Bioinformatics Advance Access published April 24, 2015

chipPCR: an R Package to Pre-Process Raw Data of Amplification Curves

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

Associate Editor: Dr. Ziv Bar-Joseph

ABSTRACT

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

Results: *chipPCR* is an **R** package for the pre-processing and quality analysis of raw data of amplification curves. The package takes advantage of **R**'s *S4* object model and offers an extensible environment. *chipPCR* contains tools for raw data exploration: normalization, baselining, imputation of missing values, a powerful wrapper for amplification curve smoothing and a function to detect the start and end of an amplification curve. The capabilities of the software are enhanced by the implementation of algorithms unavailable in **R**, such 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 qPCR and qIA experiments are part of the package. Core functionalities are integrated in GUIs (web-based and standalone *shiny* applications), thus streamlining analysis and report generation.

Availability: http://cran.r-project.org/web/packages/chipPCR
newline 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 used for nucleic acid amplification. 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). The 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., the Taguchi method) of amplification reactions (Cobb and Clarkson, 1994). **R** is widely used in bioinformatics and an early adopter of 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 amplification analysis steps 1 and 3–5 cited above are available (Perkins *et al.*, 2012; Gehlenborg *et al.*, 2013; McCall *et al.*, 2014; Pabinger *et al.*, 2014). However, **R** packages for the pre-processing and quality analysis of raw data of amplification curves are unavailable. Pre-processing in most commercial cyclers is a black box, which restrains reproducible research (Leeper, 2014). The development and optimization of equipment would benefit from the availability of a software capable of pre-processing raw data. Pre-processing algorithms remove stochastic errors and artefacts (Suppl. Sect. 2) and provide the means for raw data inspection and transformation in a format suitable 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 and a manual alteration is contradictory to reproducible research.

The *chipPCR* ("Lab-on-a-Chip" & PCR) package was developed to automatize pre-processing, analysis, visualization, and quality control of qPCR and qIA experiments. **R** enables sophisticated statistical and reproducible cross-platform analysis, and quick adaptation to changing experimental setups. Moreover, it is advantageous to set up workflows in an open environment, which offers GUIs, downstream analyses facilities, powerful data visualizations and report generation. The target audience encompasses developers and users who process raw data from 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* (Rödiger *et al.*, 2015), 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

*to whom correspondence should be addressed

range detection, baseline correction and normalization), a single-blinded randomized rating function, quality analysis summary functions, an amplification efficiency function, an amplification curve simulator and a report generation function (Suppl. Sect. 4). The supplemental material uses Donald Knuth's literate programming principle (Knuth, 1984) to conveniently present the source code. *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 running of codes 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 have implemented core functionality of our package in selected GUI technologies available in **R** (Rödiger *et al.*, 2012) as a desktop application or web-based service. *chipPCR* offers the means to run the GUI applications as a service on a server without installing **R** (e.g., http://michbur.shinyapps.io/MFIaggr_gui, on a 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.*, 2015). *chipPCR* presents *S4* objects with tailored summary and plot methods. Since data sets are an essential element of reproducible research (Leeper, 2014), we have included 22 data sets from commercial and experimental cyclers to this package.

3 EXAMPLE: QUALITY ANALYSIS

MFIaggr is a versatile analytical and graphical tool for fast multiple comparison of cycle-dependent signal dispersion and distribution (Fig. 1; Ruijter *et al.* (2013) data set). 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 the pre-processing and quality analysis of amplification curve raw data. In addition, we implemented standard methods for amplification curve processing. The *chipPCR* functions are embeddable in customized routines with other packages (see Suppl.), such as the *RDML* and *MBmca* packages. The modular package structure enables flexible data analysis adaptable to the requirements. For example, solely the functions *inder* and *smoother* are needed to estimate the Cq (SDM). *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 for users without any **R** experience.

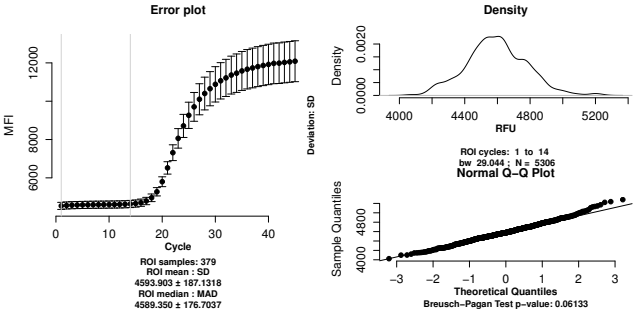

Fig. 1. *MFIaggr* plot for 379 replicate amplification curves. Cycles 1 to 14 were selected as region of interest (ROI) to analyze the cycle-dependent variance (left panel), the density plot (top-right panel) and quantile-quantile analysis (bottom-right panel), including a comprehensive statistical analysis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

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

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

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*, **30**(16), 2310–6.

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*, **1**(1), 23–33.

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.

Rödiger, S., Burdukiewicz, M., Blagodatskikh, K. A., and Schierack, P. (2015). R as an Environment for the Reproducible Analysis of DNA Amplification Experiments. *The R Journal*, **7**(2), 1–24.

Ruijter, J. M., Pfaffl, M. W., Zhao, S., Spiess, A. N., Boggy, G., Blom, J., Rutledge, R. G., Sisti, D., Lievens, A., De Preter, K., Derveaux, S., Hellemans, J., and

Vandesompele, J. (2013). Evaluation of qPCR curve analysis methods for reliable biomarker discovery: bias, resolution, precision, and implications. *Methods (San Diego, Calif.)*, **59**(1), 32–46.

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*, **61**(2), 379–388.

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*, **28**(11), 1546–1548.