BIOINFORMATICS

Vol. 00 no. 00 2014
Pages 1–2

# chipPCR: an R Package to Pre-Process Amplification Curve Data

Stefan Rödiger [1]*, Michał Burdukiewicz [2] and Peter Schierack [1]

[1]Faculty of Natural Sciences, Brandenburg University of Technology Cottbus–Senftenberg, Senftenberg, Germany
[2]Department of Genomics, Faculty of Biotechnology, University of Wrocław, Wrocław, Poland

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Associate Editor: XXXXXXX

**ABSTRACT**

**Motivation:** The quantitative real-time polymerase chain reaction (qPCR) and quantitative isothermal amplification (qIA) are standard methods for nucleic acids quantification. Numerous real-time read-out technologies with different technical foundations have been developed. Despite the continuous interest in amplification based techniques, there are only few tools for amplification data pre-processing. Especially, during development of new instruments a transparent tool for precise control of raw data is indispensable.

**Results:** *chipPCR* is an **R** package for amplification curve data pre-processing and quality analysis. The package is takes advantage of **R**'s *S4* object model and offers an extensible environment. *chipPCR* contains tools for the raw data exploration: normalization, baselining, imputation of missing values, a powerful wrapper for the amplification curve smoother and a function to detect the start and end of an amplification curve. Capabilities of the software are enhanced by implementation of algorithms yet not present in **R**, as a 5-point stencil for derivative interpolation. Simulation tools, statistical tests, plots for data quality management, amplification efficiency/quantification cycle calculation, and data sets from various qPCR and qIA experiments are also part of the package. The core functionalities of *chipPCR* are integrated in GUIs (web-based and standalone *shiny* applications) streamlining analysis and report generation.

**Availability:** Stable: http://cran.r-project.org/web/packages/chipPCR
Source code: https://github.com/michbur/chipPCR

**Contact:** stefan.roediger@b-tu.de

**Supplementary:** Supplementary data are available at Bioinformatics online.

## 1 INTRODUCTION

Quantitative polymerase chain reaction (qPCR) and quantitative isothermal amplification (qIA) are standard methods to amplify nucleic acids. qPCR and qIA are used in real-time monitoring technologies, such as our previously reported VideoScan technology (Rödiger *et al.*, 2013a; Spiess *et al.*, 2015), microfluidics and point-of-care devices to quantify nucleic acids by specific curve parameters like the quantification point (Cq) (Pabinger *et al.*, 2014; Rödiger *et al.*, 2014). Fundamental steps of amplification curve analysis are: 1) raw data read-in, 2) pre-processing (e.g., noise reduction), 3) amplification curve processing (e.g., Cq calculation), 4) post-processing and 5) data export/report generation. Reliable data flow between all steps is a requirement for the proper optimization (e.g., Taguchi method) of amplification reactions (Cobb and Clarkson, 1994). **R** is widely used bioinformatics and an rapid adopter for novel technologies (e.g., digital PCR, NanoString nCounter Platform) (Waggott *et al.*, 2012; Pabinger *et al.*, 2014). Available **R** packages focus on the read-in and (post)-processing of data from commercial qPCR systems. **R** packages for the steps 1 and 3–5 are available (Perkins *et al.*, 2012; Gehlenborg *et al.*, 2013; McCall *et al.*, 2014; Pabinger *et al.*, 2014). However, there is no **R** package for raw amplification curve data pre-processing and quality analysis. Pre-processing in most commercial cyclers is a black box, which sets severe limits to reproducible research (Leeper, 2014). Developmental equipment depends on software to pre-process the raw data. Pre-processing algorithms remove stochastic errors and artefacts (Suppl. Sect. 2) and provide means for raw data inspection and transformation in a format for successive analysis steps (e.g., smoothing, imputation), data reduction (e.g., removal of invalid sets) and data quality management. Misinterpretations are more likely if arbitrary corrections are performed. A manual alteration is in contradiction to reproducible research.

The *chipPCR* package ("Lab-on-a-**Chip**" & **PCR**) was developed to automatize pre-processing, analysis, visualization, and quality control of qPCR and qIA experiments. **R** offers sophisticated statistical tools and for reproducible cross-platform analysis and the adoption to changing experimental setups. Moreover, it is desirable to set up workflows in an open environment, which offers GUIs, downstream analyses facilities, powerful tools for data visualizations and report generation. The target audience encompasses developers and users who process raw data of commercial systems.

## 2 IMPLEMENTATION

We implemented the *chipPCR* package in the **R** software environment. *chipPCR* is a relative of the *MBmca* (Rödiger *et al.*, 2013b), the *RDML* (Blagodatskikh *et al.*, 2014), and the *dpcR* (Pabinger *et al.*, 2014) packages but focusses on pre-processing of amplification curves. The package contains pre-processor functions (smoothing, imputation, background range detection, baseline correction and normalization), a single-blinded randomized rating function, quality analysis summary functions, an amplification efficiency function, an amplification curve simulation function

*to whom correspondence should be addressed

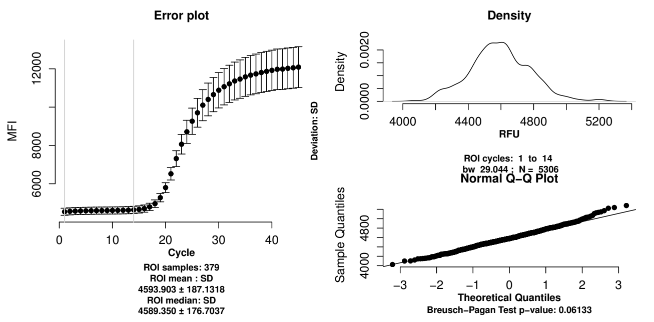


**Fig. 1.** *MFIaggr* plot for 379 replicate amplification curves. Cycles 1 to 14 were selected as region of interest (ROI) to analyze the cycle-dependent variance (left panel), the density plot (top-right panel) and quantile-quantile analysis (bottom-right panel), including a comprehensive statistical analysis as textual output (not shown). Plots indicate that the data of the background range are normal distributed. The heteroscedasticity is not significant.

and report generation (Suppl. Sect. 4). The supplemental material uses Donald Knuth's literate programming principle (Knuth, 1984) to present the source code conveniently. *chipPCR*'s naming convention is *period.separated* (Bååth, 2012). We use **R**'s object model *S4* class system (see Supplement) to separate between interface and implementation. *S4* classes require a higher effort than *S3*, but assures better control on the object structure and the method dispatch. For fast code in high-throughput applications, we avoided loops and left options for partially parallel computing usage (e.g., *smoother* function). *chipPCR* includes a set of classes for plotting. The output of our custom made plots is minimalistic, but many parameters can be adjusted directly or by the ellipse parameter.

We aim to make our software available for researchers not fluent in **R**. Therefore, we implemented core functionality of our package in selected GUI technologies available in **R** (Rödiger *et al.*, 2012) as desktop application or web-based service. *chipPCR* offers means to run the GUI applications as service on a server without installing **R** (e.g., http://michbur.shinyapps.io/MFIaggr_gui), on the local desktop (e.g., Fig. S2, S6), or as deployed from an external source for a local **R** installation. The functions *AmpSim*, *th.cyc*, *bg.max* and *amptester* are part of online GUIs. We aimed to build monolithic systems to parse, pre-process and analyze amplification curve data in a combined work-flow.

*chipPCR* relies solely on the native **R** workspace and dedicated **R** packages as default data import and export format (Perkins *et al.*, 2012; Rödiger *et al.*, 2012; Blagodatskikh *et al.*, 2014). *chipPCR* presents *S4* objects with tailored summary and plot methods. Since data sets are an essential element of reproducible research (Leeper, 2014), we included 22 data sets from commercial and experimental cyclers (e.g., helicase dependent amplification (HDA)).

## 3 EXAMPLE: QUALITY ANALYSIS

*MFIaggr* is a versatile analytical and graphical tool for fast multiple comparison of the cycle dependent signal dispersion and distribution (Fig. 1). The continuous explanatory variable $x$ (cycle number) is used to describe its relationships to $n$ continuous predictor variables $y_i$ (fluorescence values), where $i \in \{1, ..., n\}$. Use cases include the comparison of independent reaction vessels or the analysis of replicate experiments (Suppl. Sect. 6). In particular, this function might be useful for quality management during the development of high-throughput technologies. An analysis via the *shiny MFIaggr.gui* app is shown in Fig. S7.

## 4 RESULTS AND CONCLUSIONS

*chipPCR* is the first **R** package for amplification curve raw data pre-processing and quality analysis. *chipPCR* primarily targets pre-processing but standard methods to process amplification curves were also implemented. Functions of *chipPCR* are embeddable in customized routines with other packages (see Suppl.), such as the *RDML* and *MBmca* packages. We claim that the modular structure of *chipPCR* package allows users to perform flexible data analysis adjusted to their needs. Users can do estimations by hand. For example for Cq (SDM) estimation, solely the *chipPCR* functions *inder* and *smoother* are needed. *smoother* will be a method of smoothing in *inder* and by putting data in the *bg* object with summary method *summary-der* the user obtains the Cq. Thanks to GUI's it should be easy even for an user without any **R** experience omitting a big limitation of **R** packages related to qPCR and qIA.

## ACKNOWLEDGEMENT

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

*Funding*: This work was funded by the BMBF InnoProfile–Transfer–Projekt 03 IPT 611X.

*Conflict of Interest*: none declared.

## REFERENCES

Bååth, R. (2012). The State of Naming Conventions in R. *The R Journal*, **4**(2), 74–75.

Blagodatskikh, K. A., Roediger, S., and Burdukiewicz, M. (2014). Importing real-time thermo cycler (qPCR) data from RDML format files.

Cobb, B. D. and Clarkson, J. M. (1994). A simple procedure for optimising the polymerase chain reaction (PCR) using modified Taguchi methods. *Nucleic Acids Research*, **22**(18), 3801–3805.

Gehlenborg, N., Noble, M. S., Getz, G., Chin, L., and Park, P. J. (2013). Nozzle: a report generation toolkit for data analysis pipelines. *Bioinformatics*, **29**(8), 1089–1091.

Knuth, D. E. (1984). Literate Programming. *The Computer Journal*, **27**(2), 97–111.

Leeper, T. J. (2014). Archiving Reproducible Research and Dataverse with R. *The R Journal*, **6**(1), 151–158.

McCall, M. N., McMurray, H. R., Land, H., and Almudevar, A. (2014). On non-detects in qPCR data. *Bioinformatics*.

Pabinger, S., Rödiger, S., Kriegner, A., Vierlinger, K., and Weinhäusel, A. (2014). A survey of tools for the analysis of quantitative PCR (qPCR) data. *Biomolecular Detection and Quantification*.

Perkins, J. R., Dawes, J. M., McMahon, S. B., Bennett, D. L., Orengo, C., and Kohl, M. (2012). ReadqPCR and NormqPCR: R packages for the reading, quality checking and normalisation of RT-qPCR quantification cycle (Cq) data. *BMC Genomics*, **13**(1), 296.

Rödiger, S., Friedrichsmeier, T., Kapat, P., and Michalke, M. (2012). RKWard: A Comprehensive Graphical User Interface and Integrated Development Environment for Statistical Analysis with R. *Journal of Statistical Software*, **49**(9), 1–34.

Rödiger, S., Schierack, P., Böhm, A., Nitschke, J., Berger, I., Frömmel, U., Schmidt, C., Ruhland, M., Schimke, I., Roggenbuck, D., Lehmann, W., and Schröder, C. (2013a). A highly versatile microscope imaging technology platform for the multiplex real-time detection of biomolecules and autoimmune antibodies. *Advances in Biochemical Engineering/Biotechnology*, **133**, 35–74.

Rödiger, S., Böhm, A., and Schimke, I. (2013b). Surface Melting Curve Analysis with R. *The R Journal*, **5**(2), 37–53.

Rödiger, S., Liebsch, C., Schmidt, C., Lehmann, W., Resch-Genger, U., Schedler, U., and Schierack, P. (2014). Nucleic acid detection based on the use of microbeads: a review. *Microchimica Acta*, **181**(11-12), 1151–1168.

Spiess, A.-N., Deutschmann, C., Burdukiewicz, M., Himmelreich, R., Klat, K., Schierack, P., and Rödiger, S. (2015). Impact of smoothing on parameter estimation in quantitative DNA amplification experiments. *Clinical Chemistry*.

Waggott, D., Chu, K., Yin, S., Wouters, B. G., Liu, F.-F., and Boutros, P. C. (2012). NanoStringNorm: An Extensible R Package For the Pre-Processing of NanoString mRNA and miRNA Data. *Bioinformatics*, pages 1546–1548.