

dpcReport: web server and software suite for unified analysis of digital PCRs and digital assays

Michał Burdukiewicz¹, Jim Huggett², Alexandra Whale², Piotr Sobczyk³, Paweł Mackiewicz¹, Andrej-Nikolai Spiess³, Peter Schierack⁵, and Stefan Rödiger⁵

¹University of Wrocław, Department of Genomics,

²Molecular and Cell Biology Team, LGC, Teddington, United Kingdom,

³Wrocław University of Science and Technology, Faculty of Pure and Applied Mathematics,

⁴University Medical Center Hamburg-Eppendorf, Hamburg, Germany,

⁵Brandenburg University of Technology Cottbus-Senftenberg, Institute of Biotechnology

Outline

dPCR software

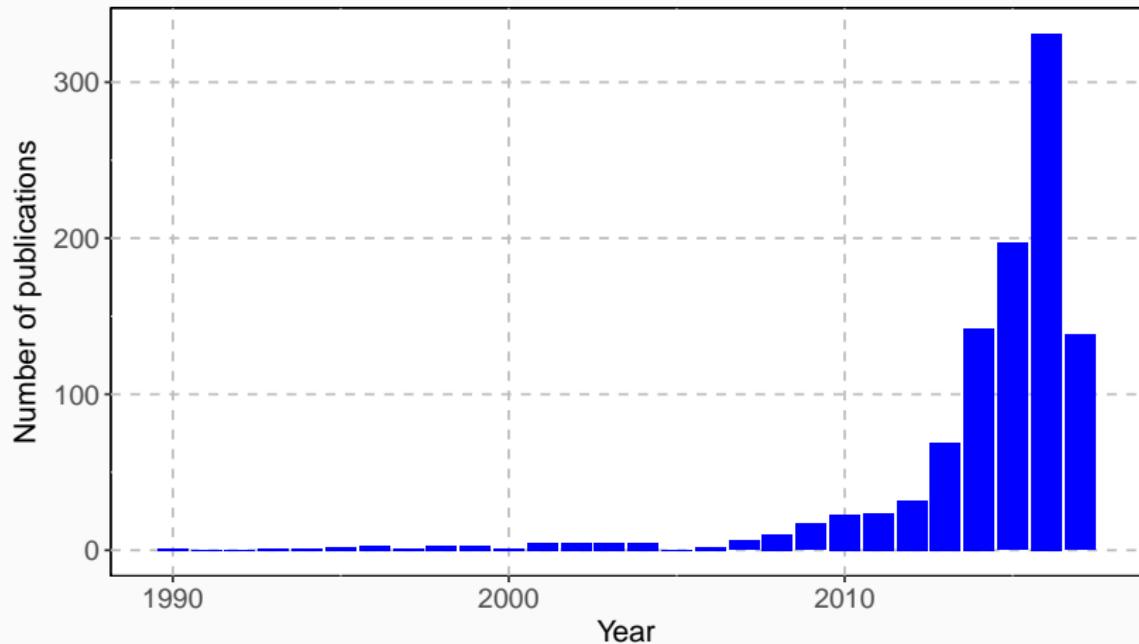
Aim

Methods

Reproducibility

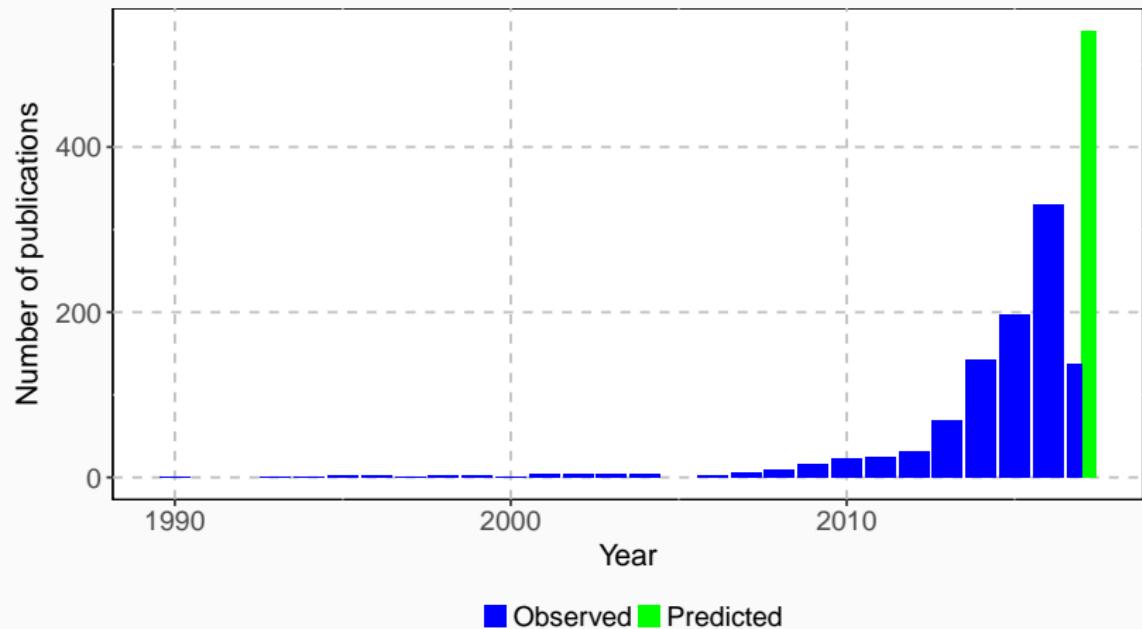
dPCR software

dPCR-related publications



Number of publications with words "digital PCR" or "dPCR" in the title/abstract.

dPCR-related publications



541 expected publications in 2017.

Vendor-provided dPCR software

Graphical user interfaces:

- QuantaSoft™ (Biorad),
- OpenArray® Digital PCR Software (Thermo Fisher),
- Digital PCR Analysis software (Fluidigm).

Vendor-provided dPCR software

Graphical user interfaces:

- QuantaSoft™ (Biorad),
- OpenArray® Digital PCR Software (Thermo Fisher),
- Digital PCR Analysis software (Fluidigm).

Closed-source software tied only to the vendor-specific data format.

Other dPCR software

Scripts and smaller tools tied to very specific task:

- **Mathematica** (Strain et al., 2013),
- **MS EXCEL** (Dobnik et al., 2015),
- **R** (Dreo et al., 2014; Trypsteen et al., 2015; Dorazio and Hunter, 2015; Vynck et al., 2016).

R: a software environment and a programming language (R Core Team, 2016), extensively used in bioinformatics and biostatistics.

Other dPCR software

Web servers (limited to published web servers):

- `definetherain` (Jones et al., 2014),
- `ddpcr` (Attali et al., 2016) (also the **R** package),

Aim

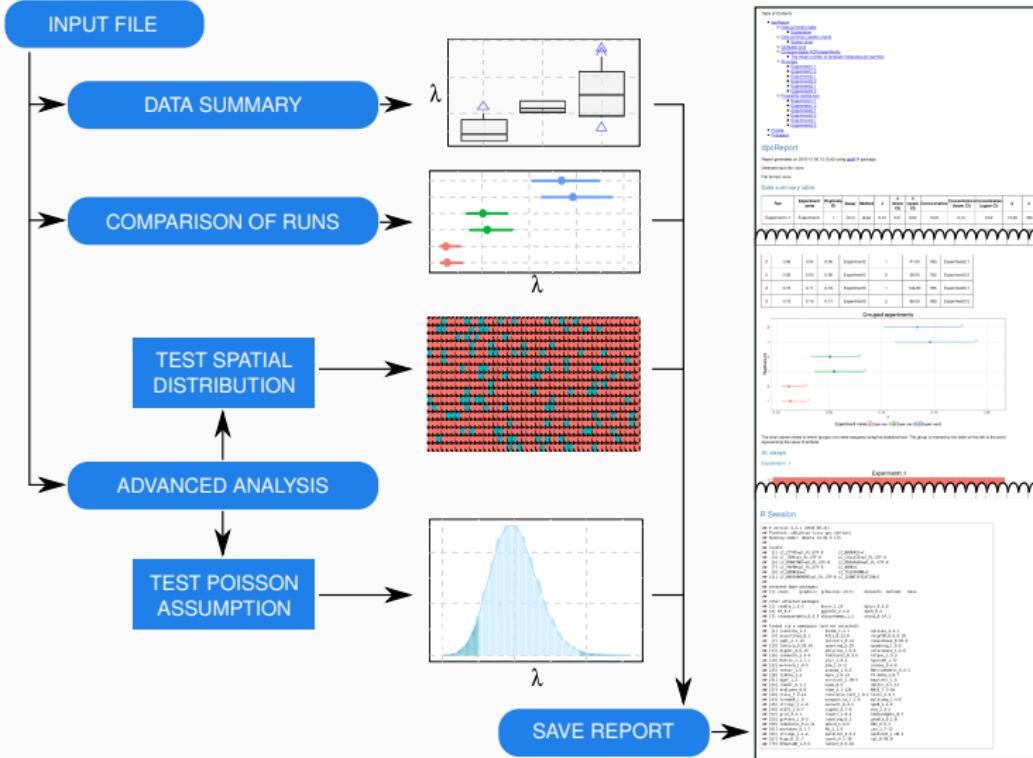
Aim

Existing software	Desired software
closed-source (vendors), open-source (scientific)	open-source
partially reproducible	fully reproducible
tied to a specific platform	multi-platform
solving a specific problem	focused on common tasks
scattered	integrated

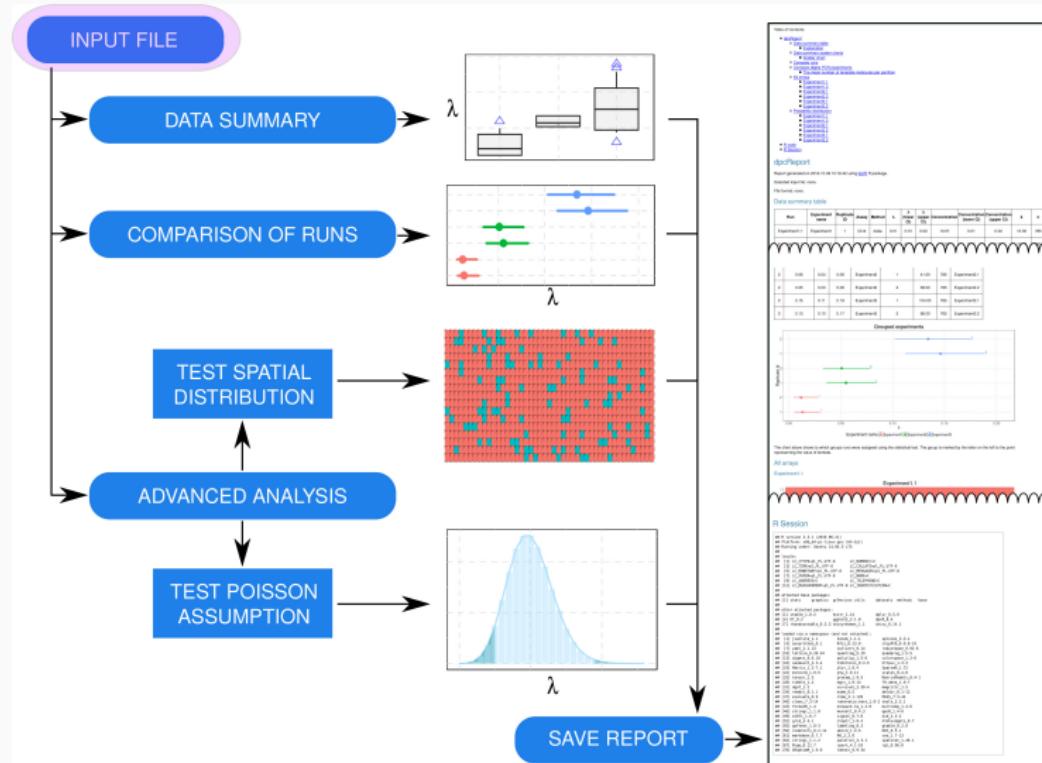
Aim

Existing software	dpcReport
closed-source (vendors), open-source (scientific)	open-source
partially reproducible	fully reproducible
tied to a specific platform	multi-platform
solving a specific problem	focused on common tasks
scattered	integrated

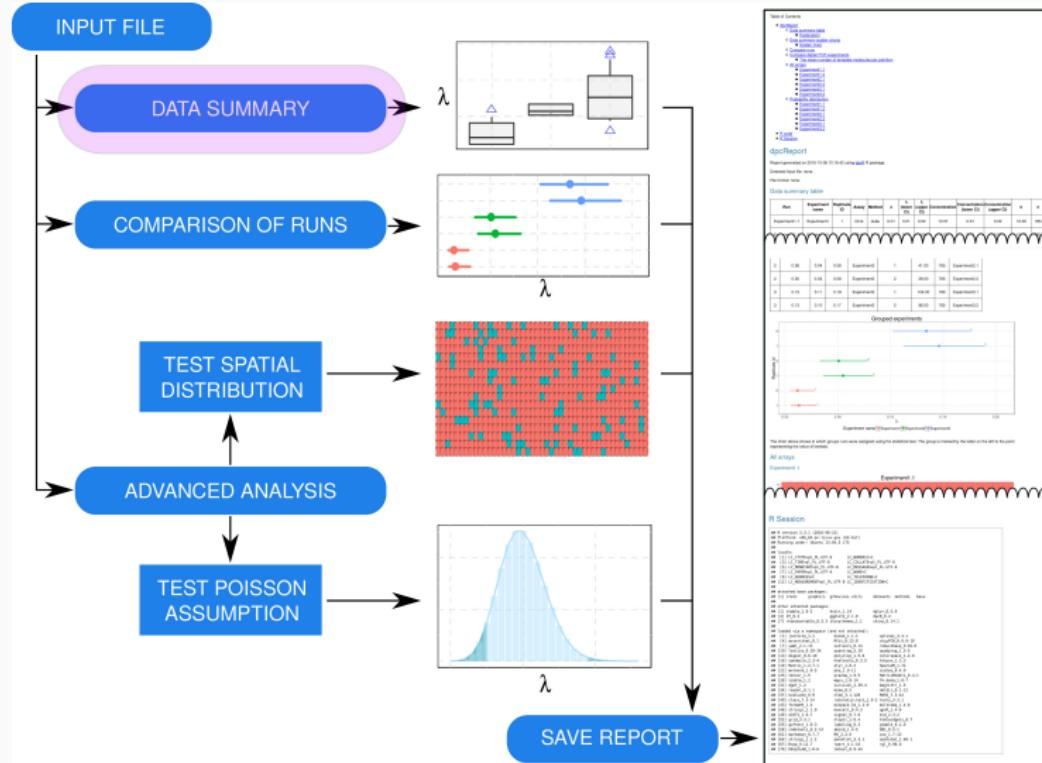
dpcReport framework



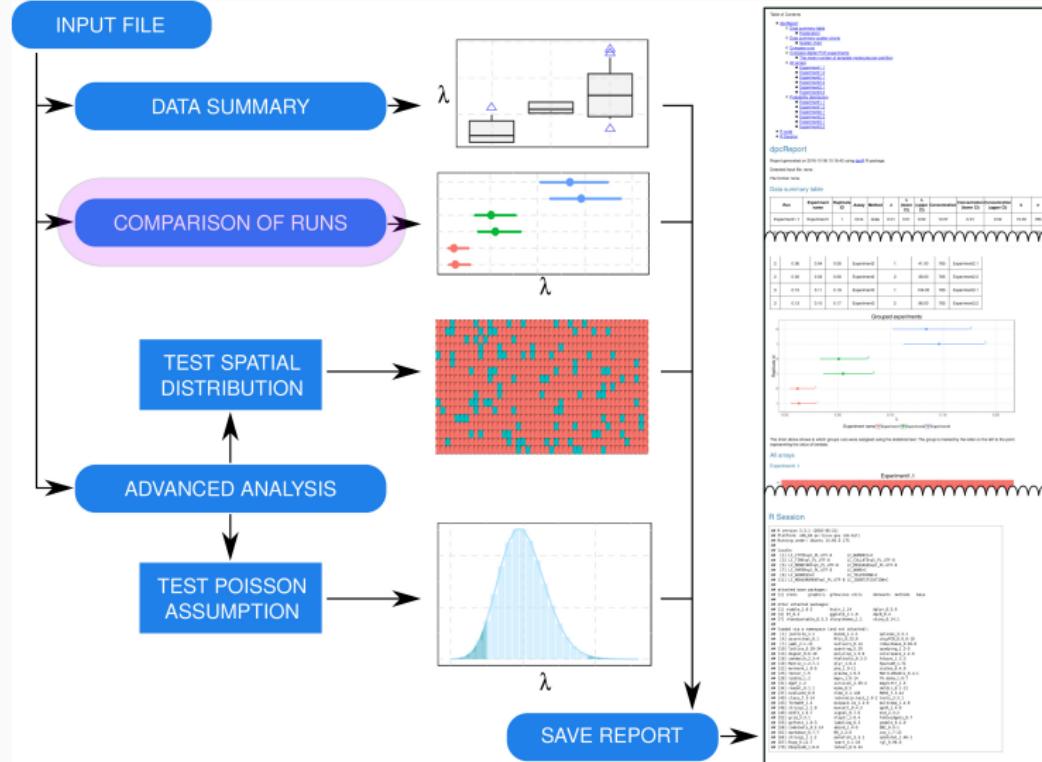
dpcReport framework



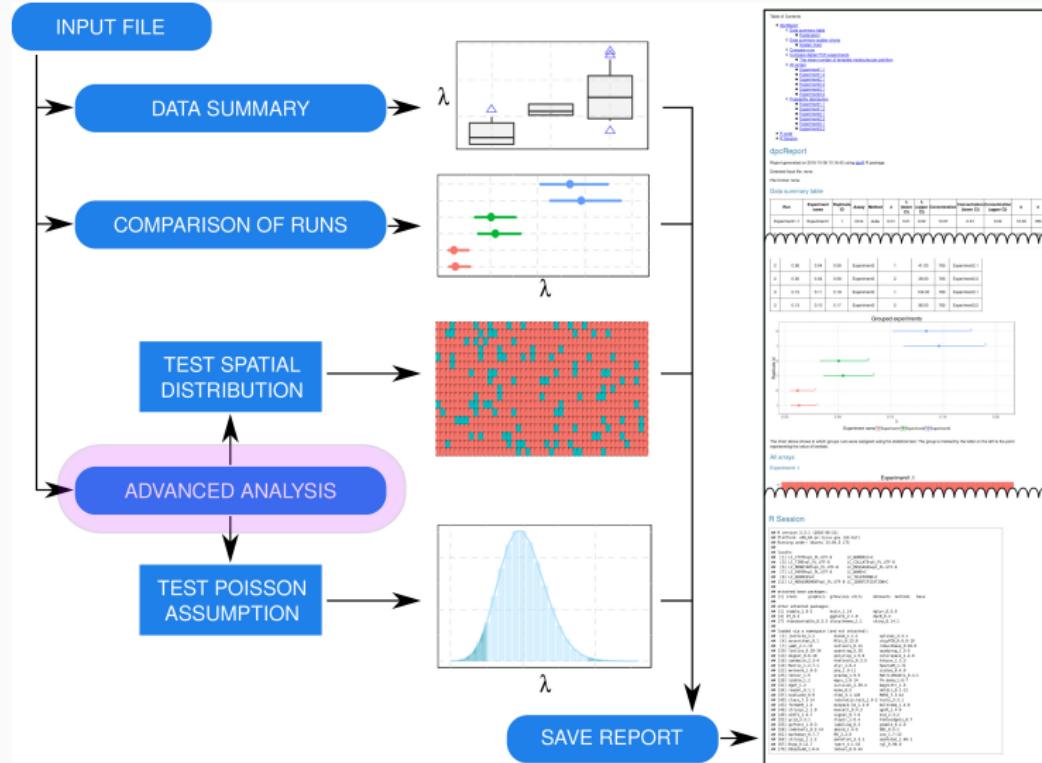
dpcReport framework



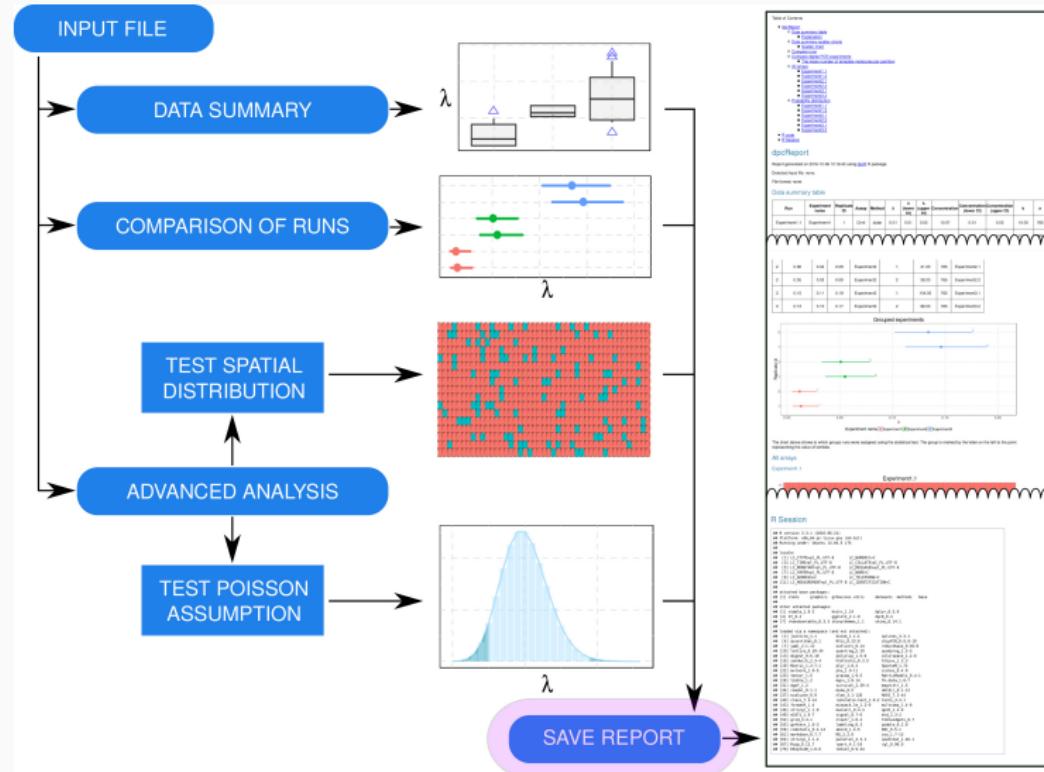
dpcReport framework



dpcReport framework



dpcReport framework



Methods

Input file

There are no universal format for dPCR data. Each system provides output in a different format. File formats differ between systems provided by the same vendor (QX100 vs QX200).

dpcReport supports analysis of data in following file formats:

- QX100 (Biorad),
- QX200 (Biorad),
- BioMark (Fluidigm).

REDF

REDF (Raw Exchange Digital PCR format): universal and minimal format for dPCR data analysis.

experiment	replicate	assay	k	n	v	uv	threshold	panel_id
Experiment1	1	Chr4	11	765	1	0	1	1
Experiment1	2	MYC	9	765	1	0	1	2
Experiment2	1	Chr4	39	765	1	0	1	3
Experiment2	2	MYC	42	765	1	0	1	4
Experiment3	1	Chr4	92	765	1	0	1	5
Experiment3	2	MYC	85	765	1	0	1	6

Multiple comparison

The most common task for dPCR is comparison of multiple reactions.

Multiple comparison

Uniformly most powerful test

Multiple comparison

Confidence intervals

Array quality control

Reproducibility

Reproducibility

Scientific software must support reproducibility, otherwise it is not scientific.

A report should contain enough information to allow the full reproduction of the conducted analysis.

Reports

Table of Contents

- [dpcReport](#)
 - [Data summary table](#)
 - [Explanation](#)
 - [Data summary scatter charts](#)
 - [Scatter chart](#)
 - [Compare runs](#)
 - [Compare digital PCR experiments](#)
 - [The mean number of template molecules per partition](#)
- [R code](#)
- [R Session](#)

dpcReport

Report generated on 2017-03-27 07:12:26 using [dpcR](#) R package.

Detected input file: 20130918_Dilution_log10.csv.

md5 checksum of the input file: 6bf3306199dd7af439d1c8acd08e23c1

The input was modified manually in dpcReport application.

File format: QX100.

Data summary table

Run	Experiment name	Replicate ID	Assay	Method	λ	λ (lower CI)	λ (upper CI)	Concentration	Concentration (lower CI)	Concentration (upper CI)	k	n
ABS1.A09.gDNA + P 10^4	ABS1	A09.gDNA + P 10^4	ileS	dube	0.07	0.07	0.08	970.44	0.08	0.09	936	13346
ABS1.A09.gDNA + P 10^4	ABS1	A09.gDNA + P 10^4	ileS	bhat	0.07	0.07	0.08	970.44	0.08	0.09	936	13346
ABS1.B09.gDNA		B09.gDNA										

Date and time

Table of Contents

- [dpcReport](#)
 - [Data summary table](#)
 - [Explanation](#)
 - [Data summary scatter charts](#)
 - [Scatter chart](#)
 - [Compare runs](#)
 - [Compare digital PCR experiments](#)
 - [The mean number of template molecules per partition](#)
- [R code](#)
- [R Session](#)

dpcReport

Report generated on 2017-03-27 07:12:26 using [dpcR](#) R package.

Detected input file: 20130918_Dilution_log10.csv.

md5 checksum of the input file: 6bf3306199dd7af439d1c8acd08e23c1

The input was modified manually in dpcReport application.

File format: QX100.

Data summary table

Run	Experiment name	Replicate ID	Assay	Method	λ	λ (lower CI)	λ (upper CI)	Concentration	Concentration (lower CI)	Concentration (upper CI)	k	n
ABS1.A09.gDNA + P 10 ⁴	ABS1	A09.gDNA + P 10 ⁴	ileS	dube	0.07	0.07	0.08	970.44	0.08	0.09	936	13346
ABS1.A09.gDNA + P 10 ⁴	ABS1	A09.gDNA + P 10 ⁴	ileS	bhat	0.07	0.07	0.08	970.44	0.08	0.09	936	13346
ABS1.B09.gDNA		B09.gDNA										

Input file name

Table of Contents

- [dpcReport](#)
 - [Data summary table](#)
 - [Explanation](#)
 - [Data summary scatter charts](#)
 - [Scatter chart](#)
 - [Compare runs](#)
 - [Compare digital PCR experiments](#)
 - [The mean number of template molecules per partition](#)
- [R code](#)
- [R Session](#)

dpcReport

Report generated on 2017-03-27 07:12:26 using [dpcR](#) R package.

Detected input file: 20130918_Dilution_log10.csv.

md5 checksum of the input file: 6bf3306199dd7af439d1c8acd08e23c1

The input was modified manually in dpcReport application.

File format: QX100.

Data summary table

Run	Experiment name	Replicate ID	Assay	Method	λ	λ (lower CI)	λ (upper CI)	Concentration	Concentration (lower CI)	Concentration (upper CI)	k	n
ABS1.A09.gDNA + P 10 ⁴	ABS1	A09.gDNA + P 10 ⁴	ileS	dube	0.07	0.07	0.08	970.44	0.08	0.09	936	13346
ABS1.A09.gDNA + P 10 ⁴	ABS1	A09.gDNA + P 10 ⁴	ileS	bhat	0.07	0.07	0.08	970.44	0.08	0.09	936	13346
ABS1.B09.gDNA		B09.gDNA										

Input file checksum

Table of Contents

- [dpcReport](#)
 - [Data summary table](#)
 - [Explanation](#)
 - [Data summary scatter charts](#)
 - [Scatter chart](#)
 - [Compare runs](#)
 - [Compare digital PCR experiments](#)
 - [The mean number of template molecules per partition](#)
- [R code](#)
- [R Session](#)

dpcReport

Report generated on 2017-03-27 07:12:26 using [dpcR](#) R package.

Detected input file: 20130918_Dilution_log10.csv.

md5 checksum of the input file: 6bf3306199dd7af439d1c8acd08e23c1

The input was modified manually in dpcReport application.

File format: QX100.

Data summary table

Run	Experiment name	Replicate ID	Assay	Method	λ	λ (lower CI)	λ (upper CI)	Concentration	Concentration (lower CI)	Concentration (upper CI)	k	n
ABS1.A09.gDNA + P 10^4	ABS1	A09.gDNA + P 10^4	ileS	dube	0.07	0.07	0.08	970.44	0.08	0.09	936	13346
ABS1.A09.gDNA + P 10^4	ABS1	A09.gDNA + P 10^4	ileS	bhat	0.07	0.07	0.08	970.44	0.08	0.09	936	13346
ABS1.B09.gDNA		B09.gDNA										

Changes in case of the manual alteration of the input file.

Manual alterations inside dpcReport

Table of Contents

- [dpcReport](#)
 - [Data summary table](#)
 - [Explanation](#)
 - [Data summary scatter charts](#)
 - [Scatter chart](#)
 - [Compare runs](#)
 - [Compare digital PCR experiments](#)
 - [The mean number of template molecules per partition](#)
- [R code](#)
- [R Session](#)

dpcReport

Report generated on 2017-03-27 07:12:26 using [dpcR](#) R package.

Detected input file: 20130918_Dilution_log10.csv.

md5 checksum of the input file: 6bf3306199dd7af439d1c8acd08e23c1

The input was modified manually in dpcReport application.

File format: QX100.

Data summary table

Run	Experiment name	Replicate ID	Assay	Method	λ	λ (lower CI)	λ (upper CI)	Concentration	Concentration (lower CI)	Concentration (upper CI)	k	n
ABS1.A09.gDNA + P 10^4	ABS1	A09.gDNA + P 10^4	ileS	dube	0.07	0.07	0.08	970.44	0.08	0.09	936	13346
ABS1.A09.gDNA + P 10^4	ABS1	A09.gDNA + P 10^4	ileS	bhat	0.07	0.07	0.08	970.44	0.08	0.09	936	13346
ABS1.B09.gDNA	...	B09.gDNA

R Session

dpcReport is based on the **R** package *dpcR*. All functionalities of dpcReport, including table and figure generation, are affected by changes in **R** and *dpcR*.

R Session

```
## R version 3.3.3 (2017-03-06)
## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: Ubuntu 14.04.5 LTS
##
## locale:
## [1] LC_CTYPE=pl_PL.UTF-8      LC_NUMERIC=C
## [3] LC_TIME=pl_PL.UTF-8      LC_COLLATE=pl_PL.UTF-8
## [5] LC_MONETARY=pl_PL.UTF-8   LC_MESSAGES=pl_PL.UTF-8
## [7] LC_PAPER=pl_PL.UTF-8     LC_NAME=C
## [9] LC_ADDRESS=C              LC_TELEPHONE=C
## [11] LC_MEASUREMENT=pl_PL.UTF-8 LC_IDENTIFICATION=C
##
## attached base packages:
## [1] stats      graphics    grDevices   utils       datasets    methods    base
##
## other attached packages:
## [1] xtable_1.8-2      knitr_1.15.1    digest_0.6.12
## [4] dplyr_0.5.0       DT_0.2        ggplot2_2.2.1
## [7] rhandsontable_0.3.4 shinythemes_1.1.1 shiny_1.0.0
## [10] dpcR_0.4
##
## loaded via a namespace (and not attached):
## [1] jsonlite_1.2        binom_1.1-1      splines_3.3.3
## [4] assertthat_0.1       highr_0.6        Rfit_0.23.0
## [7] chipPCR_0.0.8-11    yaml_2.1.14     outliers_0.14
## [10] robustbase_0.92-7   lattice_0.20-34  quantreg_5.29
## [13] quadprog_1.5-5      polyclip_1.5-6   colorspace_1.3-1
## [16] sandwich_2.3-4      htmltools_0.3.5  httpuv_1.3.3
## [19] Matrix_1.2-8        plyr_1.8.4      SparseM_1.74
## [22] mvtnorm_1.0-5       ptw_1.9-11     scales_0.4.1
## [25] tensor_1.5          pracma_1.9.5    MatrixModels_0.4-1
## [28] tibble_1.2           mgcv_1.8-16    TH.data_1.0-7
## [31] dgof_1.2            lazyeval_0.2.0   survival_2.40-1
## [34] magrittr_1.5          readxl_0.1.1    mime_0.5
## [37] deldir_0.1-12       evaluate_0.10   nlme_3.1-131
```

Reproducibility of the workflow

An analysis conducted in a GUI-based software, as *dpcReport*, is challenging to reproduce.

Reproducibility of the workflow

Choose elements of the report

- Data summary table
- Data summary scatter charts
- Compare runs
- Visualise and analyze individually each array
- Visualise and analyze probability distribution of each run
- R code used in the report generation

Be patient. The generation of the report may take few minutes.

 Save report

 Save input data (.csv)

dpcReport exports all steps of the analysis, including parameters adjusted manually by the user, in form of the **R** code that recreates the whole workflow.

Reproducibility of the workflow

R code

The R code below may be used to recreate reported results.

```
# Load packages
library(dpcR)
# if you do not have dpcR package, install it from GitHub:
# devtools::install_github("michbur/dpcR")
library(ggplot2) # ggplot2 library for nice plots
# Define theme for plots
cool_theme <- theme(plot.background=element_rect(fill = "transparent", colour = "transparent"),
panel.grid.major = element_line(colour="lightgrey", linetype = "dashed"), panel.background =
element_rect(fill = "white", colour = "black"), legend.background = element_rect(fill="NA"),
legend.position = "bottom", axis.text = element_text(size = 14), axis.title.x = element_text(size=17,
vjust = -0.1), axis.title.y = element_text(size = 17, vjust = 1), strip.text = element_text(size = 17,
face = "bold"), strip.background = element_rect(fill = "#9ecae1", colour = "black"), legend.text =
element_text(size=14), legend.title = element_text(size = 17), plot.title = element_text(size = 22),
legend.key = element_rect(fill = "white", colour = "black", linetype = "dashed", size = 0.5))
# Read and adjust data
# The input file is assumed to be in the current R working directory
input_data <- read_dpcr("20130918 Dilution_log10.csv", format = "QX100")

#####
# Print only table from summary.dpcr function
summary(input_data, print = FALSE)[["summary"]]

#####
# Prepare data for plots
plot_data <- summary(input_data, print = FALSE)[["summary"]]
plot_data <- plot_data[plot_data[["method"]] == "dube", ]
ggplot(plot_data, aes(x = experiment, y = lambda, ymax = lambda.up, ymin = lambda.low)) + geom_point(size =
4, alpha = 0.6, shape = 2, colour = "blue") + cool_theme + geom_boxplot(outlier.colour = NA, fill =
adjustcolor("lightgrey", alpha.f = 0.25), shape = 15) + ggtile(paste0("Experiment boxplot\nCI method: ",
```

Summary

dpcReport is an open-source, integrated software for the reproducible analysis of dPCR data.

Getting started

Web server: <http://tinyurl.com/dpcReport2>.

The screenshot shows the dpcReport web application. At the top, there is a blue header bar with the following navigation items: 'dpcReport' (highlighted in white), 'Input file', 'Data summary ▾', 'Comparison of runs', 'Advanced analysis ▾', 'Save report', and 'About'. Below the header, the main content area has a light gray background. It features a large blue header 'Welcome to dpcReport'. Underneath, a text block says 'For in-depth description of the dpcReport, please refer to the **About** panel.' A section titled 'Upload data' is present, with the sub-instruction 'Upload your data using the button below or analyze the preloaded data.' Below this, 'Accepted data formats:' is listed with a bulleted list: raw data - comma-separated .csv file (for array digital PCR and droplet digital PCR), QX100 - data from QX100 Droplet Digital PCR System (Life Technologies), BioMark (Detailed Table Results) - BioMark (Fluidigm), BioMark (Summary Table Results) - BioMark (Fluidigm), and amplification data: compressed (.zip) directory with amplification data from QX series. At the bottom of this section, it says 'See also [exemplary data files](#)'. Another 'Accepted file formats:' section follows, listing .CSV, .XLSX, and .XLS.

Getting started

Local instance: <https://github.com/michbur/dpcR>.

```
install.packages("dpcR")
library(dpcR)
dpcReport()
```

Acknowledgements and funding

Collaborators:

- Jim Hugget and Alexandra Whale (LGC).
- Boris Fehse (University of Hamburg).
- Mario Menschikowski (Technical University of Dresden).
- Stefan Rödiger (Brandenburg Technical University).

Funders:

- KNOW Consortium Wrocław Center for Biotechnology,
- National Science Center (2015/17/N/NZ2/01845),
- COST action "Harmonising standardisation strategies to increase efficiency and competitiveness of European life-science research".

Local instance: <https://github.com/michbur/dpcR>.

Web server: <http://tinyurl.com/dpcReport2>.

Slides: <http://tinyurl.com/dpcReport-Freising>.

dpcReport and dpcR are part of pcRuniveRsum:

<http://michbur.github.io/pcRuniveRsum/> - everything for PCR in **R**.

References

- Attali, D., Bidshahri, R., Haynes, C., and Bryan, J. (2016). Ddpcr: An R package and web application for analysis of droplet digital PCR data. *F1000Research*, 5:1411.
- Dobnik, D., Spilsberg, B., Bogožalec Košir, A., Holst-Jensen, A., and Žel, J. (2015). Multiplex Quantification of 12 European Union Authorized Genetically Modified Maize Lines with Droplet Digital Polymerase Chain Reaction. *Analytical Chemistry*, 87(16):8218–8226.

References II

- Dorazio, R. M. and Hunter, M. E. (2015). Statistical Models for the Analysis and Design of Digital Polymerase Chain Reaction (dPCR) Experiments. *Analytical Chemistry*, 87(21):10886–10893.
- Dreo, T., Pirc, M., Ramšak, Ž., Pavšič, J., Milavec, M., Žel, J., and 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.

References III

- Jones, M., Williams, J., Gärtner, K., Phillips, R., Hurst, J., and 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(100):46–53.
- R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Strain, M. C., Lada, S. M., Luong, T., Rought, S. E., Gianella, S., Terry, V. H., Spina, C. A., Woelk, C. H., and Richman, D. D. (2013). Highly precise measurement of HIV DNA by droplet digital PCR. *PloS one*, 8(4):e55943.

References IV

- Trypsteen, W., Vynck, M., De Neve, J., Bonczkowski, P., Kiselinova, M., Malatinkova, E., Vervisch, K., Thas, O., Vandekerckhove, L., and De Spiegelaere, W. (2015). ddpcRquant: Threshold determination for single channel droplet digital PCR experiments. *Analytical and Bioanalytical Chemistry*, 407(19):5827–5834.
- Vynck, M., Vandesompele, J., Nijs, N., Menten, B., De Ganck, A., and Thas, O. (2016). Flexible analysis of digital PCR experiments using generalized linear mixed models. *Biomolecular Detection and Quantification*, 9:1–13.