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

Michał Burdukiewicz[1], Jim Huggett[2], Alexandra Whale[2], Piotr Sobczyk[3], Paweł Mackiewicz[1], Andrej-Nikolai Spiess[3], Peter Schierack[5], and Stefan Rödiger[5]

[1]University of Wrocław, Department of Genomics,
[2]Molecular and Cell Biology Team, LGC, Teddington, United Kingdom,
[3]Wrocław University of Science and Technology, Faculty of Pure and Applied Mathematics,
[4]University Medical Center Hamburg-Eppendorf, Hamburg, Germany,
[5]Brandenburg University of Technology Cottbus-Senftenberg, Institute of Biotechnology

Outline

dPCR software

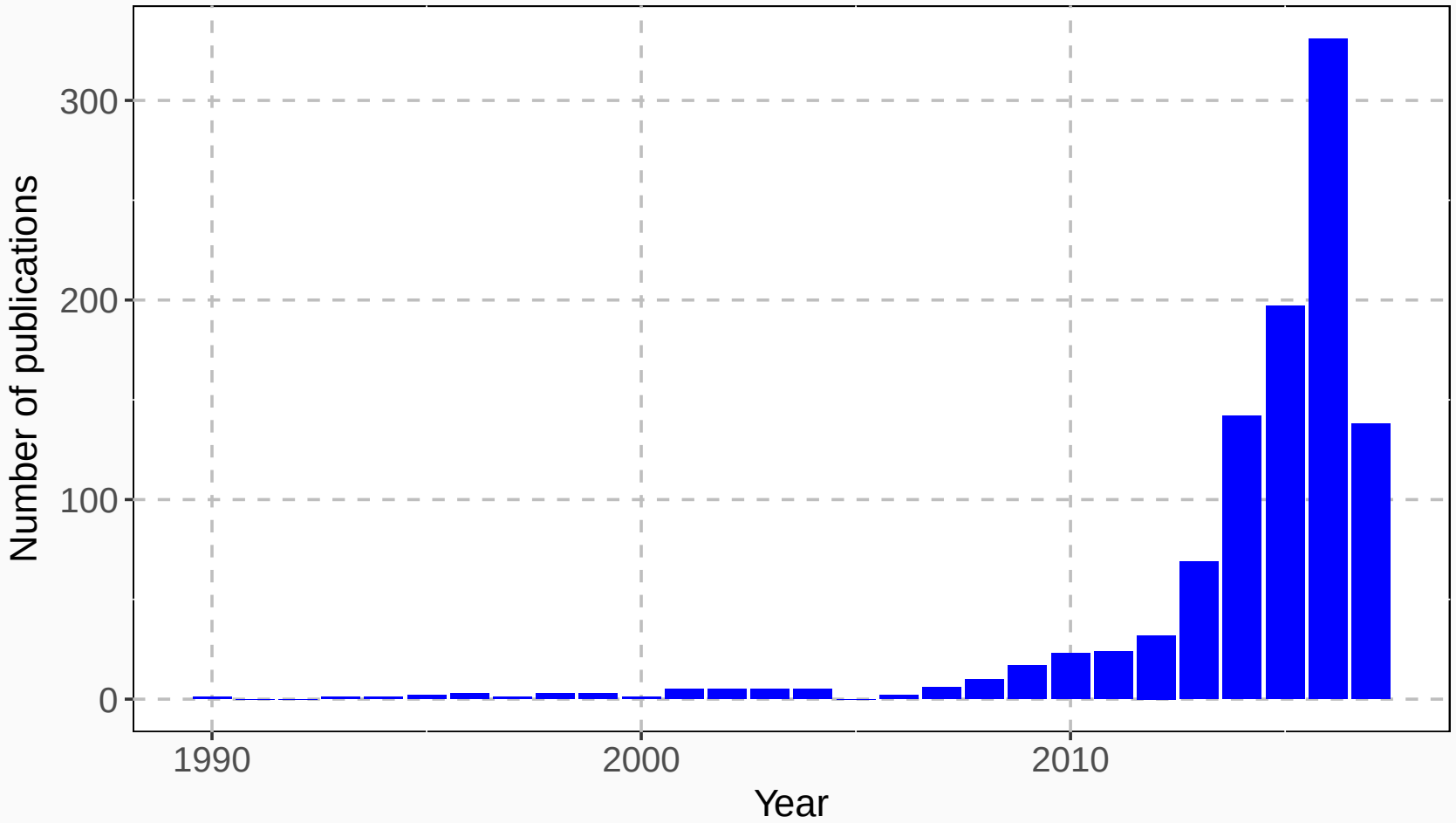
Aim

Methods

Reproducibility

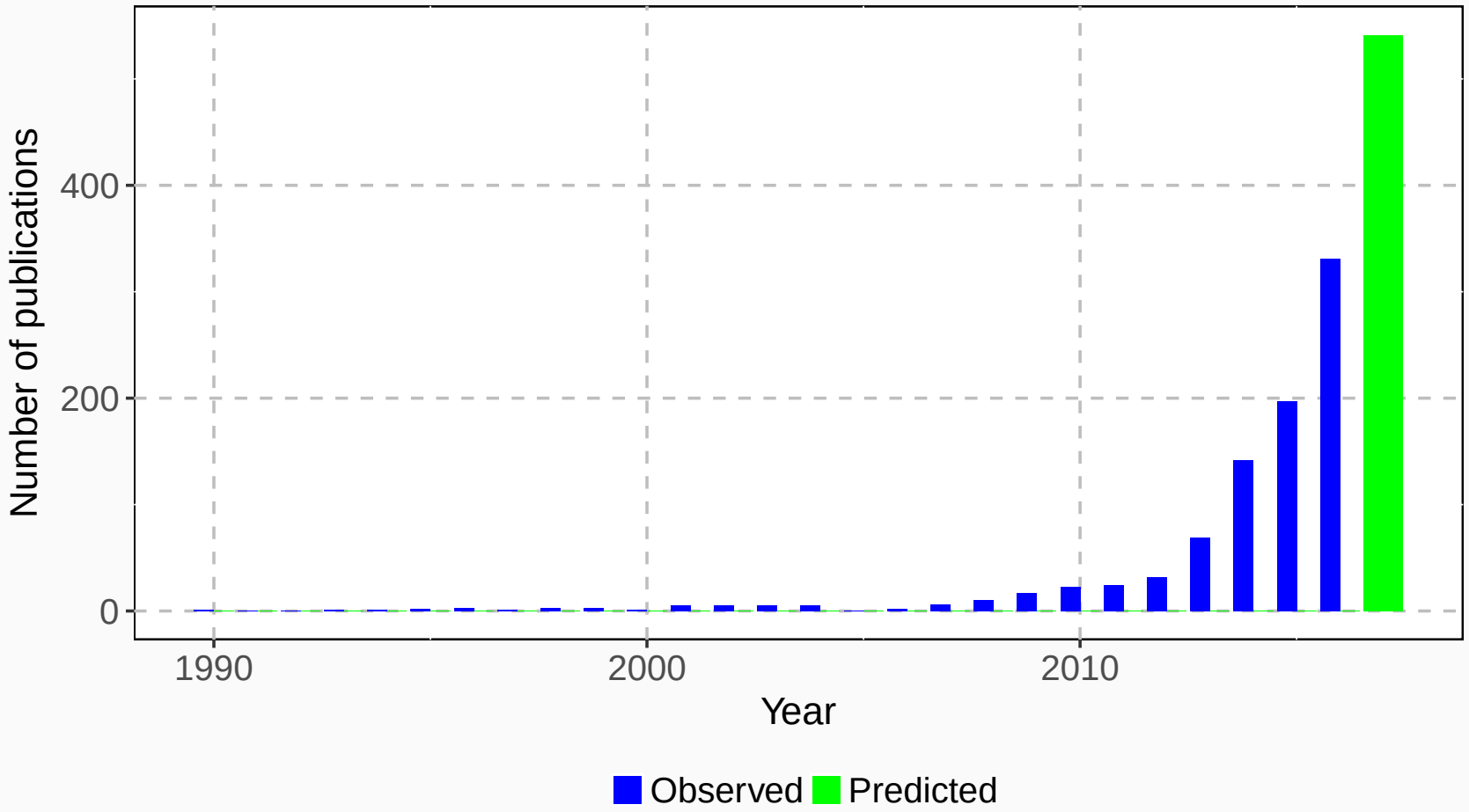
dPCR software

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

Other dPCR software

- definetherain
- dpcr package,
- twodpcr package

Aim

Aim

Create framework for dPCR data analysis:

Aim

Create framework for dPCR data analysis:

- tailored for the most common tasks,

Aim

Create framework for dPCR data analysis:

- tailored for the most common tasks,
- unified,

Aim

Create framework for dPCR data analysis:

- tailored for the most common tasks,
- unified,
- reproducible.

Methods

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

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

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

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

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

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

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 more challenging to reproduce.

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) + ggtitle(paste0("Experiment boxplot\nCI method: ",
```

Summary

dpcReport is the integrted environment for the analysis of dPCR data.

Availability

Web server: `www.smorfland.uni.wroc.pl/shiny/dpcReport`

R package (including your local instance of dpcReport):
`www.github.com/michbur/dpcR`

Acknowledgements and funding

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

- Stefan Rödiger,
- Boris Fehse.

References

Família, C., Dennison, S. R., Quintas, A., and Phoenix, D. A. (2015). Prediction of Peptide and Protein Propensity for Amyloid Formation. *PLOS ONE*, 10(8):e0134679.

Garbuzynskiy, S. O., Lobanov, M. Y., and Galzitskaya, O. V. (2010). FoldAmyloid: a method of prediction of amyloidogenic regions from protein sequence. *Bioinformatics (Oxford, England)*, 26(3):326–332.

References II

Kosiol, C., Goldman, N., and Buttimore, N. H. (2004). A new criterion and method for amino acid classification. *Journal of Theoretical Biology*, 228(1):97–106.

Melo, F. and Marti-Renom, M. A. (2006). Accuracy of sequence alignment and fold assessment using reduced amino acid alphabets. *Proteins*, 63(4):986–995.

Paz, M. L. d. l. and Serrano, L. (2004). Sequence determinants of amyloid fibril formation. *Proceedings of the National Academy of Sciences*, 101(1):87–92.

References III

Sawaya, M. R., Sambashivan, S., Nelson, R., Ivanova, M. I., Sievers, S. A., Apostol, M. I., Thompson, M. J., Balbirnie, M., Wiltzius, J. J. W., McFarlane, H. T., Madsen, A. , Riekel, C., and Eisenberg, D. (2007). Atomic structures of amyloid cross-spines reveal varied steric zippers. *Nature*, 447(7143):453–457.

Stephenson, J. D. and Freeland, S. J. (2013). Unearthing the root of amino acid similarity. *Journal of Molecular Evolution*, 77(4):159–169.

Walsh, I., Seno, F., Tosatto, S. C. E., and Trovato, A. (2014). PASTA 2.0: an improved server for protein aggregation prediction. *Nucleic Acids Research*, page gku399.