Electronic Supplementary Material (ESI) for Lab on a Chip.

Supporting Information

Optimized droplet digital assay (ddCFU) provides precise quantification of bacteria over dynamic range of 6 logs and beyond

Ott Scheler[ab], Natalia Pacocha[a], Pawel R. Debski[a], Artur Ruszczak[a], Tomasz S. Kaminski[a] and Piotr Garstecki[a]*

[a]Institute of Physical Chemistry, Polish Academy of Sciences, Kasprzaka 44/52, 01-224 Warsaw, Poland

[b]Institute of Molecular and Cell Biology, University of Tartu, Riia 23 51010 Tartu, Estonia

*corresponding author: *garst@ichf.edu.pl*

Note S1. How to design and understand the results of ddCFU assay

This instruction details how to analyze the readout from an optimized *rational droplet digital* assay described in this manuscripts. In digital methods, the positive signal yielded by a compartment or partition $d_i v_i$ is interpreted to mean, that in the said partition the number of molecules of the analyte is at least one. In order to translate this information into a useful form of the probability distribution of initial concentration having caused the recorded result, we start with the probability of obtaining the positive signal. This probability is a function of initial concentration C and the properties of the compartment (i.e. $d_i v_i$):

$$p_i(s_i = 1 \mid C) = 1 - e^{-Cd_i v_i}$$

A negative signal yielded by a compartment or partition $d_i v_i$ states, that in the said partition there were no particles of the analyte, which can be interpreted with another probability function of C:

$$p_i(s_i = 0|C) = e^{-Cd_i v_i}$$

To analyze the results of the assay, first, the binary values (digital readouts) of $\{s_i\}$ for the partitions belonging to all the libraries $\{d_i v_i\}$ must be collected.

The basis of the design *rational droplet digital* assays is setting the sub-assays of compartments in a geometric sequence of volumes v and/or dilutions d of compartments belonging to consecutive-sub-assays The change of the product dv effectively changes (for a fixed concentration) the expected number of CFUs per one compartment: $m_{CFU} = Cdv$.

Therefore, in order to prepare a *rational droplet digital* assays a series sub-assays is produced so that the corresponding series of expected numbers of CFUs per single compartment is a geometric series with quotient x. The said series can be obtained as follows:

1. By changing the volumes v_i of compartments belonging to consecutive sub-assays so that they form a geometric series with quotient x: $v_i = v_o z_i = v_o x^{i-1}$.
2. By changing the dilutions d_i of compartments belonging to consecutive sub-assays so that they form a geometric series with quotient x: $d_i = d_o z_i = d_o x^{i-1}$.
3. More generally, one can use a combination of methods 1) and 2), in this case the products of consecutive dilutions and volumes $d_i v_i$ are adjusted so as to form a geometric series with quotient x: $d_i v_i = d_o v_o z_i = d_o v_o x^{i-1}$.

Then, knowing the design of a rational droplet digital assay, the analysis is done as follows:

1. From every positive partition (i.e. $s_i = 1$) with modification factor $d_i v_i$ we construct the probability function p_i:
   $p_i(s_i = 1 \mid C) = 1 - e^{-Cd_i v_i}$
2. We also construct probabilities of obtaining negative signals for all negative compartments:
   $p_i(s_i = 0|C) = e^{-Cd_i v_i}$
3. Then, we calculate the probability $P(\{s_i\}|C)$ of obtaining the recorded state of the *rational design* digital assay, which is a product of the probability functions for all the compartments:
   $$\mathrm{P}(\{s_i\} \mid C) = \prod_{i=0}^{N-1} p_i(s_i \mid C)$$
4. This can be also written as:
   $$\mathrm{P}(\{s_i\} \mid C) = \prod_{i=0}^{N-1} \left\{ \left(1 - e^{-Cd_i v_i}\right)^{s_i} \cdot \left(e^{-Cd_i v_i}\right)^{1-s_i} \right\}$$
   Where $i = 0,\ldots,N-1$ count all the compartments of the assay

5. Then, we use Bayesian formalism to invert the product probability to the probability distribution $P(C|\{s_i\})$ of the initial concentration C (under the condition that no *a priori* information about the distribution of C is available):
   $$P(C \mid \{s_i\}) = \mathrm{P}(\{s_i\} \mid C) / \int_0^\infty \mathrm{P}(\{s_i\} \mid C) dC$$
   for practical reasons (numerical calculation of the integral), the upper limit of integration can be finite, but should be at least one order of magnitude larger than the upper limit of the dynamic range of the assay

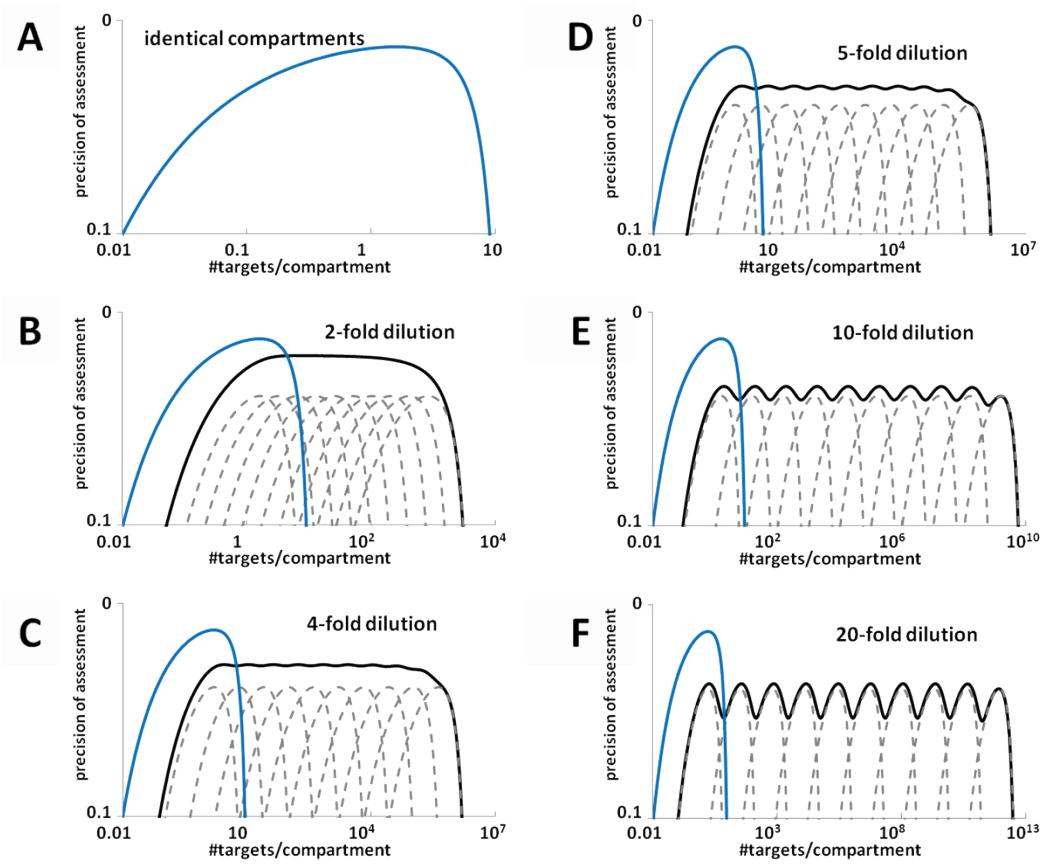
6. Knowing the distribution $P(C \mid \{s_i\})$, one can estimate the initial concentration of the sample as the expected value of this distribution:
   $$E(C) = \int_0^\infty C \cdot \mathrm{P}(C|\{s_i\}) dC$$
   again, the upper limit of integration can be finite, but should be at least one order of magnitude larger than the upper limit of the dynamic range of the assay
7. The precision of the estimate can be calculated as the spread (relative standard deviation) of the distribution $P(C \mid \{s_i\})$:
   $\sigma(C) = \sqrt{E(C^2) - (E(C))^2} / E(C)$

where

$$E(C^2) = \int_0^\infty C^2 \cdot \mathrm{P}(C|\{s_i\})dC$$

Additionally, it is useful to check the performance of the designed assay numerically. For example, one can run Monte Carlo simulations that take as an input the initial concentration of the analyte (chosen randomly from the dynamic range of the assay), then using random number generators determine the number of molecules of the analyte in every single partition (which has a Poisson distribution with $\lambda = Cd_iv_i$), marking positive and negative ones, and then calculate the estimate of the initial concentration using the instructions given in 1-6. As a result, one gets the dependence between the initial concentrations and calculated estimated concentration, which is usually 1:1. In some cases, however (especially for small assays), it is slightly tilted. Knowing this dependence from multiple MC simulations, one can use it as a correction function for the estimate of initial concentration from real experiments.

Detailed analysis of the precision of RDD assays



The comparison between the precision provided by different RDD assays (b-f) and the precision provided by a single-volume digital assay (a). The blue lines describe the performance of a single-volume digital assay which comprises of 10,000 identical compartments. The black lines describe the

performance of various RDD assays, each comprising of 10 sub assays having 1,000 compartments (10,000 compartments in total). The grey dashed line describe the behaviour of every sub-assay comprising the RDD assay. The ratio of dilutions between the compartments belonging to consecutive compartments equals 2-fold (b), 4-fold (c), 5-fold (d), 10-fold (e), and finally 20-fold (f). It is worth noticing, that if the ratio of dilutions between of compartments belonging to consecutive sub-assays is small, the functions describing precision overlap, which results in the flattening of the distribution of the final precision. If the ratio is increased (in order to cover wider dynamic range of the estimate), the functions overlap less, which results in the wavy behaviour of the observed precision. Still, amplitude of so-called waves is relatively small, and therefore do affect significantly the results.
Here we discuss in detail the behaviour of the precision of assessment provided by the RDD assay. The design of the RDD assay is based on arranging a set of classic, single-volume sub-assays that differ in dilution and/or volume of compartments (the ratio of dilutions and/or volumes of compartments belonging to consecutive sub-assays is fixed, which results in a geometric sequence of volumes/dilutions.
All the sub-assays are used simultaneously for the assessment of the number of targets. The variance of the estimate of the initial number of targets equals the sum of variances provided by all the sub-assays. The precision of the assessment used in this paper equals the relative standard deviation of the estimate, i.e. the square root of the total variance divided by the estimated number of targets.
As a result, the relative precision of the estimate is uniform in the whole dynamic range of assessment because it comes from uniformly distributed digital assays. If the difference between the consecutive sub-assays is small, i.e. the dilution ratio is close to unity, we obtain an almost flat precision. However, if the dilution ratio is higher, i.e. the sub-assays differ significantly, the functions describing precision provided by consecutive sub-assays do not overlap, and therefore the final function shares some of the characteristics of component precisions, which results in its wavy behavior.

Details of the Monte Carlo simulations

The Monte Carlo simulations were used to establish the 'design' formulas for droplet digital assays. They are based on the methods described in detail in the paper: Debski, P.R., and Garstecki, P., *Designing and interpretation of digital assays: concentration of target in the sample and in the source of sample*, Biomolecular Detection and Quantification, 2016, 10, 24-30. There were run by means of the random number generators provided by the ROOT framework (https://root.cern.ch/). We used the terms canonical and grand canonical Monte Carlo simulations analogically to canonical and grad canonical ensembles to differentiate between two situations:

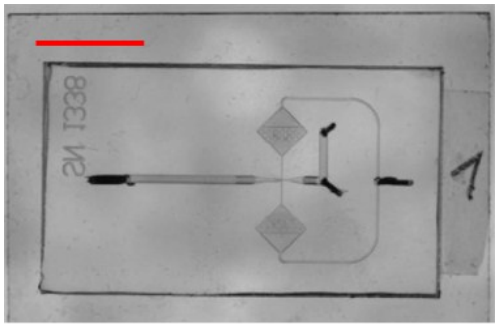
i) The grand canonical Monte Carlo simulations used as an input the initial concentration C of the analyte (in this case, the concentration of CFUs) in the assay. Then, each compartment with dilution d and volume v was treated individually, i.e. the each compartment gained randomly a positive signal with probability $p(s=1) = 1 - e^{-Cdv}$ or negative signal with probability $p(s=0) = e^{-Cdv}$. Therefore, there was a randomness of the distribution of the number of CFUs M_A in the assay (i.e. Poisson distribution with expected value CV_{assay}) and the distribution of these molecules among the compartments. These simulations were used to produce probability distributions $p(K|C)$ that were later used for Bayes' method. This method was used for the analysis of the results of droplet digital CFU assay presented in this work.

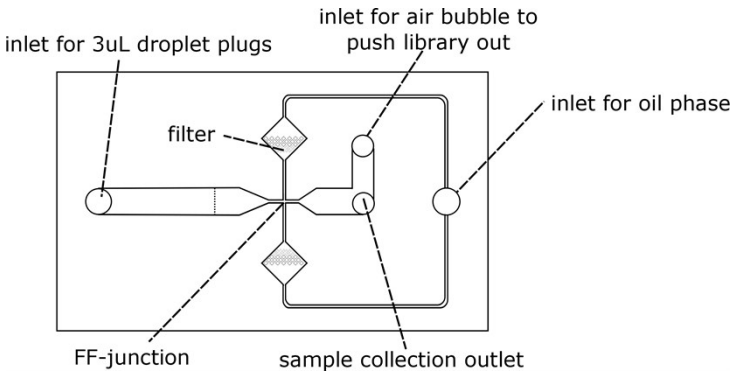
ii) The canonical Monte Carlo simulations used as an input the hard-fixed initial number M_A of CFUs in the assay. Then, each CFU was randomly distributed among the compartments with uniform probability. At the end, the compartments containing at least one molecule of the analyte were given positive signal or negative signal otherwise. Therefore, there was a randomness only in the distribution of the CFUs among the compartments. These simulations were used to produce probability distributions $p(K|M_A)$ that were later used for Bayes' method. They provide the assessment of the initial number of CFUs in the sample, however, they underestimate the error of the initial concentration in the source, and therefore they were not used in this work.

Note S2. Schematics of i) droplet generation chip and ii) droplet counting chip

Red scale bar on figures is 1cm. Schematics on right are representative and do not show exact size or proportions.

i. Droplet generation chip

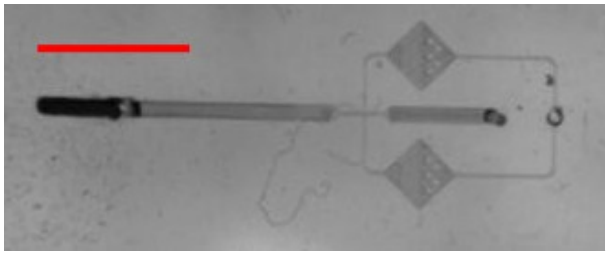


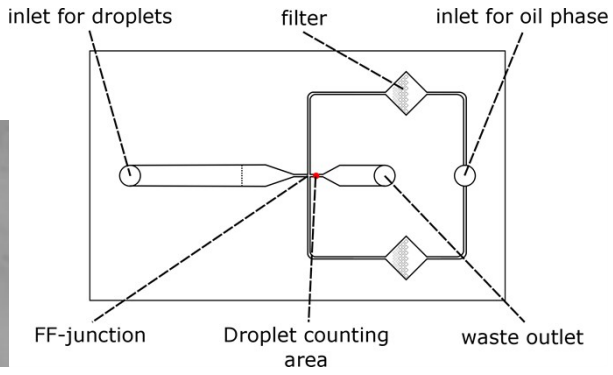


Channels have following dimensions (width x height):

- Main inlet channel: start w:800 x h:800 µm, end at vertical bar position 100x120 µm
- Outlet channel: 800x800 µm
- Oil delivering channel: 200x200 µm
- Flow-focusing junction: 100x120 µm

ii. Droplet counting chip





Channels have following dimensions (width x height):

- Droplet inlet channel: 1200x1200 µm at inlet, 1200x100 at FF-junction (linear depth change until vertical dashed line)
- Flow-focusing junction: 124x100 µm
- Secondary oil delivering channel: 114x100 µm

Note S3. Droplet size and dispersity calculations

In order to measure the size dispersion of the droplets generated from 3µL plugs, we recorded the droplet formation using Photron Fast-Cam 1000K (Japan) camera that was mounted on Nikon SMZ1000 stereoscope (Japan). We measured the whole droplet formation process of a 3µL plug. We picked randomly five different plugs from three different dilution series.

	Average of 5 plugs	Plug 1	Plug 2	Plug 3	Plug 4	Plug 5
Average droplet volume [nL]	**1.00**	1.05	1.07	0.92	1.00	0.97
Number of droplets generated	**3001.4**	2850	2792	3273	3011	3081
Coefficient of variation	**5.63%**	3.08%	3.37%	1.98%	2.86%	2.21%

We measured the area of each generated droplet in Image J software and used this data to calculate the volume of each droplet according to the model described previously (1-3). Briefly, the volume can be calculated by the following formula: $V=(\pi/12)[2D^3-(D-h)^2(2D+h)]$, where D is the diameter of a droplet and h is the height of the channel.

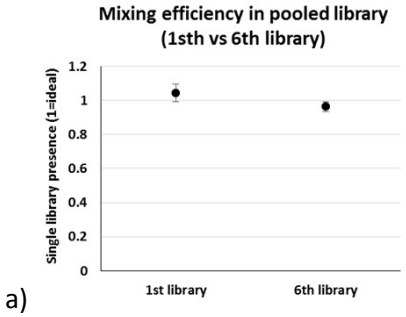
(1) Nie *et al*., Microfluid. Nanofluidics 2008, 5, 585–594.
(2) Li *et al*., Soft Matter 2008, 4 (2), 258.
(3) Kaminski *et al*., Lab Chip 2012, 12 (20), 3995.

Note S4. Library mixing efficiency of pooled droplet libraries in ddCFU assay

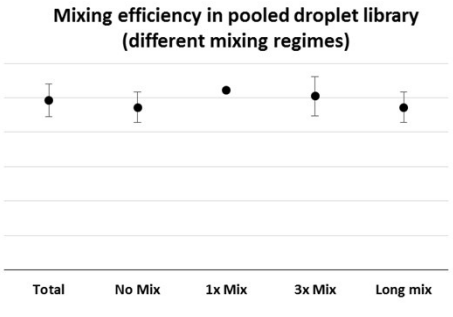
In our ddCFU assay experiments we collected droplet libraries together in 1.5 mL test-tube for incubation. After the incubation we did not analyze all of the ~33000 droplets as some fraction was always lost during transfer and handling. Thus we investigated if the pooled droplet libraries are mixed enough for precise analysis.

In order to do that we prepared 11 3µL plugs and split them into 1nL droplets as in ddCFU assay. One of the libraries was labelled with 25 µg/mL Dextran Cascade Blue (Molecular Probes, Life Technologies, USA). We then tested if that labelled population of droplets is present in the whole droplet population as 1/11 fraction. In our experiments we called this ideal situation with value "1". In all cases there are five replicate experiments.

Firstly, we tested if the order of dilution has an effect on single library presence in analysis. We compared two samples where labelled droplets were generated either first or as sixth library in sequence. We mixed the droplets by turning test-tube 90 degrees once both clock- and anticlockwise. We did not observe any significant difference as the labelled fraction was near perfect value 1 in both cases (Fig S4a).



a)



b)

	1st library	6th library
droplets analyzed (AVG)	24263.6	26523.4
labelled fraction (1= ideal)	1.04	0.96
coefficient of variation (CV)	5.07%	2.91%

	Total	No Mix	1x Mix	3x Mix	Long mix
droplets analyzed (AVG)	27381	27950	28158	27522	25895
labelled fraction (1= ideal)	0.98	0.94	1.04	1.01	0.94
coefficient of variation (CV)	7.74%	9.31%	0.88%	11.24%	9.51%

Secondly, we investigated if the mixing regime has an effect on single library presence in analysis. We compared samples where first library was labelled. Different mixing regimes were: i) no mixing, ii) turning test-tube 90 degrees once both clock- and anticlockwise, iii) turning test-tube 90 degrees three times both clock- and anticlockwise and iv) long mixing where test-tube was positioned horizontally on a slow-regime rocker for one minute. We did not observe any significant difference as the labelled fraction was near perfect value 1 in all cases. We did however observed occasionally the slight decrease in droplet stability with libraries that were either rocked or turned (Fig S4b).

Note S5. Data for experiments shown in main text

Here we show the data that was used in i) Comparison with conventional plate counting and ii) Antibiotic time-kill test. At least three replicates were done for each experiment. CV stands for Coefficient of variation (CV).

i) Comparison with conventional plate counting

	ddCFU Assay		**Plate count**		**ddCFU assay/Plate count**
	Average	CV	Average	CV	
Sample I	1.63E+09	2%	2.92E+09	11%	55.93%
Sample II	1.68E+08	14%	3.02E+08	22%	55.68%
Sample III	2.49E+07	18%	2.14E+07	7%	116.26%
Sample IV	2.05E+06	11%	2.71E+06	16%	75.70%
Sample V	2.01E+05	35%	3.04E+05	14%	66.21%
Sample VI	2.13E+04	23%	2.65E+04	10%	80.35%

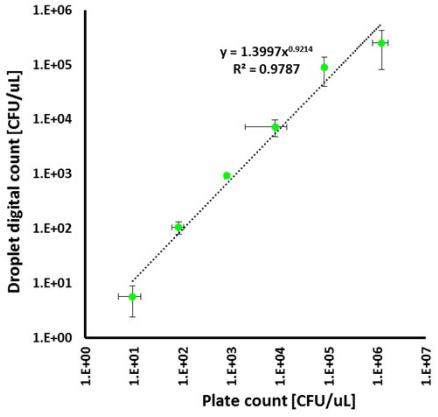
ii) Antibiotic time-kill test. In here the value "1" at the 0h stands for ~$2x10^6$ CFU/mL of viable *E. coli* bacteria that was used in experiment.

Incubation time	**Average**	**CV**
0h	1.00	3.49%
0.5h	0.37	1.14%
1h	0.22	0.78%
2h	0.04	1.10%
4h	0.00	0.00%
8h	0.00	0.00%
25h	0.00	0.00%

Note S6. ddCFU assay with 2-fold dilution libraries

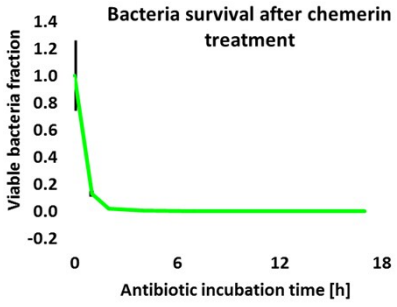
In our experiments we also tested ddCFU assay with 2-fold dilution libraries. In that case we had a sequence of dilutions that consisted of undiluted sample and 20 2-fold dilutions (21 in total). We compared this ddCFU approach with traditional plate counting (i) and time-kill test with antimicrobial peptide p4 from chemerin. Chemerin is a peptide molecule that antimicrobial properties (4-7) and is currently being investigated as potential drug. In our experiments we used 100 µM of p4 chemerin (ThinkPeptides ProImmune, UK).

i) Comparison with conventional plate counting (2-fold dilutions).



	ddCFU assay (2-fold)		Plate count		ddCFU assay/Plate count
	Average	CV	Average	CV	
Sample I	2.52E+05	67.47%	1.23E+06	34.96%	20.51%
Sample II	8.93E+04	55.04%	8.20E+04	8.66%	108.86%
Sample III	7.26E+03	34.23%	7.96E+03	75.38%	91.21%
Sample IV	9.25E+02	7.51%	8.10E+02	8.77%	114.23%
Sample V	1.06E+02	24.98%	8.33E+01	27.85%	126.78%
Sample VI	5.67E+00	58.23%	9.43E+00	48.78%	60.09%

ii) Antimicrobial kill-test with chemerin peptide. In here the value "1" at the 0h stands for ~1.6x10^6 CFU/mL of viable *E. coli* bacteria that was used in experiment. Final concentration of p4 chemerin was 100 µm in PBS with 1% of LB media. After p4 treatment the cells were washed twice with fresh media (LB media with 100 µg/mL ampicillin and 1 mM IPTG) before transferring to fresh media for ddCFU assay.



Incubation	Average	CV
0h	1.0000	26.09%
1h	0.1289	17.15%
2h	0.0174	22.55%
4h	0.0064	13.81%
7h	0.0016	83.77%
17h	0.0002	141.42%

(4) Kulig *et al.*, Journal of Immunology 2011, 187 (3), 1403-1410
(5) Banas *et al.*, PloS One 2013, 8(3), e58709
(6) Zabel *et al.*, Am J Clin Exp Immunol 2014, 3(1), 1-19
(7) Banas *et al.*, PloS One 2015, 10(2), e0117830