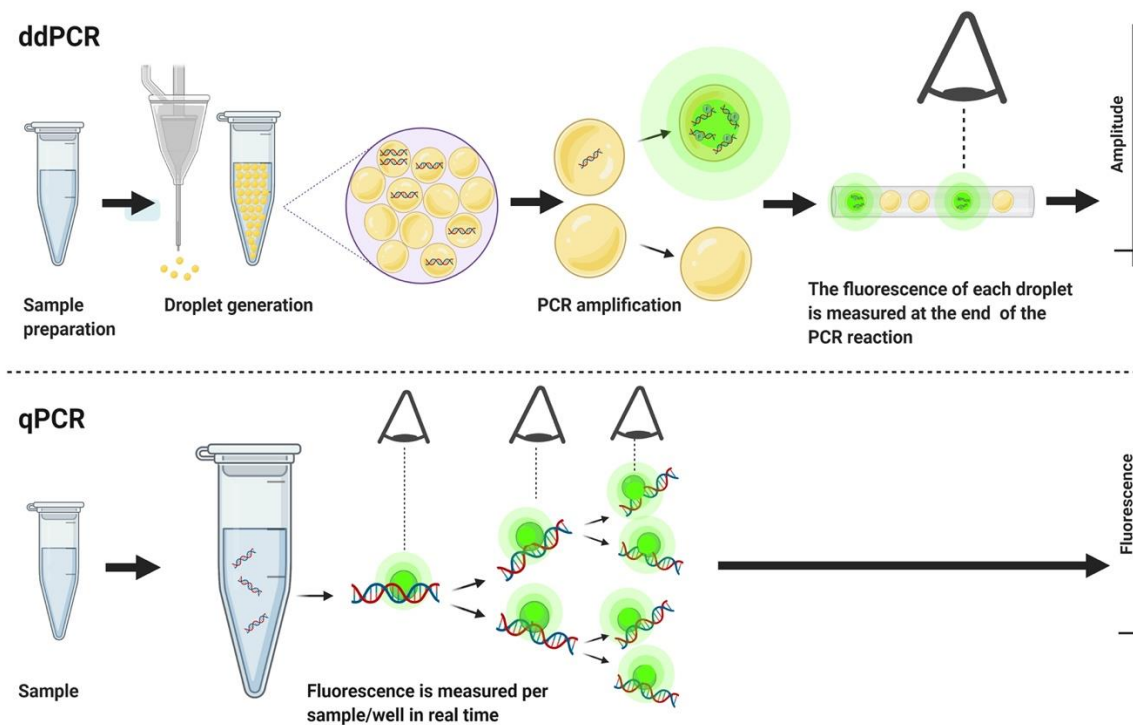


Figure 1. Droplet digital PCR (ddPCR) versus quantitative PCR (qPCR).

In ddPCR, samples are partitioned into thousands of droplets each of which undergo amplification. After PCR cycling, the fluorescence of each droplet is measured at the end of the PCR reaction. As an output, droplets are grouped as positive or negative based on their fluorescence amplitude levels (y axis). In contrast, qPCR, fluorescence gives one measure per sample in real time, at the end of each PCR cycle. Due to such fundamental differences in the approach, guidelines for standardizing qPCR assays such as the Minimum Information for publication of Quantitative real-time PCR Experiments (MIQE) [10] are not appropriate for ddPCR, and users are left to develop their own quality control protocols. The figure was created with BioRender (BioRender.com).

The application of ddPCR to environmental studies has lagged behind its clinical use, both due to the cost compared to qPCR (up to three times more costly per sample [11] but also poor resolution in environmental samples [2]. The chemical and biological complexity of environmental samples can make it difficult, even impossible, to discriminate among positive and negative reactions. In successful ddPCR assays, positive and negative droplets aggregate into respective “clouds” based on the intensity of the fluorescence when a fluorescent signal is emitted during the amplification of target DNA. In assays with mutually exclusive positive and negative clouds, the assignment of ‘positives’ is straightforward. In reality, it can be difficult to achieve discrete clouds, particularly in environmental samples, which often exhibit droplets of intermediate fluorescence (for example [12, 13]). These intermediate droplets are referred to as ‘rain’ as they do not clearly affiliate with positive or negative clouds and it can represent a significant proportion of droplets in environmental samples.

There is no uniform cause, nor approach for dealing with ‘rain’ as it may have different causes [14]. For example, droplet coagulation [13] or variation in droplet size could affect fluorescence [15]. Variation in target amplification efficiency may also arise through physical problems of template DNA [16] including fragmentation of DNA incurred during extraction or storage [17]. Degraded DNA from dead organisms may also contribute to midlevel amplification of droplets [18]. Intermediate fluorescence arising from assay conditions are more problematic for environmental assays because these reactions are likely to contain inhibitors (e.g. humic acids) [19], resulting in delayed or reduced amplification efficiency or ‘rain’ [20]. While ddPCR is more robust than qPCR to these inhibitors because of sample partitioning and post-cycle data collection [21], inhibitors may still reduce amplification efficiency [22].

Here, we first outline the factors that contribute to ambiguous results that can occur in environmental samples along with practices that can reduce ambiguity, improving both accuracy and reproducibility. These are largely due to issues that prevent separation of droplets into groups containing target DNA or not (‘clouds’) and result in intermediate levels of fluorescence (‘rain’) that make defining thresholds difficult. Below, we use original work to illustrate how cycling conditions and threshold determination can significantly affect the quality of ddPCR assays for environmental samples. We then propose a novel protocol for applying ddPCR to terrestrial environmental samples.

2. Materials and Methods

To determine best practises for environmental samples, we conducted a series of experiments to test the effect of altering cycling conditions on the presence of ‘rain’. We used additional controls to help identify the conditions that best improve successful target amplification and reduce partial, non-target amplification. We also compared the available methods for determining thresholds including manually determined threshold.

2.1 Hardware and software

We used the QX100™ droplet digital PCR system (Bio-Rad, Mississauga, ON, Canada) for droplet production and analysis, following the manufacturer instructions. For PCR parameters optimization, we used the C1000 Touch Thermal Cycler (Bio-Rad, Mississauga, ON, Canada). Boxplots were produced using the package ggplot [23] in R studio (Version 1.0.136 – © 2009–2016 RStudio, Inc.). Optimization figures were produced using the QuantaSoft Analysis Pro (version 1.0.596) software and further modified at Microsoft word 2016 for label addition.

2.2 Origin of environmental DNA used to optimize ddPCR

In order to optimize ddPCR for environmental samples, we used DNA samples associated with another field study (for details see [24]). In particular, we tested root and soils samples which were positive or negative for our target organism, the arbuscular mycorrhizal (AM) fungus *Rhizoglossus irregularis* Schenck & Smith (synonym *Rhizophagus irregularis*, *Glomus intraradices*) strain DAOM197198 [25]. Samples included field soils which had been inoculated (environmental positive control, n = 20), field soils that had not been inoculated and field soil originated from natural ecosystems (Saskatchewan, Canada) where the targeted fungus was not naturally present (environmental negative control, n = 20), and root samples (from inoculated and non inoculated *Lens culinaris* - common name Lentils belonging both in the environmental positive control and environmental negative control categories, n = 40) from both field categories. To control for the effect of soil inhibitors, we also included DNA samples of an organismal control extracted from pure spores of *R. irregularis* (organismal positive control, n = 3). Finally, we

included a no-template control (n=3) consisting only of the PCR reaction reagents and nuclease free water instead of template DNA.

2.3 DNA extraction and primers/probe

In all cases DNA was extracted using the PowerSoil® DNA Isolation Kit (MoBio) following the manufacturers instructions. The sequence of primers and probes used are specific to *R. irregularis* strain DAOM197198 and the sequences after are:

Forward: 5'- AGCAAATCTAAGTTCCTCAGAG -3'

Reverse: 5'- ACTTCTATGGCTTTGTACAGG-3'

Probe: FAM/CCCACCAGG/ZEN/GCAGATTAAATCTTCCTT/3IABKFQ-3

Standardized cycling conditions and ddPCR reaction mixture

The standardized cycling conditions based on the manufacturer's recommendations are: 10 min at 95°C for enzyme activation, followed by 40 cycles of 30 seconds at 94°C and 1 min annealing/extension step at the appropriate temperature based on the primer/probe set (in our case at 59°C). The final step is 10 min at 98°C for enzyme deactivation followed by infinite hold at 4°C. For each step, a ramp rate of 2°C/sec is applied. The standardise PCR mixture per sample based on the manufacturers recommendations are: 10 µl of 1X Supermix for Probes (Bio-Rad), 1 µl of primers/probe mixture PrimeTime® Std qPCR Assay (500 rxn) with a final concentration of 500 nM primers and 250 nM probe in the reaction, 7 µl of DNase free water and 2 µl of sample DNA.

2.4 Parameters tested

In a series of experiments using DNA from the field soils and plant roots we manipulated thermocycler steps individually and combined, in order to examine the additive effect of altered cycling condition to the ddPCR optimization. We also examined the effect of primer/probe concentration and sample concentration in reaction optimization. For all reactions, along with organismal positive control and no-template controls, we examined the usability of an environmental positive and an environmental negative control on the reaction optimization and threshold determination.

Thermocycler conditions: It is unknown whether manipulating multiple cycling condition simultaneously has an additive effect to the ddPCR optimization. For all reactions, the initial step consisted of 10 min at 95°C for enzyme activation, Following enzyme activation we manipulated annealing/extension temperature by applying a temperature gradient for ranging from 57°C to 67°C (the temperature range depends on the primer/probe set and has to be adjusted accordingly). Furthermore, we examined the effect of number of PCR cycles with minimum of 40 cycles (recommended by BioRad) up to a maximum of 45 cycles. We also examined the effect of time extension during denaturation (from 30 sec which is the recommended BioRad time to 1 min) and annealing/extension step (from 1 min which is the recommended BioRad time to 2 min). Finally, we manipulated the ramp rate and we examined differences between a ramp rate of 2°C/sec (recommended by Bio-Rad) and 1°C/sec. The analysis was performed using all the available soil and root DNA samples.

2.4.1 Primer/probe concentration

We examined the effect of reduced primer/probe concentration, below the manufacturers recommended quantity, on the reaction's efficiency. The PCR mixture for the ddPCR

reaction per sample contained: 10µl of 1X Supermix for Probes (Bio-Rad), 1µl primers/probe mixture (PrimeTime® Std qPCR Assay (500 rxn), 7 µl of DNase free water and 2µl of DNA. We used three different primer/probe quantities: 20X (recommended concentration) that yields a final concentration of 500 nM primers and 250 nM probe, 10X (reduced concentration) that yields a final concentration of 250 nM primers and 125 nM probe, and finally, 1X (reduced concentration) that yields a final concentration of 50 nM primers and 25 nM probe. We used the organismal positive control to demonstrate the effect of primer concentration.

2.4.2 Sample DNA concentration

We tested whether sample dilution (up to 100-time dilution) and spiking unknown environmental samples with known quantity of organismal DNA can facilitate the discrimination between true or false positive samples. For this purpose, we used unknown environmental samples that contained extreme amount of “rain” to the level that separation of negative and positive clouds was not possible. For this assay we used multiple field soil (n=10) and root (n=8) DNA sample with a heavy rain profile.

2.4.3 Automated algorithm comparison

First, to identify current threshold determination approaches for environmental studies using ddPCR, we searched Web of Science (Clarivate Analytics 2018). We identified 75 papers including applications in monitoring of invasive species [26,27], detection of microbes in soil, plant tissue and water [28,29,30,31] and testing food for contamination by microbes or non-declared species [32](Table S1). The most common strategy for threshold assignment was the automatic threshold set by *QuantaSoft™*, the default software provided by BioRad Technologies, Inc. (2017) followed by manually set threshold, but for the majority of the studies without explained parameters, via the *QuantaSoft™ software*. Additional threshold determination approaches include the algorithms, *Define the Rain*, *ddPCRquant* and *Umbrella*.

2.4.3.1 Define the Rain is a Java script that uses a k-nearest neighbour algorithm to define a distinct positive and negative cloud separately for a positive control sample and subsequently identify extreme values (droplets) that are outliers [33]. The two cut-off thresholds (positive and negative) are defined as three times the standard deviation of the mean droplet amplitude within each cluster. These thresholds are then applied to all samples and the outliers that fall between them are classified as rain. Final concentration is then calculated using a Poisson distribution of positive and negative droplets with intermediate ‘rain’ removed. This algorithm has been shown to be a significant improvement in accuracy compared to the *QuantaSoft®* software, at least for copy number detection in clinical samples (Jones et al., 2014). *Define The Rain* base thresholds on a positive control derived from pure culture. Such controls may have tighter clustering and higher fluorescence of positive droplets due to the absence of inhibitors that exist in environmental samples. This can increase the number of false positives, as significant quantities of true positive events were excluded when discarding ‘rain’ as per *Define The Rain* [33,34] or another algorithm [35]. An ‘environmental positive’ or environmental sample spiked with target DNA may more accurately represent the distribution of positive and negative clouds in environmental samples and represent a better control.

2.4.3.2 ddPCRquant uses single thresholds defined by distribution of fluorescence amplitude of negative clouds in non-template controls (NTC)[13]. Originally proposed by defining a global threshold of the mean fluorescence of three NTC plus six times their standard deviation, then improved upon this idea by incorporating extreme value theory to define positive droplets (the ‘extreme’ events separate from a strong cloud of NTC droplets)[13]. Because *ddPCRquant* does not rely on positive controls and takes into

consideration baseline fluorescence, *ddPCRquant* may be more robust for environmental samples. However, to date there are no published studies citing the use of *ddPCRquant* for analysis of environmental samples.

2.4.3.3 *Umbrella* is the newest alternative for threshold determination and does so by maximizing precision across an experiment [36]. *Umbrella*, an open source R package, assumes that each sample follows a mixture model with two components representing a negative and positive cloud. It then uses Bayes' theorem to determine the likelihood that each droplet belongs to one of two groups. This algorithm draws a threshold based on the distribution of all droplets in a given well with no assumptions about how an assay might affect droplet fluorescence. Samples without identifiable positive clouds will be returned with a high measure of uncertainty. Thus, this approach may be appropriate only for samples with clearly delineated clouds. There are no studies using *ddPCR* for environmental samples to date that cite the use of *Umbrella*.

We used information from six samples, one field soil environmental negative control, one organismal positive control, one field soil unknown environmental sample and three non-template control samples (NTC's) (required for *ddPCRquant* analysis) to compare threshold determination of the following algorithms as opposed to a manually determined threshold. Specifically, our treatments were:

2.4.3.4 *Manual* threshold: We used the QuantaSoft Analysis Pro (version 1.0.596) software in order to define the manual threshold as described on the section 5.4. Briefly, manual threshold determination was calculated based on the NTC's, the organismal positive control and the environmental negative control. *QuantaSoft* automatic threshold: We used the QuantaSoft Analysis Pro (version 1.0.596) software in order to automatically determine the threshold using the automatic threshold determination for multiple well selection. *ddPCRquant* threshold: In order to determine the threshold using '*ddPCRquant*' we used the online available tool "<http://statapps.ugent.be/dPCR/ddpcrquant/>" with the following parameters: Block size = 150, 10 repeats, 201 μ l of volume MIX, 1 μ l of template volume, threshold determination. *Umbrella* threshold: We were unable to use the '*Umbrella*' algorithm since we could not locate the open source code for R which was not yet available. "*Define The Rain*" threshold: We used the online available tool <http://definetherain.org.uk/> and the automatic threshold determination after uploading the sample information.

3. Results

We summarize the parameters optimized for the following analyses in Table 1.

Table 1. Optimizing ddPCR assays for environmental samples. For the purpose of this study, we examined the additive effect of optimized cycling conditions suggested by the literature and the effect of the primer/probe and sample concentration on the efficiency of the reaction. We also recommend the use of multiple controls that can further optimize the annealing/extension temperature and facilitate threshold determination.

	Variable	Recommendations
Optimizing Cycling conditions	Denaturation	Increase to 1 min
	Annealing/ extension	Increase to 2 min
	Ramp rate	Decrease to 1°C/sec
	Number of cycles	Increase to 44 - 50
Optimizing Annealing/ extension temperature	Annealing/ extension temperature	Using organismal positive control
		Use a high abundance positive environmental sample
		Use a low in abundance positive environmental sample
		Use a negative environmental sample
		Use a non-template control (NTC)
Adjusting primer/probe concentration	Primer/ probe concentration	Use recommended concentration
Adjusting sample concentration	Sample concentration	Dilute very concentrated samples as needed when there is no clear cluster separation or a lot of rain between clusters
Threshold determination	Manually set threshold (should be determine taking under consideration multiple control)	Using a high abundance positive environmental sample
	Minimum accepted positive droplet threshold	Using a negative environmental sample
		Use three – five NTC samples. Use both a positive control dilution series and a spiked environmental sample with same dilution series

3.1 Optimizing cycling conditions.

The optimized protocol increased amplitude levels and cloud clustering efficiency (Fig. 2). After full optimization, the resolution between the clouds is increased and clouds become more tightly clustered. Also, the “rain” decreases from 2394 droplets to 254 droplets in the optimized protocol.

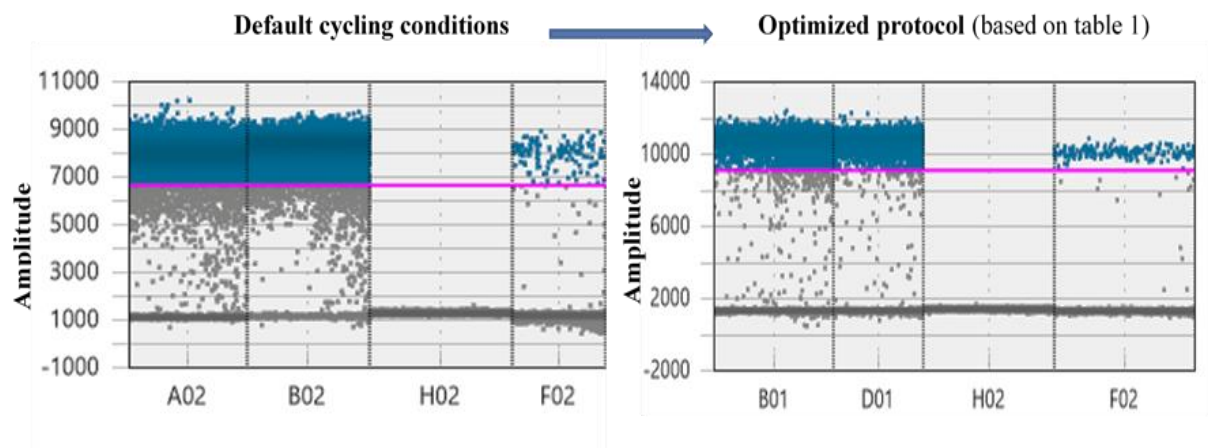


Figure 2. The effect of protocol optimization based on the suggestions presented in Table 1 on the output from a ddPCR assay examining environmental samples at two different template concentrations. Blue dots indicate putative positive droplets whereas the grey represent negatives. The first two wells represent two environmental sample with different template concentrations (A02 field soil DNA sample and B02 root DNA sample), and the last two wells represent a non template and an organismal positive control, respectively. The pink line is a manually set high threshold based on the organismal positive control of the optimized protocol and it is used to demonstrate the increase in amplitude levels after the optimization. The width of each well is related to the number of final droplets analyzed (accepted samples need to have more than 10000 droplets).

3.2 Optimizing Annealing/ extension temperature

Multiple controls are required to determine optimum annealing and extension temperature (Fig. 3). Using only organismal (a) and environmental (b) positive controls suggests an optimum annealing temperature between 57-59° C to maximize separation of the clouds and minimize 'rain' (Fig.3). However, this is not true for environmental negative controls (c). Here, 59°C significantly decreased **non-specific amplification**, limiting the possibility for false positives. Although 61°C almost eliminates rain, it is not optimal for the environmental positive control (Fig 3b).

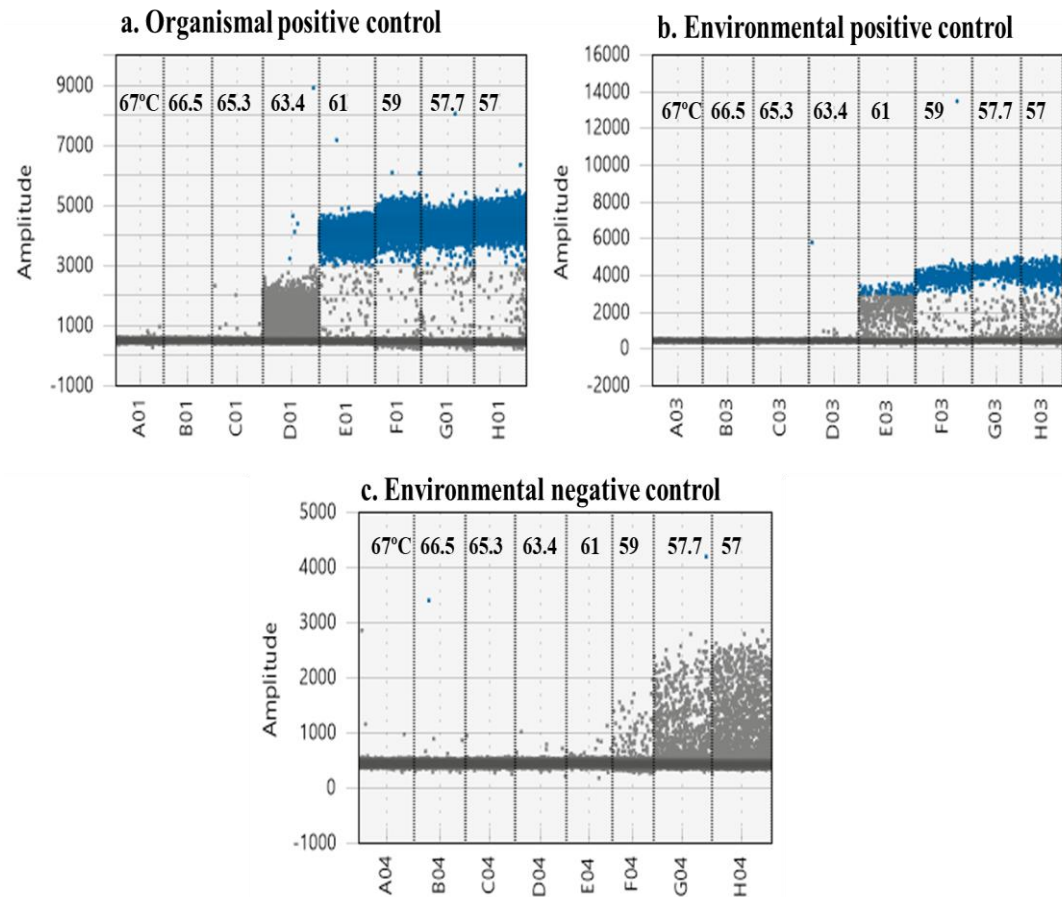


Figure 3. Multiple controls are required to determine optimum annealing and extension temperature. Organismal (a) , environmental (b) positive controls and (c) environmental negative controls (c). Multiple positive controls improve discrimination between true positives and “rain” or false positives.

3.3 Adjusting primer/probe concentration

Higher primer concentration improves resolution and cloud clustering (Fig. 4). We used an organismal positive control to demonstrate the primer concentration effect was an organismal positive control. Using 1X primer and probe concentration (which yields 25 nM primers and 12.5 nM probe), results in positive cloud amplitude around 500.(4a) Primer and probe concentration of 10x shifts the amplitude of the positive cloud closer to 3000. (4b). The recommended 20x primer and probe concentration shifts the positive cloud to 4500 (4c). Red circles emphasize the amplitude levels of each primer/probe concentration used and indicate the effect of changing the primer/probe concentration on the amplitude level.

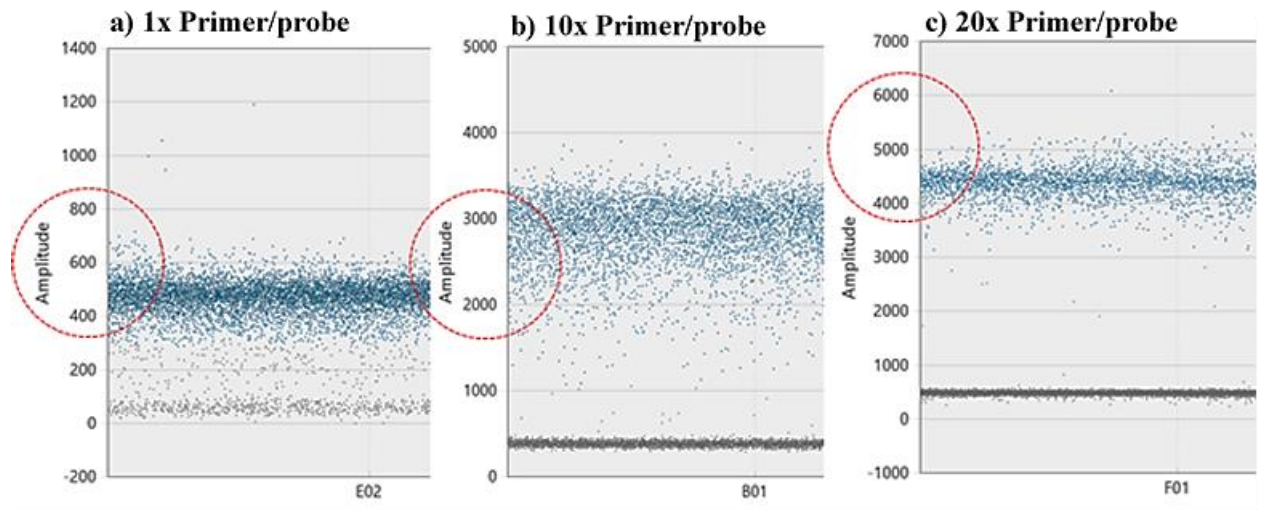


Figure 4. Higher primer concentration improves resolution and cloud clustering. We used an organismal positive control to demonstrate the primer concentration effect was an organismal positive control. a) using 1X primer and probe concentration (which yields 25 nM primers and 12.5 nM probe), results in positive cloud amplitude around 500. b) Primer and probe concentration of 10x shifts the amplitude of the positive cloud closer to 3000. c) The recommended by the manufacturer 20x primer and probe concentration, shifts the positive cloud to 4500 (c). Red circles emphasize the amplitude levels of each primer/probe concentration used and indicate the effect of changing the primer/probe concentration on the amplitude level.

3.4 Adjusting sample concentration

Dilution and spiking of unknown environmental samples, help distinguish between false and true positive samples (Fig. 5). A highly concentrated sample prevents discrete cloud clustering and results in reduced amplitude levels (5a). By diluting the sample, it is easier to discriminate between false and positive droplets. When dilution does not result in cluster separation, we can distinguish between non-specific amplification or strong inhibition, by spiking the sample with a known quantity of our target (5b).

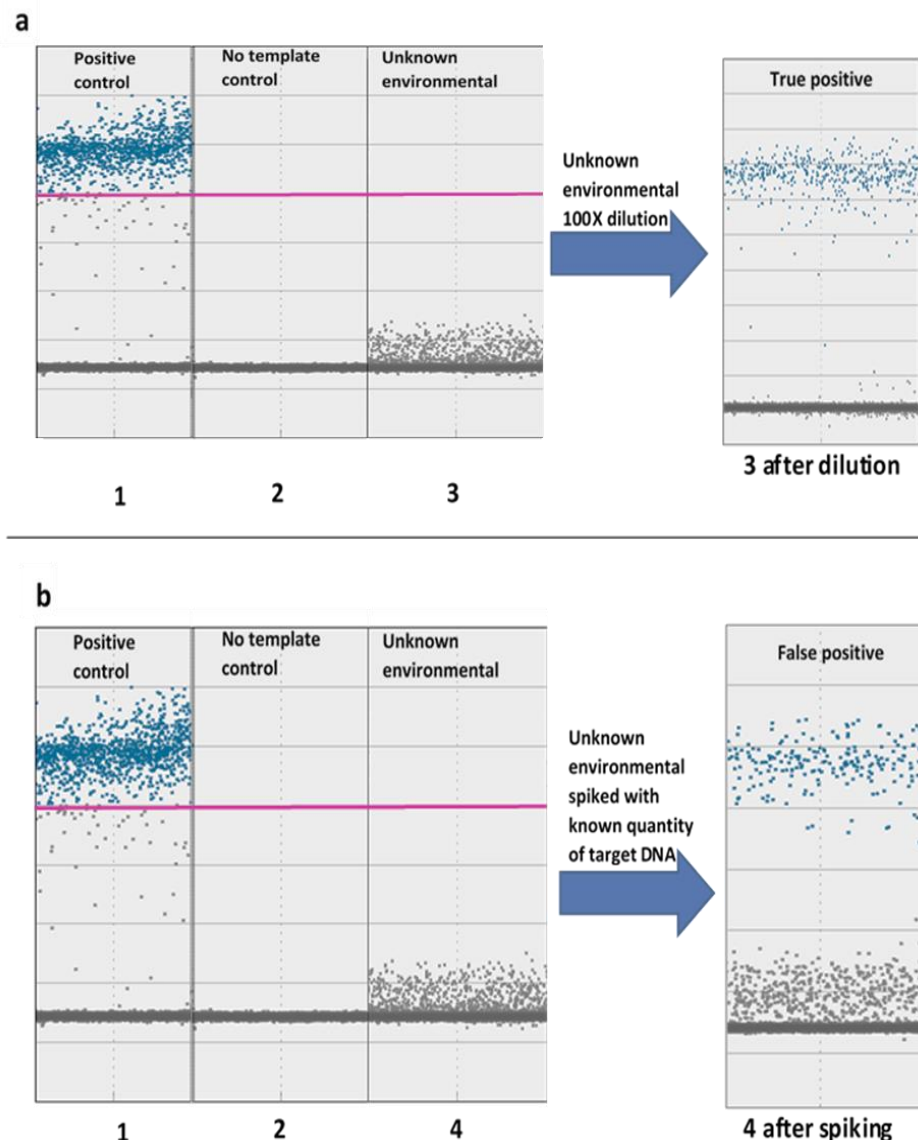


Figure 5. Dilution (a) and spiking (b) of unknown environmental samples help distinguish between false and true positive samples. a) The low cloud in sample a3 clearly separates into a true positive cloud and a negative cloud after dilution. b) The unknown environmental sample (sample b4) is indeed negative because the added known quantity of our target appears to the amplitude levels of the positive while unknown cloud remains. The low cloud produced in sample 4 may be due to nonspecific or partial amplification or the reflective property of proteins existing in the environmental sample. Samples a1 and b1 corresponds to an organismal positive control while samples a2 and b2 represent a no-template control. The pink line is a high threshold determined based on the positive cloud of the organismal positive control.

3.5 Threshold determination

A high threshold based on a positive control may exclude positive droplets in positive environmental samples which may have reduced amplitude levels (Fig. 6). A medium threshold defined by non-positive environmental samples could avoid non-specific amplification but also include positive samples with lower amplitude due to inhibition or high target quantity. Setting a low threshold based on the NTC (common practice) can

erroneously identify negative environmental samples as positives due to low level of non-specific amplification.

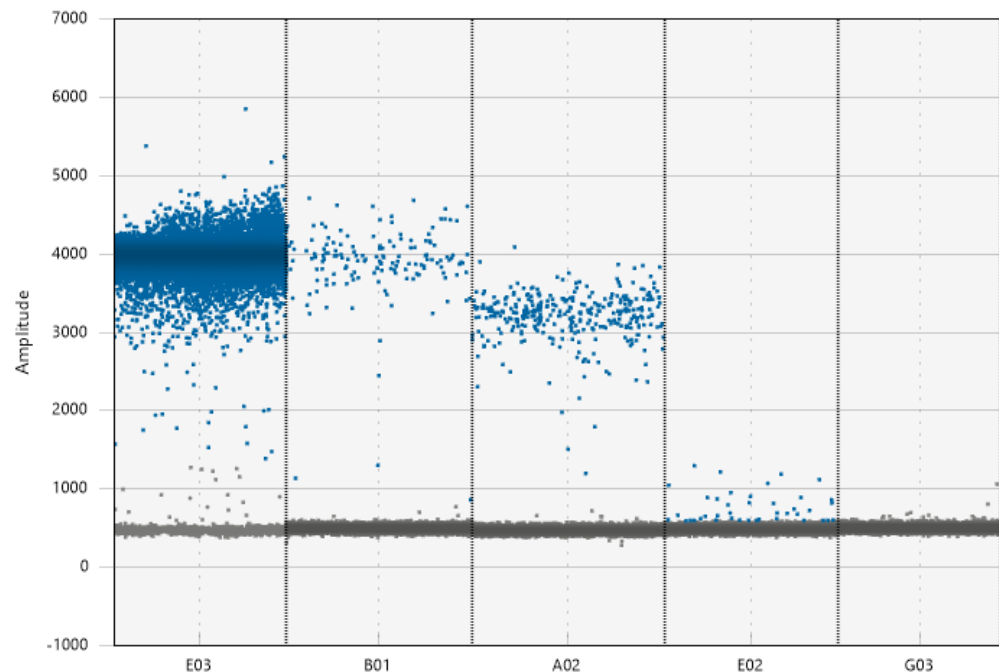


Figure 6. A high threshold (E03) may exclude positive droplets in positive environmental samples which may have reduced amplitude levels (A02). A medium threshold, based on the non-positive environmental samples, could avoid non-specific amplification but also include positive samples with lower amplitude due to inhibition or high target quantity. Setting a low threshold based on the NTC (common practice) can erroneously identify negative environmental samples as positives (E03) due to low level of non-specific amplification. The positive environmental is root DNA sample while the positive environmental 2 is a field soil DNA sample.

In addition to threshold amplitude, we used four sample categories “Environmental negative”, “No-template control” (NTC), “Organismal positive” and “Unknown environmental” to examine how the manual threshold determination based on multiple controls (define also as expected value) compares in output to the available threshold algorithms (Fig. 7). Droplets were assigned to negative (NEG), positive (POS) and rain (RAIN). Only *Define the rain* assigned droplets into the rain category. For the environmental negative control and the unknown environmental sample, none of the automatic threshold determination algorithms matched the expected value as defines by the manual threshold based on multiple controls.

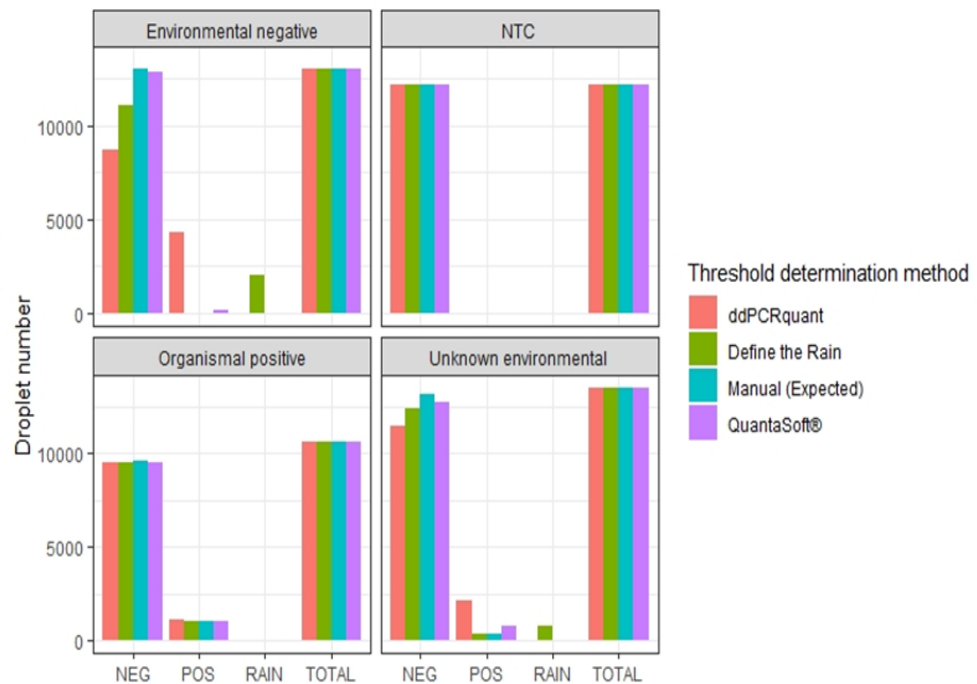


Figure 7. Comparing automatic threshold determination for algorithms and manual threshold determination. We used four sample categories “Environmental negative”, “No-template control” (NTC), “Organismal positive” and “Unknown environmental” to examine how the manual threshold determination based on multiple controls (define also as expected value) compares in output to the available threshold algorithms. Droplets are assigned to negative (NEG), positive (POS) and rain (RAIN).

4. Discussion

The results are presented as QuantaSoft software output since this is the only available software coupled to the ddPCR system and it is the first exposure to data produced by ddPCR. To facilitate interpretation of the figures we created an explanatory model figure (Fig. S1). The figure presented in this study aim to illustrate the effectiveness of the optimization step on example environmental samples that contain significant amount of rain and do not demonstrate conclusive tests.

4.1 Cycling conditions

Adjusting thermocycling conditions, such as altering the duration of denaturation and annealing/extension [37], increasing the number of cycles [39] and decreasing the ramp rate [38] can improve the efficiency of the reaction, which will result in improved droplet clustering and fluorescence levels (Table 1). While altering individual steps has been shown to optimize the ddPCR reaction, the additive effect of combined thermocycling condition alteration in a single protocol remained unknown. We demonstrate that the influence of each step is additive and while a fully optimized protocol can increase the duration of the PCR reaction (e.g., 4 – 5 hours) it can greatly improve droplet resolution for environmental samples (Fig 2). In particular, optimization of annealing and extension temperature may have significant effects on cloud separation. While in standard PCR protocols, the annealing/extension temperature is optimized using target sequence positive and no-template controls, with environmental samples we recommend the use of two additional controls: an **environmental positive** and **environmental negative control** (Fig 3).

Unlike typical positive controls, which are represented by DNA from pure cultures or cloned template, **environmental positive controls** represent DNA samples where the

target organism has been added to the matrix from which you hope to distinguish it. For example, if the target were a soil fungus, similar to our experiment, the environmental control would constitute soil which had been inoculated with the target fungus. **Environmental positive controls** will have less efficient amplification and lower fluorescence of positive droplets compared to organismal positive controls comprised of laboratory material due to the presence of inhibitors in environmental samples. Thus, they allow better assessment of reasonable thresholds for positive droplets for environmental samples (**Fig 3a and b**). **Environmental negative controls** allow discrimination among positives and non-target sequences from the environmental matrix. These controls represent the substrate alone, without target DNA. They are particularly useful when using low annealing temperatures, which can favour non-specific amplification. An environmental negative control is useful for separating such false positives from the droplet cloud (**Fig 3**).

4.2 Primer/probe concentration

Increasing the quantity of primer and probe above recommended guidelines does not affect the amplitude levels or clustering efficiency [38]. However, while decreasing the quantity of primer/probe may reduce the cost of the reaction, its effect on assay efficiency has not been examined. We found that reduction of primer/probe quantity in the reaction may decrease clustering and should be avoided, as it increases the risk for an erroneously set threshold. We suggest working with the recommended quantity of primer and probe (20x which yields 500 nM primers and 250 nM probe). (**Fig 4**).

4.3 Sample concentration

Diluting template DNA can result in tighter clustering because of increased reaction efficiency through decreased inhibitor concentration [34]. It is difficult to determine optimum DNA concentration for environmental samples *a priori*, since samples are comprised of DNA from organisms in addition to the target. It is therefore often necessary to optimize DNA concentration for samples individually. In our example, we show an environmental sample with low-level amplitude droplets that may be either non-target amplification, or incomplete amplification of target DNA due to inhibitors (**Fig 5a – 3**). Diluting such samples may resolve ambiguous droplets (**Fig 5a – 3 after dilution**). If dilution does not resolve low-level amplification, we recommend that an environmental sample can be spiked with a low, known quantity of the target DNA. If the introduced target DNA does not amplify, then inhibition is likely the cause of the low-level droplets, and they are most likely true positives. If the introduced DNA amplifies above the environmental sample, then inhibition is not a problem, and the intermediate droplets are probably non-target DNA (**Fig 5b**).

4.4 Threshold determination

For assays with high target abundance and little rain, threshold determination (manually or automatically) should not affect assay outcomes (BioRad Technologies, Inc. 2018). For environmental samples, however, discrimination of low abundance target from environmental controls is necessary to avoid false calls and spurious interpretation. For example, **setting a low threshold based on NTCs** can lead to false positives since it is not unusual to observe low-level, non-specific amplification in environmental samples (**Fig 6**). Similarly, **setting a high threshold based on positive controls** can underestimate target quantity (**Fig 6**) since positive droplets from environmental samples might exhibit lower fluorescence levels due to inhibition (e.g., Fig 6, well A02). We propose a manual threshold approach by setting a threshold that excludes droplets from **negative environmental controls** and includes the reduced fluorescence positive cloud produced by **positive environmental controls** (**Fig 6**).

Our analysis demonstrates that 'Define The Rain' is the most accurate method for determining thresholds. 'Define The Rain' produced results similar to the manually determined threshold based on multiple controls (**Fig. 7**) (**Fig S2**). However, 'Define The Rain' identifies droplets as 'rain', meaning the classification of the rain to positive or negative relies on the researcher's judgment. In samples with a high proportion of "rain" droplets a manual approach might be as effective.

QuantaSoft® software automatic threshold failed to determine between the environmental negative and true positive, and erroneously calculated the targets quantity (**Fig. 7**). While the threshold was determined after considering the group of samples, the threshold appeared random in each sample (**Fig S3**). While the QuantaSoft® software algorithm is the most commonly algorithm used (see Table S1), it should be avoided for environmental samples.

'ddPCRquant' algorithm poorly estimated target quantity and failed to identify the environmental negative as true negative (**Fig. 7**). The 'ddPCRquant' algorithm simply merged the three NTS's and draw a line right above the dense droplet cloud created from the merged sample. Every droplet above that 'low threshold' was considered as positive (**Fig S4**). Setting a low threshold, based on non-template controls, can misestimate the targets quantity, and identify false positives droplets as true positives (see section "Threshold determination"). 'ddPCRquant' should only be used when the negative cloud amplitude is the same as the NTCs cloud. Because this is rarely the case for environmental samples, 'ddPCRquant' algorithm should be avoided for environmental samples.

4.5 Limit of detection determination

Even after the threshold determination, the possibility of false positives exists, especially for rare environmental targets. When samples have a low number of droplets above the determined threshold, they may be considered positives via misinterpretation of fluorescence of foreign particles or artefacts produced by the instruments [40,41,42]. In order to define these droplets, knowledge on instrument limit of detection (LOD) is needed. The majority of strategies used to determine this threshold result in misestimation of targets presence in environmental samples [40]. In order to define the LOD a dilution series of a known quantity positive control is needed [26,40,43]. The purpose is to start the reaction with very low number of known target quantity (e.g. 1 to 10 copies per 20 mL PCR reaction) and examine the LOD of the instrument. Regardless the options, LOD threshold should be clearly defined in studies that use ddPCR. While using algorithms allows for rapid analysis of multiple samples, they cannot compensate for a poor assay. All assays, particularly those with significant rain (e.g., environmental samples), should attempt to reduce the rain through assay optimization before threshold determination.

5. Conclusions

Regardless of target, for both clinical and environmental samples the single most important strategy to reduce 'rain' is proper assay optimization. Even after meticulous assay optimization, however, most environmental samples will contain significant levels of intermediate droplets. It is important to determine whether these droplets represent non-target organisms (and are negatives) or reduced amplification efficiency (and are actually positives). Unfortunately, there is no easy way to discriminate among these possibilities. Threshold determination should result from a series of decisions spanning assay conditions to threshold analysis. Importantly, the threshold that is ultimately used should be a single value, not a range of values. A single, fixed manually determined threshold through the use of appropriate environmental controls prevents subjectivity and allows straightforward interpretation by the reader.

Supplementary Materials: The following are available online at www.mdpi.com/xxx/s1, Figure S1: A key to interpreting ddPCR output figures. Figure S2. Manual threshold determination, Figure S3. Automatically determined threshold using the QuantaSoft® software, Table S1. Web of Science (Clarivate Analytics 2018) search for environmental studies using ddPCR.

Author Contributions: VK and MMH designed the research, VK performed the research, and all co-authors wrote the paper.

Funding: This work was supported by the National Sciences and Engineering Research Council of Canada discovery grant (2018).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hindson, B. J., Ness, K. D., Masquelier, D. A., Belgrader, P., Heredia, N. J., Makarewicz, A. J., ... Colston, B. W. (2011). High-Throughput Droplet Digital PCR System for Absolute Quantitation of DNA Copy Number. *Analytical Chemistry*, 83(22), 8604–8610. doi: 10.1021/ac202028g
2. Basu, A. S. (2017, August 1). Digital Assays Part I: Partitioning Statistics and Digital PCR. *SLAS Technology*, Vol. 22, pp. 369–386. SAGE Publications Inc. doi: 10.1177/2472630317705680
3. Hindson, C. M., Chevillet, J. R., Briggs, H. A., Gallichotte, E. N., Ruf, I. K., Hindson, B. J., ... Tewari, M. (2013). Absolute quantification by droplet digital PCR versus analog real-time PCR. *Nature Methods*, 10(10), 1003–1005. doi: 10.1038/nmeth.263
4. Beaver, J. A., Jelovac, D., Balukrishna, S., Cochran, R. L., Croessmann, S., Zabransky, D. J., ... Park, B. H. (2014). Detection of cancer DNA in plasma of patients with early-stage breast cancer. *Clinical Cancer Research*, 20(10), 2643–2650. doi: 10.1158/1078-0432.CCR-13-2933
5. Liu, R., Zhao, X., Guo, W., Huang, M., Qiu, L., Zhang, W., ... Chen, Z. (2020). Dynamic monitoring of HER2 amplification in circulating DNA of patients with metastatic colorectal cancer treated with cetuximab. *Clinical and Translational Oncology*, 22(6), 928–934. doi: 10.1007/s12094-019-02215-7
6. Fu, W., Wang, C., Zhu, P., Xu, W., Li, X., & Zhu, S. (2020). A universal analytical approach for screening and monitoring of authorized and unauthorized GMOs. *LWT*, 125, 109176. doi: 10.1016/j.lwt.2020.109176
7. Milavec, M., Dobnik, D., Yang, L., Zhang, D., Gruden, K., & Žel, J. (2014, August 16). GMO quantification: Valuable experience and insights for the future. *Analytical and Bioanalytical Chemistry*, Vol. 406, pp. 6485–6497. Springer Verlag. doi: 10.1007/s00216-014-8077-0
8. Hamza, I. A., & Bibby, K. (2019, April 1). Critical issues in application of molecular methods to environmental virology. *Journal of Virological Methods*, Vol. 266, pp. 11–24. Elsevier B.V. doi: 10.1016/j.jviromet.2019.01.008
9. Rački, N., Dreo, T., Gutierrez-Aguirre, I., Blejec, A., & Ravnikar, M. (2014). Reverse transcriptase droplet digital PCR shows high resilience to PCR inhibitors from plant, soil and water samples. *Plant Methods*, 10(1). doi: 10.1186/s13007-014-0042-6
10. Bustin, S. A., Benes, V., Garson, J. A., Hellemans, J., Huggett, J., Kubista, M., ... Wittwer, C. T. (2009). The MIQE guidelines: Minimum information for publication of quantitative real-time PCR experiments. *Clinical Chemistry*, 55(4), 611–622. doi: 10.1373/clinchem.2008.112797
11. Yang, R., Paparini, A., Monis, P., Ryan, U., 2014. Comparison of next-generation droplet digital PCR (ddPCR) with quantitative PCR (qPCR) for enumeration of *Cryptosporidium* oocysts in faecal samples. *Int. J. Parasitol.* 44, 1105–1113. <https://doi.org/10.1016/j.ijpara.2014.08.004>
12. Demeke, T., & Dobnik, D. (2018). Critical assessment of digital PCR for the detection and quantification of genetically modified organisms. *Analytical and Bioanalytical Chemistry*, 410(17), 4039–4050. doi: 10.1007/s00216-018-1010-1
13. Trypsteen, W., Vynck, M., de Neve, J., Bonczkowski, P., Kiselina, M., Malatinkova, E., ... De Spiegelaere, W. (2015). ddpcRquant: threshold determination for single channel droplet digital PCR experiments. *Analytical and Bioanalytical Chemistry*, 407(19), 5827–5834. doi: 10.1007/s00216-015-8773-4
14. Zink, F. A., Tembrock, L. R., Timm, A. E., Farris, R. E., Perera, O. P., & Gilligan, T. M. (2017). A droplet digital PCR (ddPCR)

- assay to detect *Helicoverpa armigera* (Lepidoptera: Noctuidae) in bulk trap samples. *PLoS ONE*, 12(5). doi: 10.1371/journal.pone.0178704
15. Dobnik, D., Demšar, T., Huber, I., Gerdes, L., Broeders, S., Roosens, N., ... Žel, J. (2018). Inter-laboratory analysis of selected genetically modified plant reference materials with digital PCR. *Analytical and Bioanalytical Chemistry*, 410(1), 211–221. doi: 10.1007/s00216-017-0711-1
 16. Pipan, B., Zupančič, M., Blatnik, E., Dolničar, P., & Meglič, V. (2018). Comparison of six genomic DNA extraction methods for molecular downstream applications of apple tree (*Malus X domestica*). *Cogent Food & Agriculture*, 4(1). doi: 10.1080/23311932.2018.1540094
 17. Meijerink, J. P. P., Mandigers, C., Van De Locht, L., Tönnissen, E., Goodsaid, F., & Raemaekers, J. (2001). A novel method to compensate for different amplification efficiencies between patient DNA samples in quantitative real-time PCR. *Journal of Molecular Diagnostics*, 3(2), 55–61. doi: 10.1016/S1525-1578(10)60652-6
 18. Emerson, J. B., Adams, R. I., Román, C. M. B., Brooks, B., Coil, D. A., Dahlhausen, K., ... Rothschild, L. J. (2017, August 16). Schrödinger's microbes: Tools for distinguishing the living from the dead in microbial ecosystems. *Microbiome*, Vol. 5, p. 86. doi: 10.1186/s40168-017-0285-3
 19. Kreader, C. A. (1996). Relief of amplification inhibition in PCR with bovine serum albumin or T4 gene 32 protein. *Applied and Environmental Microbiology*, 62(3), 1102–1106.
 20. Dingle, T. C., Sedlak, R. H., Cook, L., & Jerome, K. R. (2013, November). Tolerance of droplet-digital PCR vs real-time quantitative PCR to inhibitory substances. *Clinical Chemistry*, Vol. 59, pp. 1670–1672. doi: 10.1373/clinchem.2013.211045
 21. Taylor, S. C., Laperriere, G., & Germain, H. (2017). Droplet Digital PCR versus qPCR for gene expression analysis with low abundant targets: From variable nonsense to publication quality data. *Scientific Reports*, 7(1), 1–8. doi: 10.1038/s41598-017-02217-x
 22. Cavé, L., Brothier, E., Abrouk, D., Bouda, P. S., Hien, E., & Nazaret, S. (2016). Efficiency and sensitivity of the digital droplet PCR for the quantification of antibiotic resistance genes in soils and organic residues. *Applied Microbiology and Biotechnology*, 100(24), 10597–10608. doi: 10.1007/s00253-016-7950-5
 23. Wickham H (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-3-319-24277-4, <https://ggplot2.tidyverse.org>.
 24. Kokkoris, V., Li, Y., Hamel, C., Hanson, K., & Hart, M. (2019). Site specificity in establishment of a commercial arbuscular mycorrhizal fungal inoculant. *Science of The Total Environment*, 660, 1135–1143. doi: 10.1016/j.scitotenv.2019.01.100
 25. Sieverding, E., Berndt, R., Oehl, F., da Silva, G. A., Berndt, R., & Oehl, F. (2014). Rhizoglosum, a new genus of the Glomeraceae. *Mycotaxon*, 129(2), 373–386. doi: 10.5248/129.373
 26. Doi, H., Uchii, K., Takahara, T., Matsushashi, S., Yamanaka, H., & Minamoto, T. (2015). Use of droplet digital PCR for estimation of fish abundance and biomass in environmental DNA surveys. *PLoS ONE*, 10(3). doi: 10.1371/journal.pone.0122763
 27. Simmons, M., Tucker, A., Chadderton, W. L., Jerde, C. L., & Mahon, A. R. (2015). Active and passive environmental DNA surveillance of aquatic invasive species. *Canadian Journal of Fisheries and Aquatic Sciences*, 73(1), 76–83. doi: 10.1139/cjfas-2015-0262
 28. Cooley, M. B., Carychao, D., & Gorski, L. (2018). Optimized co-extraction and quantification of DNA from enteric pathogens in surface water samples near produce fields in California. *Frontiers in Microbiology*, 9(MAR). doi: 10.3389/fmicb.2018.00448
 29. Martínez-Diz, M. del P., Andrés-Sodupe, M., Berbegal, M., Bujanda, R., Díaz-Losada, E., & Gramaje, D. (2020). Droplet Digital PCR Technology for Detection of *Ilyonectria liriodendri* from Grapevine Environmental Samples. *Plant Disease*, 104(4), 1144–1150. doi: 10.1094/PDIS-03-19-0529-RE
 30. Rosa, D., Pogiatis, A., Bowen, P., Kokkoris, V., Richards, A., Holland, T., & Hart, M. (2020). Performance and Establishment

- of a Commercial Mycorrhizal Inoculant in Viticulture. *Agriculture*, 10(11), 539. doi: 10.3390/agriculture10110539
31. Kokkoris, V., Chagnon, P.-L., Yildirim, G., Clarke, K., Goh, D., MacLean, A. M., ... Corradi, N. (2021). Host identity influences nuclear dynamics in arbuscular mycorrhizal fungi. *Current Biology*. doi: 10.1016/j.cub.2021.01.035
 32. McMahon, T. C., Blais, B. W., Wong, A., & Carrillo, C. D. (2017). Multiplexed single intact cell droplet digital PCR (MuSIC ddPCR) method for specific detection of enterohemorrhagic E. coli (EHEC) in food enrichment cultures. *Frontiers in Microbiology*, 8(MAR). doi: 10.3389/fmicb.2017.00332
 33. Jones, M., Williams, J., Gärtner, K., Phillips, R., Hurst, J., & Frater, J. (2014). Low copy target detection by Droplet Digital PCR through application of a novel open access bioinformatic pipeline, “definetherain.” *Journal of Virological Methods*, 202, 46–53. doi: 10.1016/j.jviromet.2014.02.020
 34. Dreo, T., Pirc, M., Ramšak, Ž., Pavšič, J., Milavec, M., Žel, J., & Gruden, K. (2014). Optimising droplet digital PCR analysis approaches for detection and quantification of bacteria: A case study of fire blight and potato brown rot. *Analytical and Bioanalytical Chemistry*, 406(26), 6513–6528. doi: 10.1007/s00216-014-8084-1
 35. Strain, M. C., Lada, S. M., Luong, T., Rought, S. E., Gianella, S., Terry, V. H., ... Richman, D. D. (2013). Highly Precise Measurement of HIV DNA by Droplet Digital PCR. *PLoS ONE*, 8(4), e55943. doi: 10.1371/journal.pone.0055943
 36. Jacobs, B. K. M., Goetghebeur, E., Vandesompele, J., De Ganck, A., Nijs, N., Beckers, A., ... Clement, L. (2017). Model-Based Classification for Digital PCR: Your Umbrella for Rain. *Analytical Chemistry*, 89(8), 4461–4467. doi: 10.1021/acs.analchem.6b04208
 37. Witte, A. K., Mester, P., Fister, S., Witte, M., Schoder, D., & Rossmannith, P. (2016). A systematic investigation of parameters influencing droplet rain in the *Listeria monocytogenes* prfA assay -reduction of ambiguous results in ddPCR. *PLoS ONE*, 11(12), 1–13. doi: 10.1371/journal.pone.0168179
 38. Witte, A. K., Fister, S., Mester, P., Schoder, D., & Rossmannith, P. (2016). Evaluation of the performance of quantitative detection of the *Listeria monocytogenes* prfA locus with droplet digital PCR. *Analytical and Bioanalytical Chemistry*, 408(27), 7583–7593. doi: 10.1007/s00216-016-9861-9
 39. Köppel, R., & Bucher, T. (2015). Rapid establishment of droplet digital PCR for quantitative GMO analysis. *European Food Research and Technology*, 241(3), 427–439. doi: 10.1007/s00217-015-2475-1
 40. Hunter, M. E., Dorazio, R. M., Butterfield, J. S. S., Meigs-Friend, G., Nico, L. G., & Ferrante, J. A. (2017). Detection limits of quantitative and digital PCR assays and their influence in presence–absence surveys of environmental DNA. *Molecular Ecology Resources*, 17(2), 221–229. doi: 10.1111/1755-0998.12619
 41. Kiselina, M., Pasternak, A. O., De Spiegelaere, W., Vogelaers, D., Berkhout, B., & Vandekerckhove, L. (2014). Comparison of droplet digital PCR and seminested real-time PCR for quantification of cell-associated HIV-1 RNA. *PLoS ONE*, 9(1). doi: 10.1371/journal.pone.0085999
 42. Yang, R., Paparini, A., Monis, P., & Ryan, U. (2014). Comparison of next-generation droplet digital PCR (ddPCR) with quantitative PCR (qPCR) for enumeration of *Cryptosporidium* oocysts in faecal samples. *International Journal for Parasitology*, 44(14), 1105–1113. doi: 10.1016/j.ijpara.2014.08.004
 43. Baker, C. S., Steel, D., Nieukirk, S., & Klinck, H. (2018). Environmental DNA (eDNA) from the wake of the whales: Droplet digital PCR for detection and species identification. *Frontiers in Marine Science*, 5(APR). doi: 10.3389/fmars.2018.00133