BIOINFORMATICS

Vol. 00 no. 00 2005
Pages 1–3

dpcrR: anR package for the analysis of digital PCR

Michał Burdukiewicz [1], Jim Huggett [2], Bart K.M. Jacobs [3], Lieven Clement [3], Piotr Sobczyk [1], Andrej-Nikolai Spiess [4], Peter Schierack [5], and Stefan Rödiger [5]*

[1]Department of Genomics, Faculty of Biotechnology, University of Wrocław, Wrocław, Poland
[2]Molecular and Cell Biology Team, LGC, Teddington, United Kingdom
[3]Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Belgium
[4]University Medical Center Hamburg-Eppendorf, Hamburg, Germany
[5]Faculty of Natural Sciences, Brandenburg University of Technology Cottbus–Senftenberg, Germany

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Associate Editor: XXXXXXX

ABSTRACT

Motivation: The digital Polymerase Chain reaction is a state of the technology which is emerging in various research areas including life-sciences and diagnostics. dPCR is likely to have the same impact in quantification of nucleic acids as quantitative real-time PCR. Advantages over conventional qPCR include the possibility of absolute quantification and the drastically reduced sensitivity to inhibitors. There are different technical approaches to dPCR based on droplets/microfluidics or nano-structured chambers and different statistical analysis methods. However, a unified open software for which fits the needs for (I) data analysis and presentation in research, (II) as software frame-work for novel technical developments, (III) as platform for teaching this new technology and (IV) serves as reference for statistical methods to dPCR is lacking. Therefore we aimed to develop an**R** package which serves as Swiss-army knife in dPCR.

Results: To cover all methods of dPCR we implemented all accessible peer-review methods and plots into the *dpcR* **R** package [3] with a plug-in like architecture. This versatile package provides functions to process data degenerated by droplets and chamber based technologies. Functions included may be used to simulate dPCRs, perform statistical data analysis, plotting of the results and simple report generation. We implemented many functions with binding to the*shiny* **R** package [4] to provide means to run it as interactive web application. Features such as functions to estimate the underlying Poisson process, methods from peer-reviewed literature for calculating confidence intervals based on single samples as well as on replicates, a novel Generalized Linear Model-based procedure to compare digital PCR experiments and a spatial randomness test for assessing plate effects have been integrated. Thus, the *dpcR* package can be used by**R** novices in a graphical user interface or on expert level in R. The *dpcR* package is an open environment, which can be adopted to the growing knowledge in dPCR. The *dpcR* package can be used to build a custom-made analyzer according to the wishes of the user. The source code is open source (GPL-3 or later) and freely available from CRAN.

Availability: http://cran.r-project.org/web/packages/dpcR newline Source code: https://github.com/michbur/dpcR

Contact: stefan.roediger@b-tu.de

1 INTRODUCTION

There are three principal approaches to quantify nucleic acids. The fist is by referencing the material to an external calibrator (qPCR), The standard approach to quantify nucleic acids has been the quantitative real-time PCR (qPCR) so far [10]. It is a well established and robust technology, which allows precise quantification of DNA material in high throughput fashion at a reasonable price. However, the quantification by qPCR is challenging at very low and very high concentrations. At low concentration Monte Carlo effect play a major role and at high concentration inhibition process start to dominate the qPCR. Thus, the qPCR is only usable in the working range of the calibrator. In addition, pre-processing and data analysis is a affected by numerous adverse effects [16]. The second approach is to count the number of molecules (e.g., superSAGE or NanoStrings) (Matsumura-2006-Nature-Methods, Waggott-2012-Bioinformatics). The third is to analyze the number of positive reactions in relation to the number of total reaction (dPCR). Since approximately ten year the digital PCR (dPCR) is gaining entrance in the mainstream user-base. There is currently an intensive research on qPCR platforms with the overall aim to make to technology broadly usable, cheap, robust and to enable high sample throughput. The chemical basis of the dPCR is identical to the qPCR, which includes master-mix preparation and thermal cycling of the sample. In contrast to qPCR the amplification ration does not take place in a single reaction chamber but is rather a process of clonal amplification in small separate “compartments” (e.g., nl volume droplets of water oil emulsions, chambers on micro structured chips). The quantification of the amplification is not done by determining a Cq-value derived from an amplification curve but applying a Poisson distribution based determination of the concentration of the starting material. Therefore, the dPCR does not require an external calibration [15, 13].

*to whom correspondence should be addressed

A first proposal for digital PCR like approach and the use of the Poisson distribution to quantify the number of molecules on a "sample" was shown by Ruano et al. 1990 (PNAS) with the single molecule dilution (SMD) PCR. In 1999 Vogelstein et al. (PNAS) described the first true digital PCR [8]. Application of the dPCR cover all applications of conventional qPCR, including investigation of alleles, gene expression analysis and absolute quantification of PCR products. For absolute quantification the qPCR relied on an external calibrator (calibration curve) which was derived serial decadic dilution (e.g., 1:10 → 1:100 → 1:1000) of a known target input quantity. The real-time monitoring of the PCR product formation enabled to determine quantification points (Cq). The Cq are strictly related to the input quantity. A simple arithmetic operation (after logarithmic transformation of the concentration) is sufficient to determine any nucleic acid quantity [6].

qPCR dPCR Number of copies/DNA per volume (e.g., ng/l, copies/l) total number of compartments * ln (...)

The dPCR has some principle assumptions and fundamental properties. First of all the chemical reaction should be not affected by inhibitors. The distribution of the single molecule target regions follows a Poission distribution. The Poisson distribution appears like a normal distribution but without negative values and being zero the lowest. First a large number (n) of amplifications reactions as required to have a high statistical power. Therefore in practical terms a massive number of PCR reactions is needed. For Poission distributions an n of XY (get reference from table/text book form statistics/biostatistics?) is considered large. Second that the molecules required for the amplification amplifications reactions are randomly distributed in the compartments. Visual analysis, Ripley's K functions or ??? can be used to test for randomness of the reaction and thus to exclude the clustering of of positive reactions. A clustering of positive wells might be due to sample loading or analysis process (systematical error). The outcome of an amplification can be no amplification at all (less than 1 copy per volume), an unsaturated reaction with a binary/"multinary" amplification (usable to calculate the "concentration") or a saturated reaction where virtually all compartments are positive.

Calculation of the "Concentration" Reference to "Supplement"

Calculation of the uncertainty To determine the uncertainty of the calculations two approach have been proposed in the peer-review literature (Dube 2008, PLoS One, Bath). The uncertainty is dependent on the number of PCR reactions (reference to *dpcR* functions). Reference to "Supplement" and *dpcR* functions.

We developed the *dpcR* package which is software suite for analysis of dPCR based on the open source statistical software R. The *dpcR* includes to invitation to the scientific community to join and support the development of *dpcR* (github?). The aim of the software is to provide the scientific community a tool for teaching purposes, data analysis, theoretical research (simulation) and to accelerate the development of new approaches to dPCR. We implemented all existing statistical methods for dPCR and suggest the introduction of a standardized nomenclature for qPCR. The package enables the simulations and predictions of Poisson distribution for dPCR scenarios, the analysis of previously run dPCRs.

Interactive use and graphical representation with *shiny* [4].

Import and export of results figures and data.

There are currently two technical approaches to dPCR. dPCRs may use (microfluidic)chambers or emulsion based chambers (QX200 $^{\text{TM}}$(Bio-Rad), RainDrop $^{\text{TM}}$System (RainDance)). Chamber based dPCR systems have fixed geometries, including the volume of the reaction chambers. Despite the fact that dPCRs is an endpoint analysis the chamber based technologies allow generally the real-time monitoring of the amplification reaction and subsequent confirmation of the amplification reaction be melting curve analysis. Thus, such technologies enable easier trouble shooting and quality management of the data. However, the downside of these technologies is the fixed limited number of compartments and the price. The emulsion based dPCRs are easier to perform since the compartments are generated by microfluidic technologies and have practically no limitation regarding the number of compartments. This results in a higher statistical power to quantify small differences in sample quantities. The emulsion chambers are made of water-in-oil emulsions with similar sizes.

Two-sided exact tests and matching confidence intervals for discrete data [5]

Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing [1]

Interval Estimation for a Binomial Proportion [2]

We have chosen **R** because it is the *lingua franca* in biostatistics and broadly used in other disciplines [13]. The are many packages in existence which enable the fast development of new methods and plotting facilities. As most **R** packages depend on one or more other packages [9] depends *dpcR* on other packages, resulting in a complex network of recursive dependencies. Core packages *qpcR* [11], *shiny* [4], *MBmca* [12], *chipPCR* [14] and further packages as shown in the dependency graph (Supplement XYZ).

The GUI employs advanced plots based on *ggplot2* [7].

R has a rich set of tool to arrange data (reshape?) in order to prepare them for the analysis. This is important when it comes to the question how experiments should be treated. It is possible to analyze the PCR reaction the panels independently (effect on CI and uncertainty) or to pool/aggregate all reactions (effect on CI and uncertainty) to achieve higher sensitivity/certainty.

2 APPROACH

One basic design decision was to structure specific properties of digital PCR systems (dropet vs. chamber) in auxiliary functions and to perform central calculation specific to Poisson statistics in independent main functions. Chamber digital PCR systems fundamentally rely on the proper preprocessing of qPCR data. We have chosen to implement the core functionality by a dependency to the *qpcR* **R** package [11]. The main functions (e.g., for analysis, simulations, plotting), several auxiliary helper functions (e.g., data import) and data set of different dPCR systems are listed in Table XY. Further dependencies to 3rd party packages include *pracma*, ... See the vignette for details.

3 METHODS

4 DISCUSSION

There are currently different software solutions for dPCR analysis such as the OpenArray software (Life Technologies) or XYZ (Bio-Rad). Most of the are black boxes which prevent deep insight into the data processing step. In addition most of the software solutions are aimed to be used in very specific scenarios and a mutual exclusive to alternative platforms (e.g., droplet vs. chamber-based). We have chosen **R** because it is the *lingua franca* in biostatistics and broadly used in other disciplines [13].

5 CONCLUSION

In conclusion, *dpcR* provides means to understand how digital PCR works, to design, simulate and analyze experiments, and to verify their results (e.g., confidence interval estimation), which should ultimately improve reproducibility. We have built what we believe to be the first unified, cross-platform, dMIQE compliant, open source (GPL-3 or later) software frame-work for analyzing digital PCR experiments. Our frame-work, designated *dpcR* , is targeted at a broad user base including end users in clinics, academics, developers, and educators. We implemented existing statistical methods for dPCR and suggest the introduction of a standardized nomenclature for qPCR. Our frame-work is suitable for teaching and includes references for an elaborated set of methods for dPCR statistics. Our software can be used for (I) data analysis and visualization in research, (II) as software frame-work for novel technical developments, (III) as platform for teaching this new technology and (IV) as reference for statistical methods with a standardized nomenclature for dPCR experiments. The package enables the simulations and predictions of Poisson distribution for dPCR scenarios, the analysis of previously run dPCRs. Due to the plug-in structure of the software it is possible to build custom-made analyzers.

ACKNOWLEDGEMENT

Grateful thanks belong to the **R** community.

Funding: This work was funded by the Federal Ministry of Education and Research (BMBF) InnoProfile–Transfer–Projekt 03 IPT 611X.

Conflict of Interest: none declared.

REFERENCES

[1]Yoav Benjamini and Yosef Hochberg. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1):289–300, 1995.

[2]Lawrence D. Brown, T. Tony Cai, and Anirban DasGupta. Interval estimation for a binomial proportion. *Statist. Sci.*, 16(2):101–133, May 2001.

[3]Michał Burdukiewicz, Stefan Rödiger, Bart Jacobs, and Piotr Sobczyk. *dpcR: Digital PCR Analysis*, 2015. R package version 0.2.

[4]Winston Chang, Joe Cheng, JJ Allaire, Yihui Xie, and Jonathan McPherson. *shiny: Web Application Framework for R*, 2015. R package version 0.12.2.

[5]Michael Fay. Two-sided exact tests and matching confidence intervals for discrete data. *Proceedings of the National Academy of Sciences of the United States of America*, 2(1):53–58, June 2010.

[6]Jim F. Huggett, Simon Cowen, and Carole A. Foy. Considerations for Digital PCR as an Accurate Molecular Diagnostic Tool. *Clinical Chemistry*, page clinchem.2014.221366, October 2014.

[7]David Kahle and Hadley Wickham. ggmap: Spatial visualization with ggplot2. *The R Journal*, 5(1):144–162, 2013.

[8]Alexander A. Morley. Digital PCR: A brief history. *Biomolecular Detection and Quantification*, 1(1):1–2, 2014.

[9]Jeroen Ooms. Directions for improved dependency versioning in R. *The R Journal*, 5(1):197–207, 2013.

[10]Stephan Pabinger, Stefan Rödiger, Albert Kriegner, Klemens Vierlinger, and Andreas Weinhäusel. A survey of tools for the analysis of quantitative PCR (qPCR) data. *Biomolecular Detection and Quantification*, 1(1):23–33, 2014.

[11]Christian Ritz and Andrej-Nikolai Spiess. qpcR: an R package for sigmoidal model selection in quantitative real-time polymerase chain reaction analysis. *Bioinformatics*, 24(13):1549–1551, January 2008.

[12]Stefan Rödiger, Alexander Böhm, and Ingolf Schimke. Surface Melting Curve Analysis with R. *The R Journal*, 5(2):37–53, December 2013.

[13]Stefan Rödiger, Michał Burdukiewicz, Konstantin A. Blagodatskikh, and Peter Schierack. R as an Environment for the Reproducible Analysis of DNA Amplification Experiments. *The R Journal*, 7(2):127–150, 2015.

[14]Stefan Rödiger, Michał Burdukiewicz, and Peter Schierack. chipPCR: an R Package to Pre-Process Raw Data of Amplification Curves. *Bioinformatics*, page btv205, April 2015.

[15]David A. Selck, Mikhail A. Karymov, Bing Sun, and Rustem F. Ismagilov. Increased Robustness of Single-Molecule Counting with Microfluidics, Digital Isothermal Amplification, and a Mobile Phone versus Real-Time Kinetic Measurements. *Analytical Chemistry*, 85(22):11129–11136, November 2013.

[16]Andrej-Nikolai Spiess, Claudia Deutschmann, Michał Burdukiewicz, Ralf Himmelreich, Katharina Klat, Peter Schierack, and Stefan Rödiger. Impact of Smoothing on Parameter Estimation in Quantitative DNA Amplification Experiments. *Clinical Chemistry*, 61(2):379–388, January 2015.