

Databases and ontologies

Enabling reproducible real-time quantitative PCR research: the RDML package

Stefan Rödiger¹, Michał Burdukiewicz², Andrej-Nikolai Spiess³ and Konstantin Blagodatskikh^{4,*}

¹Institute of Biotechnology, Brandenburg University of Technology Cottbus–Senftenberg, Senftenberg, Germany, ²Department of Genomics, Faculty of Biotechnology, University of Wrocław, Wrocław, Poland, ³Department of Andrology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany and ⁴Evrogen JSC, Moscow, Russia

*To whom correspondence should be addressed.

Associate Editor: Janet Kelso

Received on May 3, 2017; revised on July 28, 2017; editorial decision on August 7, 2017; accepted on August 24, 2017

Abstract

Motivation: Reproducibility, a cornerstone of research, requires defined data formats, which include the setup and output of experiments. The real-time PCR data markup language (RDML) is a recommended standard of the minimum information for publication of quantitative real-time PCR experiments guidelines. Despite the popularity of the RDML format for analysis of quantitative PCR data, handling of RDML files is not yet widely supported in all PCR curve analysis softwares.

Results: This study describes the open-source *RDML* package for the statistical computing language **R**. *RDML* is compatible with RDML versions ≤ 1.2 and provides functionality to (i) import RDML data; (ii) extract sample information (e.g. targets and concentration); (iii) transform data to various formats of the **R** environment; (iv) generate human-readable run summaries; and (v) to create RDML files from user data. In addition, *RDML* offers a graphical user interface to read, edit and create RDML files.

Availability and implementation: <https://cran.r-project.org/package=RDML>. *rdmlEdit* server <http://shtest.evrogen.net/rdmlEdit/>. Documentation: <http://kabl原因.github.io/RDML/>.

Contact: k.blag@yandex.ru

Supplementary information: [Supplementary data](#) are available at *Bioinformatics* online.

1 Introduction

Real-time quantitative PCR (qPCR) is one of the most widely applied methods in molecular biology, diagnostics, forensics and genetic testing. The popularity of this method resulted in a vast array of diverse qPCR systems. Typically, these devices use different data processing methods and non-interchangeable storage formats (e.g. binary formats). A unified data format enables the comparisons of PCR results from different systems, and disparate storage formats prohibit the processing of raw data from one instrument in other systems. This is essentially contraproductive as data processing poses a challenge in qPCR data analysis (Spiess *et al.*, 2015, 2016).

To facilitate the comparison of experimental results from qPCRs and preserve information necessary to reproduce the runs, the real-time PCR data markup language (RDML) was developed (Lefever *et al.*, 2009). It contains the primary raw data acquired by the qPCR system as well as meta-information required to understand the experimental setup (e.g. sample annotation, qPCR protocol, probe and primer sequences). RDML, as the unified data format, facilitates the reproducibility of runs and is suggested as the standard qPCR interchange data format by the minimum information for publication of quantitative real-time PCR experiments (Bustin *et al.*, 2009). It is supported by numerous qPCR machines vendors and third-party software.

Despite the popularity of the RDML format for analysis of qPCR data, there is no open-source biostatistical computing

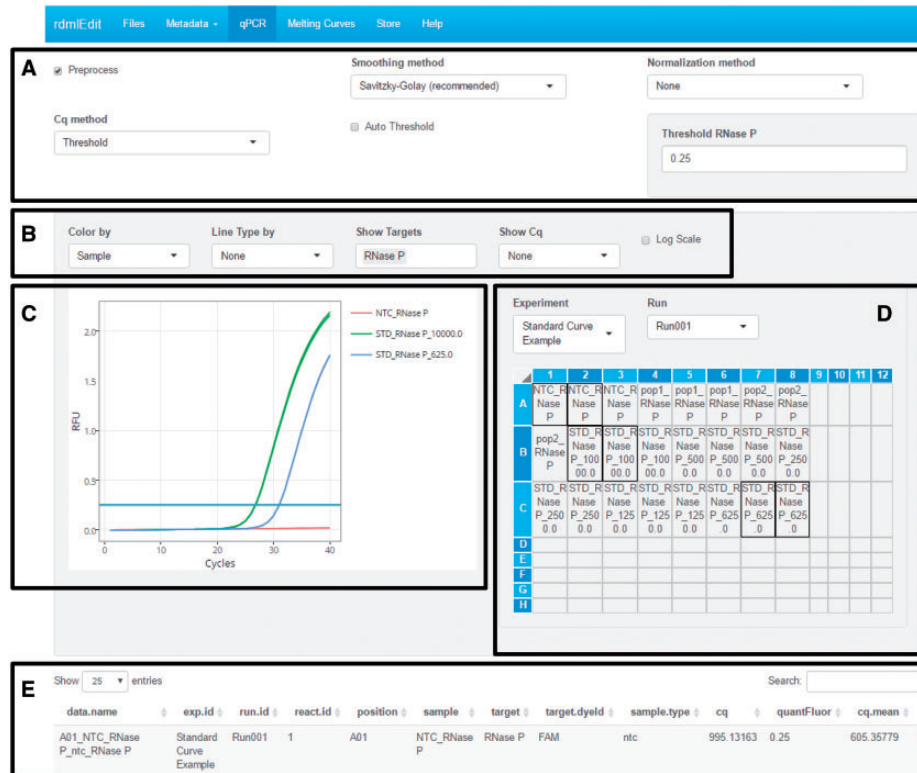


Fig. 1. The GUI has tools to view and analyze the raw data of the amplification (tab 'qPCR') and melting curves (tab 'Melting Curves'). (A) Data preprocessing and Cq computation; (B) plot settings; (C) interactive amplification plot; (D) plate view with tubes selectors and (E) summary table with filtering options

software capable of reading, processing and writing this format. Instead, users have to choose between programs from manufacturers of qPCR devices, which are limited to files generated by their own systems, or commercial software such as *qbase+*. Ruijter *et al.* (2015a) published the open-source desktop software *RDML-Ninja*, which can visualize, edit and validate RDML files. However, *RDML-Ninja* cannot transfer qPCR data to biostatistical analysis pipelines, which limits its applications. Moreover, there is a need for a *de novo* creation of RDML data in systems that do not support the RDML format (e.g. commercial systems, prototypes).

Although the cross-platform statistical computing language R includes several packages for qPCR and melting curve analysis (Pabinger *et al.*, 2014; Rödiger *et al.*, 2015b), there were previously no tools supporting the seamless processing of RDML files. The common strategy for data management in R is native formats such as R workspaces and objects. Hence, the principle of reproducible research is not warranted without a standard method of data import such as the *RDML* package. Our software enables the application of qPCR-related R tools, while working on RDML data directly derived from the qPCR system (Supplementary Information S1 and S2). The main target audience of the *RDML* R package is data analysis experts and software developers who develop customized qPCR analysis pipelines (Supplementary Information S2). However, the *RDML* package is the foundation of the shiny Graphical User Interface (GUI) *rdmlEdit* (Supplementary Information S1). This GUI addresses the needs of users with little or no programming experience (e.g. biologists).

2 Implementation

The *RDML* package allows to exchange RDML files and transform them to a human-readable format. It provides R6 classes, which

corresponds to *RDML* \leq v. 1.2 input format types. Central functionalities of the *RDML* package encompass the read-in of RDML data files and the summary generation for RDML objects.

In contrast to other RDML-importing software, the *RDML* package covers also data preprocessing (e.g. smoothing) and analysis, estimation of the Cq (quantification cycle value) based on the second derivative maximum method or cycle threshold method. It is recommended that more advanced data processing is performed (using more specialized packages) since the reported Cq values are indicative only. In addition, the package supports the creation of RDML objects from qPCR data of systems that do not support the RDML format. To create a new RDML object, one has to provide fluorescence data and a minimal description of the run (Supplementary Information S1).

Another feature unique to the *RDML* package is the ability to merge several RDML files into one single file with the *MergeRDMLs()* function and process it like a unit. For example, one can combine two runs with samples of one biological experiment or add calibration samples to a run with genes-of-interest, see Supplementary Information S1. The *RDML* package does not handle interrun variation. However, recently a study (Ruijter *et al.*, 2015b) showed how the export to the RDML format can be used in an analysis pipeline (raw fluorescence data \rightarrow amplification curve analysis \rightarrow removal of interrun variation and statistical analysis) of qPCR data. The same applies to handling of missing values, which is also a challenge during the analysis of qPCR data (Ronde *et al.*, 2017). There are further R packages on Comprehensive R Archive Network (CRAN), Bioconductor and GitHub, which can be chained with the *RDML* package to deal with non-detects (McCall *et al.*, 2014), expression analysis (Dvinge and Bertone, 2009; Matz *et al.*, 2013; Perkins *et al.*, 2012; Rödiger *et al.*, 2015a), periodicity in qPCR run data (Spiess *et al.*, 2016) and melting

curve data (Ritz and Spiess, 2008; Rödiger *et al.*, 2013). An example of an analysis pipeline is given in [Supplementary Information S2](#).

All functions provided by the package, including data manipulation and analysis, are available in a graphical user interface, **rdmlEdit** (Fig. 1), which is available as a web server. Users can also deploy their own local services on any computing system that has a working R environment running ([Supplementary Information S1](#)). Concluding, the RDML package allows:

- import of RDML data,
- export of data in the RDML format,
- basic analysis of RDML data and
- human-readable summary of RDML data.

3 Discussions and conclusions

The RDML package for R processes data from RDML \leq v. 1.2 format files and can create RDML v. 1.2 files from user provided data. It is the first dedicated software for the RDML data format that supports merging and creating new RDML files.

This package can be used as part of the qPCR processing workflow or for preliminary summaries of experiments. The largest benefit is the opening of RDML data for further statistical analysis with dedicated algorithms already provided in the R environment. The plethora of features, in addition to classical qPCR curves, allows RDML to be a foundation of machine learning methods for amplification curve analysis. To conclude, the RDML package may be used in pipelines with other R packages (e.g. *qpcR* as described in the [Supplementary Information S2](#)) or as part of web servers as shown by us and others (Mallona *et al.*, 2017). Thanks to the **rdmlEdit** web server, our software may help scientists who are less fluent in R, and thus allow them to easier access and process their experimental data that is saved in the RDML format.

Acknowledgement

Grateful thanks belong to the R community and the RDML consortium.

Funding

This work was funded by the Federal Ministry of Education and Research (BMBF) InnoProfile-Transfer-Projekt 03IPT611X and in part by ‘digilog: Digitale und analoge Begleiter für eine alternde Bevölkerung’

(Gesundheitscampus Brandenburg, Brandenburg Ministry for Science, Research and Culture).

Conflict of Interest: none declared.

References

- Bustin, S.A. *et al.* (2009) The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin. Chem.*, **55**, 611–622.
- Dvinge, H. and Bertone, P. (2009) HTqPCR: high-throughput analysis and visualization of quantitative real-time PCR data in R. *Bioinformatics*, **25**, 3325–3326.
- Lefever, S. *et al.* (2009) RDML: structured language and reporting guidelines for real-time quantitative PCR data. *Nucleic Acids Res.*, **37**, 2065–2069.
- Mallona, I. *et al.* (2017) Chainy: an universal tool for standardized relative quantification in real-time PCR. *Bioinformatics*, **33**, 1411.
- Matz, M.V. *et al.* (2013) No control genes required: Bayesian analysis of qRT-PCR data. *PLoS One*, **8**, e71448.
- McCall, M.N. *et al.* (2014) On non-detects in qPCR data. *Bioinformatics*, **30**, 2310–2316.
- Pabinger, S. *et al.* (2014) A survey of tools for the analysis of quantitative PCR (qPCR) data. *Biomol. Detect. Quantif.*, **1**, 23–33.
- Perkins, J.R. *et al.* (2012) ReadqPCR and NormqPCR: R packages for the reading, quality checking and normalisation of RT-qPCR quantification cycle (Cq) data. *BMC Genomics*, **13**, 296.
- Ritz, C. and Spiess, A.-N. (2008) qpcR: an R package for sigmoidal model selection in quantitative real-time polymerase chain reaction analysis. *Bioinformatics*, **24**, 1549–1551.
- Rödiger, S. *et al.* (2013) Surface melting curve analysis with R. *R J.*, **5**, 37–53.
- Rödiger, S. *et al.* (2015a) chipPCR: an R package to pre-process raw data of amplification curves. *Bioinformatics*, **31**, 2900–2902.
- Rödiger, S. *et al.* (2015b) R as an environment for the reproducible analysis of DNA amplification experiments. *R J.*, **7**, 127–150.
- Ronde, M. W. J. D. *et al.* (2017) Practical data handling pipeline improves performance of qPCR-based circulating miRNA measurements. *RNA*, **23**, 811–821.
- Ruijter, J.M. *et al.* (2015a) RDML-Ninja and RDMLdb for standardized exchange of qPCR data. *BMC Bioinformatics*, **16**, 197.
- Ruijter, J.M. *et al.* (2015b) Removal of between-run variation in a multi-plate qPCR experiment. *Biomol. Detect. Quantif.*, **5**, 10–14.
- Spiess, A.-N. *et al.* (2015) Impact of smoothing on parameter estimation in quantitative DNA amplification experiments. *Clin. Chem.*, **61**, 379–388.
- Spiess, A.-N. *et al.* (2016) System-specific periodicity in quantitative real-time polymerase chain reaction data questions threshold-based quantitation. *Sci. Rep.*, **6**, 38951.