R as an Environment for the Analysis of dPCR and qPCR Experiments

by Stefan Rödiger, Michał Burdukiewicz, Konstantin Blagodatskikh, Michael Jahn and Peter Schierack

Abstract There is an ever-increasing number of applications, which use quantitative PCR (qPCR) or digital PCR (dPCR) to elicit fundamentals of biological processes. Moreover, the novel amplification strategies based on quantitative isothermal amplification (qIA) have become more prominent in life sciences and point-of-care-diagnostics. Additionally, the analysis of melting data is essential during many experiments. Several software packages have been developed for the analysis of such datasets. In most cases, the software is either distributed as closed source software or as monolithic block with little freedom to perform highly customized analysis procedures. We argue, among others, that R is an excellent foundation for reproducible and transparent data analysis in a highly customizable cross-platform environment. However, for novices it is often challenging to master R or learn capabilities of the vast number of packages available. In the paper, we describe exemplary workflows for the analysis of qPCR, qIA or dPCR experiments including the analysis of melting curve data. Our analysis relies entirely on R packages available from public repositories. Additionally, we provide information related to standardized and reproducible research.

Introduction

The quantitative Polymerase Chain Reaction (qPCR) is the method of choice when a precise quantification of minute DNA traces is required. The applications include the detection and quantification of pathogens or gene expression analysis (Pabinger et al., 2014). Only few bioanalytical methods had such a significant impact on the progress of life sciences and medical sciences as the qPCR (Huggett et al., 2015). Numerous commercial and experimental monitoring platforms have been developed in the past years. This includes standard plate cyclers, capillary cyclers, microfluidic platforms and related technologies (Rödiger et al., 2013b; Viturro et al., 2014; Rödiger et al., 2014; Khodakov and Ellis, 2014; Wu et al., 2014).

In the past decades several isothermal amplification technologies emerged, such as helicase dependent amplification (HDA). Isothermal amplification methods were readily combined with real-time monitoring technologies (qIA) or digital PCR and is used in various fields like diagnostics and point-of-care-testing (Selck et al., 2013; Rödiger et al., 2014; Nixon et al., 2014).

The digital PCR (dPCR) is a novel approach for detection and quantification of nucleic acids and can be seen as a next generation method (Huggett et al., 2015). The dPCR technology breaks fundamentally with the previous concept of nucleic acid quantification. The key difference between dPCR and traditional qPCR lies in the method of measuring (absolute) nucleic acids amounts, which yields discrete information instead of the continuous signal. This is possible after "clonal DNA amplification" in thousands of small separated partitions (e.g., droplets, nano chambers) (Huggett et al., 2013; Milbury et al., 2014; Morley, 2014). The partitions with no nucleic acid remain negative and the others turn positive (e.g., Figure 1). Selected technologies (e.g., OpenArray®Real-Time PCR System) monitor amplification reactions in the chambers ("partitions") in real-time. After that, all quantification cycle (Cq) values are calculated from the amplification curves and converted into discrete events by means of positive and negative chambers. Finally, the absolute quantification of nucleic acids is done using Poisson statistics. Recently, we have published the **dpcR** package at CRAN, which is the first open source R software package for the analysis of dPCR experiments (see **dpcR** for details).

The complexity of hardware, wetware and software requires expertise to master a workflow. This comprises standards for experiment design, generation and analysis of data, multiple hypothesis testing, interpretation, reporting and storage of results (Huggett et al., 2014a; Castro-Conde and de Uña Álvarez, 2014). Scientific misconduct and fraud have shaken the scientific community on several occasions (Bustin, 2014). In particular, the scientific community works hard to uncover pitfalls of qPCR experiments. This lead to the development of peer-reviewed quantification cycle (Cq) analysis algorithms (Ruijter et al., 2013), fully characterized qPCR chemistries (Ruijter et al., 2014) and guidelines for a proper conduct of qPCR experiments as implemented in the MIQE guidelines (minimum information for publication of quantitative real-time PCR experiments) (Huggett et al., 2013; Bustin, 2014). We share the philosophy of the MIQE guidelines to increase the experimental transparency for better experimental practice and reliable interpretation of results and encourage the use of open data exchange formats like the XML-based Real-Time PCR Data Markup Language (RDML) (Lefever et al., 2009). We see the application of R in line with this philosophy.

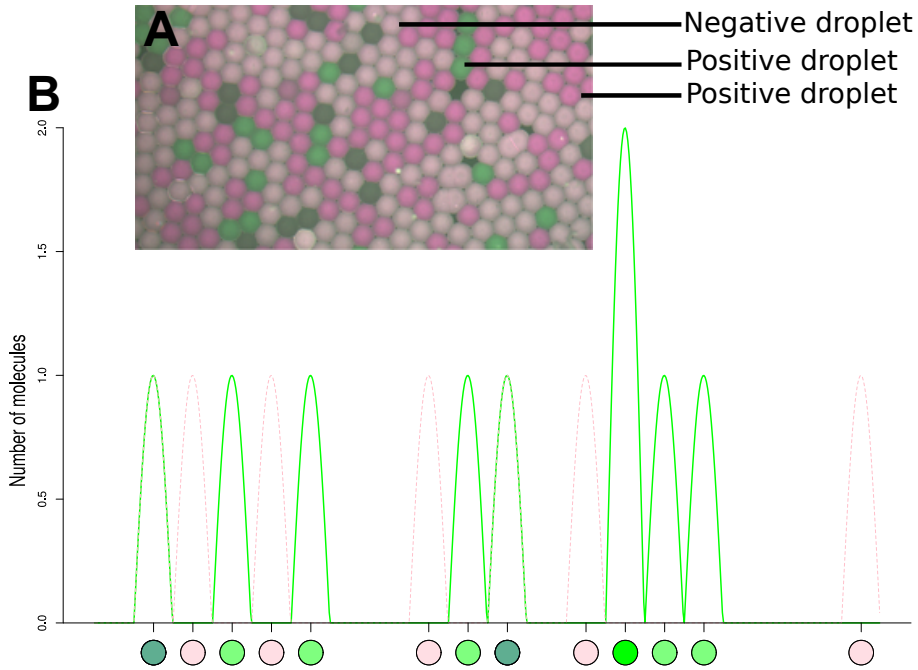

Figure 1: Scheme of a digital PCR experiment. **(A)** A droplet digital PCR reaction mix was formed in a BioRad QX100 Droplet Digital PCR System. The droplets (circa 100 μm in diameter) were subjected to custom made slide chambers for the detection and analysis in fully automatized imaging VideoScan platform (Rödiger et al., 2013b). **(B)** Subsequently the samples can be digitalized by counting number of positive and total number of droplets. The plot was generated with the `sim_ddpcr` function from the **dpcR** package.

In case of closed source software the analysis usually happens in a black box fashion tied to a specific platform. We agree that closed frameworks are not necessarily a bad thing, but should be avoided if possible (Rödiger et al., 2013a; Spiess et al., 2015). Studies by McCullough and Heiser (2008); Almiron et al. (2010); Durán et al. (2014) exemplified where the black box approach might fail in science. The same holds true for the open source software since software bugs are independent of a development model. However, at least open source gives the possibility to track and eliminate errors by an individual entity. Black-box systems often force the user to process the data by suboptimal analysis algorithms (Ruijter et al., 2013). Moreover, the data usually cannot be accessed between the steps of the analysis which restrains quality control. The visualization options are usually limited by the software vendor preferences and do not attain publication quality. In addition to this, users who have no access to the commercial software are barred. Aside from closed source software, data analysis is often performed in spreadsheets. However, this data processing approach is not advisable for research purposes. Most spreadsheets lack (or do not use) tools to validate the input, to debug implemented procedures and to automatize the workflow. These traits make them prone to errors and not well suited for complicated tasks (McCullough and Heiser, 2008; Burns, 2014).

For several reasons, R is one of the most popular tools in bioinformatics and is known as an early adopter of emerging technologies (Pabinger et al., 2014). R provides packages to build highly customized workflows, covering: data read-in, data preprocessing, analysis, post-processing, visualization and storage. As recently briefly reviewed in Pabinger et al. (2014), numerous R packages have been developed for the analysis of qPCR, dPCR, qIA and melting curve analysis experiments, including: **kulife**, **MCMC.qpcr**, **qPCR.CT**, **DivMelt**, **qpcR**, **dpcR**, **chipPCR**, **MBmca**, **RDML**, **nondetects**, **qpcrNorm**, **HTqPCR**, **SLqPCR**, **ddCt**, **EasyqpcR**, **unifiedWMWqPCR**, **ReadqPCR**, **NormqPCR**. All the packages are either available from CRAN or Bioconductor (Gentleman et al., 2004). The packages can be freely combined in a plugin-like architecture.

R is an open, operating system-independent platform for a broad spectrum of calculation options. Particularly, the visualization of experiments is one of R's pinnacles. R enables the users to create an efficient manipulation, restructuring and reshaping of data to make them readily-available for further processing. This is of the particular importance to the human–machine interface (Oh, 2014). Intrinsic properties of R such as the naming convention (Bååth, 2012) and class systems (e.g., **S3**, **S4**, reference classes and **R6**) vary considerable, depending on the package developer preferences. However, due to the open source approach, there is the common ground to track numerical errors. R offers various methods for a standardized data import/export and exchange. Workflows can embed

structured models Zeller et al. (2009), open data exchange formats (e.g., RDML), binary formats (Michna and Woods, 2013) or tools provided by the R workspace (R Development Core Team, 2012). The NetCDF binary format, available from the **RNetCDF** package, has advantages over some other binary formats (e.g., the RData format), since arbitrary array data sections of massive datasets can be processed efficiently (Michna and Woods, 2013). This might be useful for large data sets as present in high-throughput PCRs or dPCRs with large partition numbers. The R environment offers several datasets, which can be used for testing of algorithms. Therefore, others and we argue that R is suitable for reproducible research (Murrell, 2012; Gandrud, 2013; Hofmann et al., 2013; Kuhn, 2014; Leeper, 2014; Liu and Pounds, 2014).

The aim of this paper is to show case studies for qPCR, dPCR, qIA and melting curve analysis experiments. Our workflow effectively follows the principle illustrated in Figure 2. We intend to aggregate functionalities dispersed between various packages and offer a fast insight in the analysis of nucleic acid experiments with R. In particular, we describe how to:

- read-in data from a standardized file format,
- pre-process the amplification curve data,
- calculate specific parameters from the amplification curve data,
- calculate the melting temperature,
- and report the data.

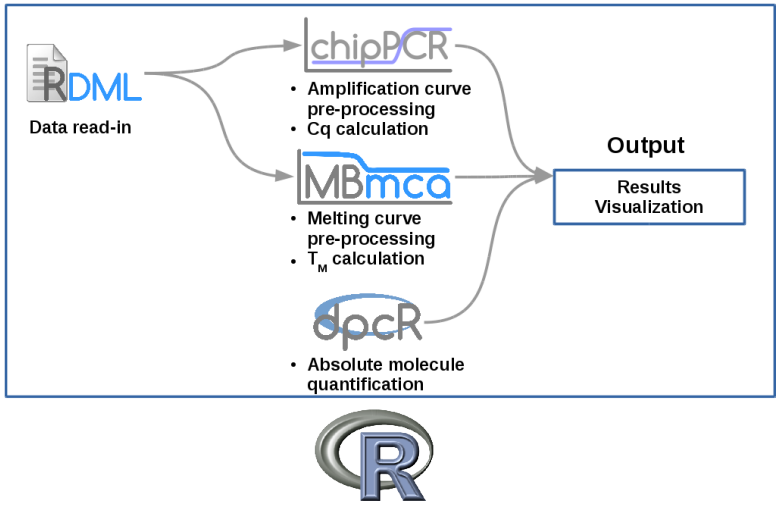

Figure 2: Exemplary workflow for quantitative PCR, digital PCR, quantitative isothermal amplification and melting curve analysis experiments in R. Core functionality is provided by the R software environment for statistical computing and graphics. In our scenario we used the **RDML** package to read-in data in standardized format. However, any format supported by R can be used. Further, processing of amplification curve data was performed with the **chipPCR** package and melting curve data were analysed with the **MBmca** package. The **dpcR** package can be embedded in the analysis of digital PCR experiments. Cq, quantification cycle; T_M, melting temperature.

Setting-up a working environment

We recommend performing the scripting in a dedicated integrated development environment (IDE) and graphical user interface (GUI) such as **RKWard** (Rödiger et al., 2012), **RStudio** (RStudio Team, 2012; Gandrud, 2013) or other technologies (Valero-Mora and Ledesma, 2012). Benefits of IDE's with GUI include syntax-highlighting, auto completion and function references for rapid prototyping of workflows. Typically, the analysis will start with data from a commercial platform. Most platforms have an option to export a CSV file or spreadsheets application file (e.g., *.xls, *.odt). The details for the data import have been described elsewhere (R Development Core Team, 2012; Rödiger et al., 2012). To keep the case study sections compact we have chosen to load datasets from the **qpcR** package (Ritz and Spiess, 2008) (v. 1.4.0) and the **RDML** package (v. 0.8-3) to our workspace. The **chipPCR** package (Rödiger et al., 2015) (v. 0.0.8-8) was used for data preprocessing, quality control and the calculation of the quantification cycle (Cq).

The Cq is a quantitative measure, which represents the number of cycles needed to reach a user defined threshold signal level, in the exponential phase of a qPCR/qIA reaction. Several Cq methods

have been described (Ruijter et al., 2013). In this study we have chosen the second derivative maximum method (Cq_{SDM}) and the "Cycle threshold" (Cq_{Ct}) method.

During a perfect qPCR reaction, the target DNA doubles (2^n; n = cycle number) at each cycle. Here the amplification efficiency (AE) is 100 %. However, in reality, numerous factors cause an inhibition of the amplification (AE < 100 %). The AE can be determined by the relation of the Cq value depending on the sample input quantity as described in (Rödiger et al., 2015; Svec et al.).

In Rödiger et al. (2013a) we described the application of R for the analysis of melting curve experiments from microbead-based assays. Since the mathematical foundation for melting curve analysis (MCA) is identical between all methods we used functions from the **MBmca** package (0.0.3-4) for an analysis of the target specific melting temperature (T_M) in experiments of the present study.

We completed our examples with case studies for the analysis of dPCR experiments. In particular, we used the **dpcR** package (0.1.4.0) to estimate the number of molecules in a sample.

Results

In the following sections we will show how R can be used (I) as a unified open software for data analysis and presentation in research, (II) as software frame-work for novel technical developments, (III) as platform for teaching of new technologies and (IV) as reference for statistical methods.

Case study one – qPCR and Amplification Efficiency Calculation

The goal of our fist case study was to calculate the Cq values and the AE from a qPCR experiment. We used the `guescini1` dataset from the **qpcR** package, where the gene *NADH dehydrogenase 1* (MT-ND1) was amplified in a LightCycler® 480 (Roche) thermocycler. Details of the experiment are described in Guescini et al. (2008). First we started with loading the required packages and datasets. A good practice for reproducible research is to track the package versions and environment used during the analysis. The function `sessionInfo` from the **utils** package provides this functionality. Assuming that the analysis starts with a clean R session it is possible to assign the required packages to an object, as shown below. The reproducibility of research can be further improved by the **archivist** package, which stores and recovers crucial data and preserves metadata of saved objects (not shown). All settings of R session can be easily saved and/or restored using **settings** package.

```
# Load the required packages for the data import and analysis.
# Load the chipPCR package for the pre-processing and curve data quality
# analysis and load the qpcR package as data resource.
require(chipPCR)
require(qpcR)

# Collect information about the R session used for the analysis of the
# experiment.
current.session <- sessionInfo()

# Next load the 'guescini1' dataset from the qpcR package to the
# workspace and assign it to the object 'gue'.
gue <- guescini1

# Define the dilution of the sample DNA quantity for the calibration curve
# and assign it to the object 'dil'.
dil <- 10^(2: -4)
```

We previewed the amplification curve raw data using the `matplot` function (see code below). The amplification curve data showed a strong signal level variation in the plateau region (Figure 3A). Therefore, all data were subjected to a minimum-maximum normalization (see Rödiger et al. (2013a)) using the `CPP` function from the **chipPCR** package. In addition, all data were baselined and smoothed (Figure 3B). The Cq values were calculated by the Cq_{SDM} and Cq_{Ct} methods as shown next.

```
# Pre-process the amplification curve data with the CPP function from the
# chipPCR package. The trans parameter was set TRUE to perform a baselining and
# the method.norm parameter was set to minm for a min-maximum normalization. All
# amplification curves were smoothed by Savitzky-Golay smoothing.

res.CPP <- cbind(gue[, 1], apply(gue[, -1], 2, function(x) {
  CPP(gue[, 1], x, trans = TRUE, method.norm = "minm", method.reg = "least",
```

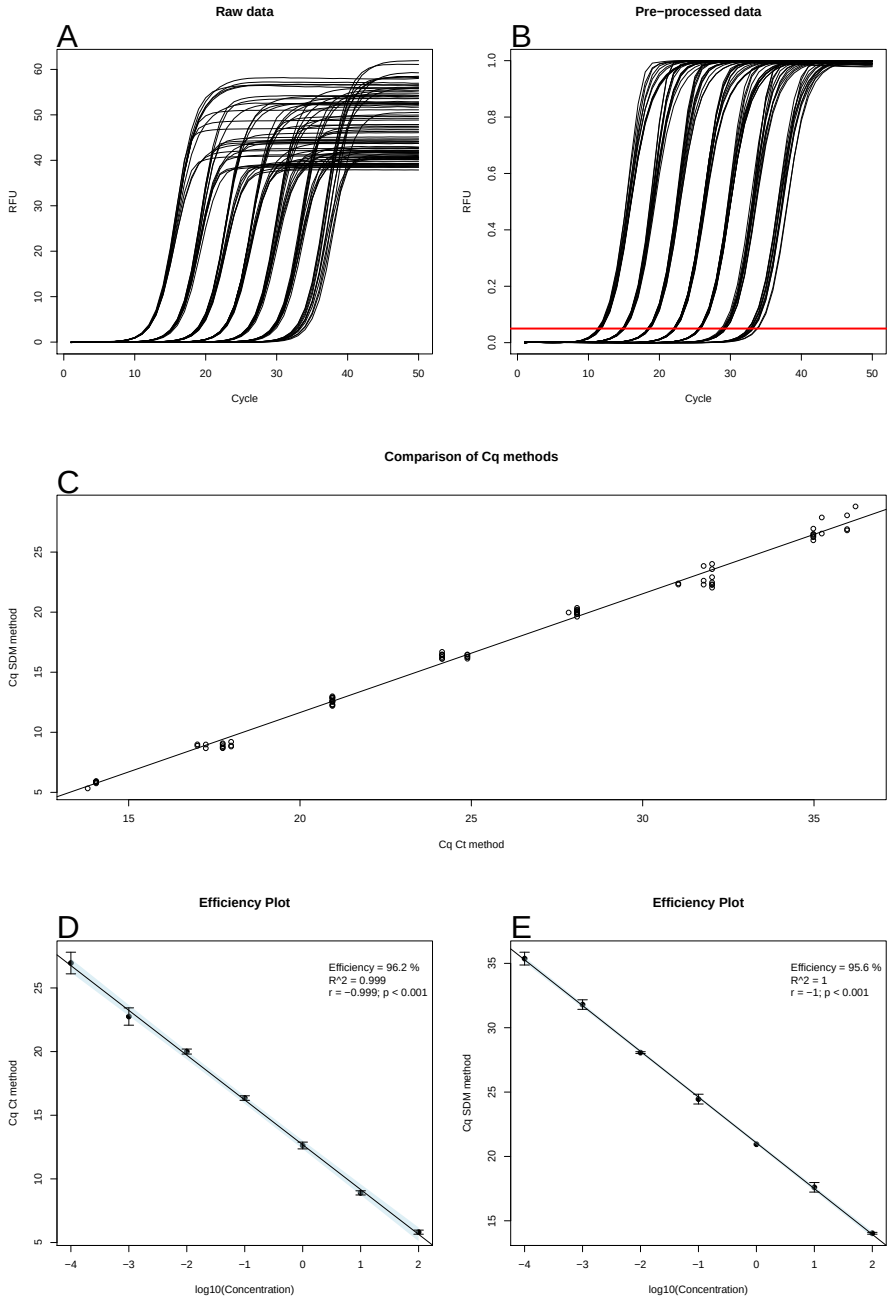

Figure 3: Analysis of the amplification curve data of the `guescini1` dataset. **(A)** Raw data from the calibration curve samples were visually inspected. The qPCR curves display a broad variation in plateau fluorescence (38 – 62 RFU). The red horizontal line (—) indicates the fluorescence level (0.05) used for the calculation of the Cq value by the "cycle threshold" method. **(B)** The `CPP` function from the **chipPCR** was used to baseline the data, to smooth the data with Savitzky-Golay smoothing filter and to normalize the data between 0 and 1. **(C)** The Cq values were calculated by the Cq_{SDM} method ("SDM method") (`inder`, **chipPCR**) and the Cq_{Ct} method ("Ct method") (`th.cyc`, **chipPCR**). The threshold value was set to $r = 0.05$. The Cq_{SDM} and Cq_{Ct} values were plotted and analysed by a linear regression ($R^2 = 0.9945$; $P < 2.2^{-16}$) and Pearson's ($r = 0.9972605$; $P < 2.2^{-16}$). The amplification efficiency based on **(D)** Cq_{Ct} values and **(E)** Cq_{SDM} values were automatically analysed with the `effcalc` (**chipPCR**) function. Cq, Quantification cycle; SDM, Second derivative maximum, R^2, Coefficient of determination; r, Pearson product-moment correlation coefficient, RFU, relative fluorescence units.

```
      bg.range = c(1,7))[["y.norm"]]
}))

# Use the th.cyc function from the chipPCR package to calculate the Cq values
# by the cycle threshold method at a threshold signal level "r" of 0.05.
Cq.Ct <- apply(gue[, -1], 2, function(x)
  th.cyc(res.CPP[, 1], x, r = 0.05)[1])

# Use the inder function from the chipPCR package to calculate the Cq values
# by the SDM method.
Cq.SDM <- apply(gue[, -1], 2, function(x)
  summary(inder(res.CPP[, 1], x), print = FALSE)[2])

# Fit a linear model to carry out a regression analysis.
res.Cq <- lm(Cq.Ct ~ Cq.SDM)
```

To compare the Cq_{SDM} and Cq_{Ct} methods we performed a regression analysis. The Cq methods are in a good agreement. However, the dispersion of the Cq_{Ct} values appeared to be higher than in the Cq_{SDM} values (Figure 3C).

```
> summary(res.Cq)

Call:
lm(formula = Cq.Ct ~ Cq.SDM)

Residuals:
    Min      1Q  Median      3Q     Max
-1.4904 -0.2730  0.0601  0.3540  1.1871

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -8.125534   0.207419  -39.17   <2e-16 ***
Cq.SDM       0.988504   0.008097  122.08   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5281 on 82 degrees of freedom
Multiple R-squared:  0.9945,  Adjusted R-squared:  0.9945
F-statistic: 1.49e+04 on 1 and 82 DF,  p-value: < 2.2e-16
```

The dilution ("dil") and the Cq ("Cq.Ct") values served as input for the calculation of the amplification efficiency (AE) with `effcalc` function from the **chipPCR** package. In our case study we needed to rearrange the "Cq.Ct" values in a matrix using the command $effcalc(dil, t(matrix(Cq.Ct, nrow = 12, ncol = 7))$. For visualization of the confidence intervals of the regression analysis we set the parameter $CI = TRUE$.

```
# Arrange and plot the results in a convenient way.
layout(matrix(c(1,2,3,3,4,5), 3, 2, byrow = TRUE))

# Store used margin parameters
def.mar <- par("mar")
layout(matrix(c(1,2,3,3,4,5), 3, 2, byrow = TRUE))
# Set bigger top margin.
par(mar = c(5.1, 4.1, 6.1, 2.1))

# Plot the raw amplification curve data.
matplot(gue[, -1], type = "l", lty = 1, col = 1, xlab = "Cycle",
        ylab = "RFU", main = "Raw data")
mtext("A", side = 3, adj = 0, cex = 2)

# Plot the pre-processed amplification curve data.
matplot(res.CPP[, -1], type = "l", lty = 1, col = 1, xlab = "Cycle",
        ylab = "RFU", main = "Pre-processed data")
mtext("B", side = 3, adj = 0, cex = 2)
abline(h = 0.05, col = "red", lwd = 2)
```

```
# Plot Cq.SDM against Cq.Ct and add the trendline from the linear regression
# analysis.

plot(Cq.SDM, Cq.Ct, xlab = "Cq Ct method", ylab = "Cq SDM method",
     main = "Comparison of Cq methods")
abline(res.Cq)
mtext("C", side = 3, adj = 0, cex = 2)

# Use the effcalc function from the chipPCR package to calculate the
# amplification efficiency.
plot(effcalc(dil, t(matrix(Cq.Ct, nrow = 12, ncol = 7))), ylab = "Cq Ct method",
     CI = TRUE)
mtext("D", side = 3, adj = 0, cex = 2)

plot(effcalc(dil, t(matrix(Cq.SDM, nrow = 12, ncol = 7))), ylab = "Cq SDM method",
     CI = TRUE)
mtext("E", side = 3, adj = 0, cex = 2)

# Resore margin default values.
par(mar = c(5.1, 4.1, 4.1, 2.1))
```

Finally, Cq values were plotted using the `layout` function (Figure 3D and E). The Cq values and amplification vary slightly between both methods. This is an expected observation and is in accordance to the findings by Ruijter et al. (2013). As shown in this case study, it is easy to set-up a streamlined workflow for data read-in, pre-processing and analysis with a few functions.

Case study two – qPCR and Melting Curve Analysis

A common task during the analysis of qPCR experiments is to distinguish between positive and negative samples (see Figure 5). If the melting temperature of a sample is known it is possible to automatize the decision by a melting curve analysis (MCA). As shown in Rödiger et al. (2013a) this can be done by interrogating the T_M. Therefore, we used a logical statement, which tests if T_M is within a tight temperature range. We used the signal height as second parameter. In line with "Case study one" we used the function `sessionInfo` to track all packages used during the analysis. Reproducible research is greatly enhanced if open data exchange formats are used. Therefore, we used the **RDML** package for data read-in. The amplification and melting curve data were measured with a CFX96 system (Bio-Rad) and then exported as RDML v 1.1 format file as 'BioRad_qPCR_melt.rdml'. Within this qPCR experiment we amplified the *Mycobacterium tuberculosis katG* gene and tried to detect a mutation at codon 315. The experiment was separated in two parts:

1. Detection of overall *M. tuberculosis* DNA (wild-type and mutant) and
2. specific detection of wild-type *M. tuberculosis* by melting of TaqMan probe (quencher – BHQ2, fluorescent reporter – Cy5) with amplified DNA (see Luo et al. (2011) for probe/primer sequences and further details).

The qPCR was conducted using EvaGreen® Master Mix (Syntol) according to the manufacturer's instructions, with 500 nM of primers and probe in a 25 μL final reaction volume. Thermocycling was conducted using a CFX96 (BioRad) initiated by 3 min incubation at 95 °C, followed by 41 cycles (15 s at 95 °C; 40 s at 65 °C) with a single read-out taken at the end of each cycle. Probe melting was conducted between 35 °C and 95 °C by 1 °C at 1 s steps.

The structure of an RDML file is quite complex. Though RDML files are XML structured files and thus intended to be readable by humans, it is hard to grasp the complex hierarchical file structure with out some basic understanding. A simple and fast method to compactly display the structure of object in R is to use the `str` or `summary` function (not shown). However, such R tools are not informative in this context. Therefore we implemented a dendrogram like view (Figure 4). According to this, the file contains different datasets, each with 3 samples, ('pos', 'ntc', 'unknown'). Only a subset of the data was used in our case study and combined to the object "qPCR".

```
# Import the qPCR and melting curve data via the RDML package.
# Load the chipPCR package for the pre-processing and curve data quality
# analysis and the MBmca package for the melting curve analysis.
require(RDML)
require(chipPCR)
```

```
require(MBmca)
require(dplyr)

# Collect information about the R session used for the analysis of the qPCR
# experiment.
current.session <- sessionInfo()

# Load the BioRad_qPCR_melt.rdml file form RDML package and assign the data to the
# object BioRad.
filename <- paste(path.package("RDML"), "/extdata/", "BioRad_qPCR_melt.rdml", sep = "")
BioRad <- RDML$new(filename)

# Fetch cycle dependent fluorescence for the EvaGreen channel and row 'D'
# (that contains target 'Cy5-2' at channel 'Cy5') of the
# katG gene and aggregate the data in the object qPCR.

qPCR <- BioRad$AsTable() %>%
  filter(target == "EvaGreen",
         grepl("^D", position))  %>%
  BioRad$GetFData(.)
```

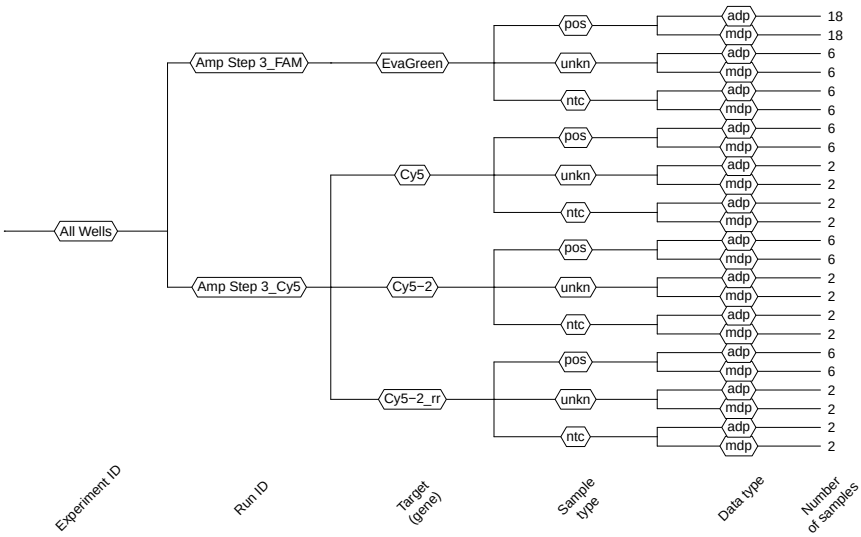

Figure 4: File structure visualization of the RDML file 'BioRad_qPCR_melt.rdml' form the **RDML** package. The file was read by the `RDML` function and the structure displayed as dendrogram by the call `BioRad$AsDendrogram()`. The names are used according to the RDML convention by Lefever et al. (2009). The object `BioRad` branches into an experiment with `Run ID` names for two fluorescence detection channels (FAM, Cy5). The targets have typical designations like `pos` (positive), `unkn` (unknown) and `ntc` (non-template control). In the deeper branches are the data types `adp` (amplification data point) and `mdp` (melting data point) shown with the number of samples (ranging from 2 to 18).

We inspected and pre-processed a subset of the amplification curve data solely using functionalities provided by the **chipPCR** package. The `plotCurves` function was used to get an overview of the curvatures. The data indicated a baseline shift in all curves with a slight negative trend (Figure 5). This observation suggested to baseline the raw data by using a linear regression model (cycles x - y; ($bg.range = c(x, y)$) in the `CPP` function.). The curvatures of "D1_Alm12..." and "D2_Alm12..." exhibited a drop in the plateau phase. However, this is not of importance during this stage of the analysis.

```
# Use plotCurves function to get an overview of the amplification curve samples.

plotCurves(qPCR[, 1], qPCR[, -1], type = "l")
```

```
# Detect positive samples - calculate Cq values by the cycle threshold method.
# The threshold signal level r was set to 10.
Cq.Positive <- t(apply(qPCR[, -1], 2, function(x)
{
  res <- CPP(qPCR[, 1], x, trans = TRUE, bg.range = c(2, 8),
             method.reg = "least")[["y.norm"]]
  # The th.cyc fails when the threshold exceeds maximum
  # observed fluorescence values, so it must be used with try()
  th.cycle <- try(th.cyc(qPCR[, 1], res, r = 10)[1], silent = TRUE)
  cq <- ifelse(class(th.cycle) != "try-error", as.numeric(th.cycle), NA)
  pos <- !is.na(cq)
  c(Cq=cq, M.Tub_positive = pos)
}
))
Cq.Positive
```

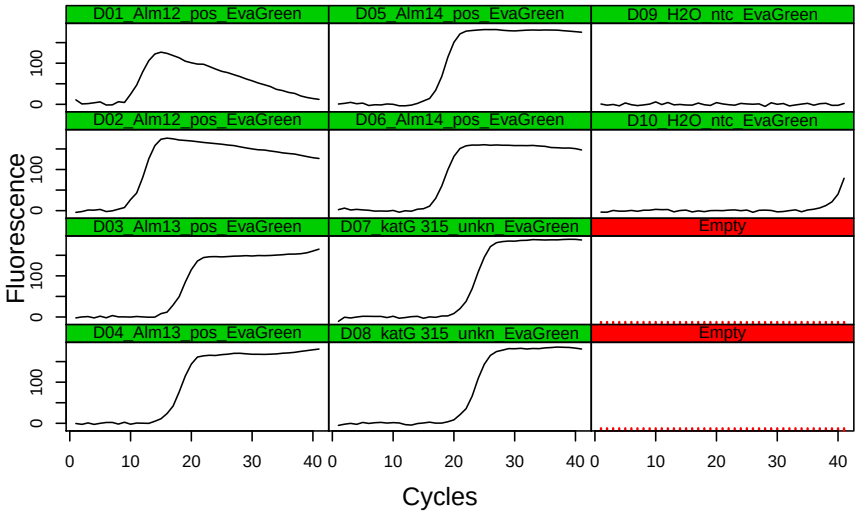

Figure 5: Analysis of the amplification curve data. The calibration curve samples were inspected by the `plotCurves` function from the **chipPCR** package. The green color code indicates that the data contain no missing values. However, the visual inspection revealed that the data are noisy. All samples ("D1_Alm12…" - "D8_Alm12…") appeared to be positive. One negative control ("D10_H20_ntc_EvaGreen") seems to be contaminated.

Since the amplification curves indicated that selected samples (except non-template-control ("NTC")) are positive, we distinguished between true positive and true negative samples by MCA (Figure 6A).

```
# Fetch temperature dependent fluorescence for the Cy5 channel of the
# probe that can hybridize with Mycobacterium tuberculosis katG gene (codon 315)
# and aggregate the data in the object 'melt'.
melt <- BioRad$AsTable() %>%
        filter(target == "Cy5-2")  %>%
        BioRad$GetFData(., data.type = "mdp")

# Calculate the melting temperature with the diffQ function from the MBmca
# package. Use simple logical conditions to find out if a positive sample with
# the expected Tm of circa 54.5 degree Celsius is found. The result of the test
# is assigned to the object 'positive'.
Tm.Positive <- matrix(nrow = ncol(melt) - 1,
                      byrow = TRUE,
                      dimnames = list(colnames(melt)[-1]),
                      unlist(apply(melt[, -1], 2, function(x) {
                        res.Tm <- diffQ(cbind(melt[, 1], x),
```

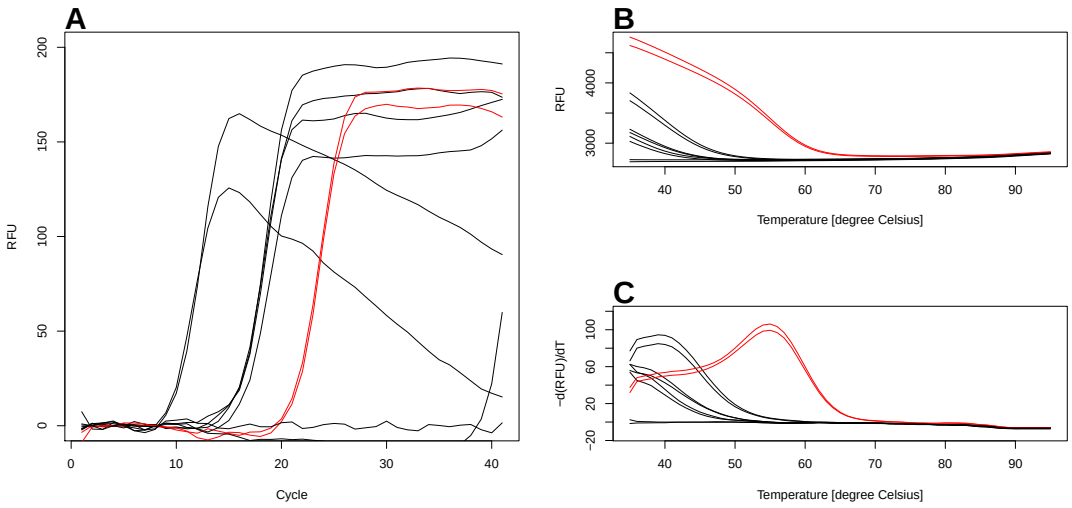

Figure 6: Amplification curve plots and melting curve plots. **(A)** The raw amplification curve data were pre-processed with the `CPP` function prior to visualization. To calculate the T_M values from the raw melting curve data **(B)** we used the `diffQ` function from the **MBmca** package. **(C)** We adjusted our algorithm to plot the true positive melting peaks in red, while negative melting peaks were labelled in black. The inspection of the plot and the output of `results.tab` showed that only D07_katG 315_unkn_EvaGreen (—) and D08_katG 315_unkn_EvaGreen (—) are true positive. RFU, relative fluorescence units; -d(RFU)/dT, negative first derivative of the melting-curve.

```
                              fct = max, inder = TRUE)
    positive <- ifelse(res.Tm[1] > 54 &
                              res.Tm[1] < 55 &
                              res.Tm[2] > 80, 1, 0)
    c(res.Tm[1], res.Tm[2], positive)
  }))))


# Present the results in a tabular output as data.frame 'results.tab'.
# Result of analysis logic is:
# Cq.Positive && Tm.Positive = Wild-type
# Cq.Positive && !Tm.Positive = Mutant
# !Cq.Positive && !Tm.Positive = NTC
# !Cq.Positive && Tm.Positive = Error
results <- sapply(1:length(Cq.Positive[, 1]), function(i) {
  if(Cq.Positive[i, 2] == 1 && Tm.Positive[i, 3] == 1)
    return("Wild-type")
  if(Cq.Positive[i, 2] == 1 && Tm.Positive[i, 3] == 0)
    return("Mutant")
  if(Cq.Positive[i, 2] == 0 && Tm.Positive[i, 3] == 0)
    return("NTC")
  if(Cq.Positive[i, 2] == 0 && Tm.Positive[i, 3] == 1)
    return("Error")
})

results.tab <- data.frame(cbind(Cq.Positive, Tm.Positive, results))
names(results.tab) <- c("Cq", "M.Tub DNA", "Tm", "Height",
                        "Tm positive", "Result")

results.tab[["M.Tub DNA"]] <- factor(results.tab[["M.Tub DNA"]],
                                     labels=c("Not Detected", "Detected"))

results.tab[["Tm positive"]] <- factor(results.tab[["Tm positive"]],
                                       labels=c(TRUE, FALSE))
results.tab
```

The results of the analysis can be invoked by the statement *results.tab* (not shown). Finally, we plotted and printed the output of our melting curve (Figure 6B) and melting peak (Figure 6C) analysis.

```
# Convert the decision from the results.tab object in a color code:
# Negative, black; Positive, red.

color <- c(Tm.Positive[, 3] + 1)

# Arrange the results of the calculations in plot.
layout(matrix(c(1,2,1,3), 2, 2, byrow = TRUE))

# Use the CPP function to preporcess the amplification curve data.
plot(NA, NA, xlim = c(1, 41), ylim = c(0,200), xlab = "Cycle", ylab = "RFU")
mtext("A", cex = 2, side = 3, adj = 0, font = 2)
lapply(2L:ncol(qPCR), function(i)
  lines(qPCR[, 1],
        CPP(qPCR[, 1], qPCR[, i], trans = TRUE,
            bg.range = c(1,9))[["y.norm"]],
        col = color[i - 1]))

matplot(melt[, 1], melt[, -1], type = "l", col = color,
        lty = 1, xlab = "Temperature [degree Celsius]", ylab = "RFU")
mtext("B", cex = 2, side = 3, adj = 0, font = 2)

plot(NA, NA, xlim = c(35, 95), ylim = c(-15, 120), xlab = "Temperature [degree Celsius]",
     ylab = "-d(RFU)/dT")
mtext("C", cex = 2, side = 3, adj = 0, font = 2)

lapply(2L:ncol(melt), function(i)
  lines(diffQ(cbind(melt[, 1], melt[, i]), verbose = TRUE,
              fct = max, inder = TRUE)[["xy"]], col = color[i - 1]))
```

According to the analysis by MCA the samples "D07_katG315…" and "D08_katG315…" were the only positive samples. The remaining samples appeared to be positive in the amplification plot in Figure 5 due to primer-dimer formation or sample contamination.

Case study three – Isothermal Amplification

Isothermal amplification is an alternative nucleic acid amplification method, which uses a constant temperature rather than cycling through denaturation, annealing and extension steps (Rödiger et al., 2014). The signal is monitored continuously on a time basis. Often the abscissa values are not uniformly spaced in qIA. Our `CPP` function gives a warning in such a case. In qIA the Cq values are dependent on the time instead of cycles. In this study we used the `th.cyc` function from the **chipPCR** package to determine the time (Cq_t) required to reach a defined threshold signal level.

We performed a quantitative isothermal amplification (qIA) with the plasmid *pCNG1* by using a Helicase Dependent Amplification (HDA). Our previously reported VideoScan platform (Rödiger et al., 2013b) was used to control the temperature and to monitor the amplification reaction. The VideoScan technology is based on a highly versatile fluorescence microscope imaging platform, which can be operated with a heating/cooling unit (HCU) for qPCR and MCA applications (Rödiger et al., 2013a,b). Since the enzyme DNA Helicase unwinds DNA, no thermal denaturation is needed. The HDA conditions were taken from the "IsoAmp III Universal tHDA Kit", Biohelix Corp, as described by the vendor. In detail, the reaction was composed of "mix A" 10 μL A. bidest, 1.25 μL 10X buffer, 0.75 μL primer (150 nM final), 0.5 μL template plasmid. Preincubation: The mixture was incubated for 2 min at 95°C and immediately placed on ice. Reaction "mix B" contained 5 μL A. bidest., 1.25 μL 10X buffer, 2 μL NaCl, 1.25 μL $MgSO_4$, 1.75 μL dNTPs, 0.25 μL EvaGreen (Biotium), 1 μL enzyme mix. The mix was covered with 50 μL mineral oil (Roth). The fluorescence measurement in VideoScan HCU started directly after adding "mix B" at 65°C. A $1x$ (D1) and a 1 : 10 dilution (D2) were tested. The resulting dataset `C81` is part of the **chipPCR** package. Two concentrations (stock and 1:10 diluted stock) of input DNA were used in the HDA. Similar to the previous case studies, we first prepared the plot of the data. Since the raw data showed a slight negative trend and an off-set of circa 0.45 MFI (mean fluorescence intensity) it was necessary to pre-process the raw data (Figure 7A).

First we had a look at the `C81` dataset with the `str` function.

```
str(C81)
```

```
'data.frame':        351 obs. of  5 variables:
 $ Cycle : int  0 1 2 3 4 5 6 7 8 9 ...
 $ t.D1  : int  0 51 73 90 107 124 140 157 174 190 ...
 $ MFI.D1: num  0.549 0.535 0.532 0.53 0.525 ...
 $ t.D2  : int  0 19 53 72 91 110 128 147 166 185 ...
 $ MFI.D2: num  0.77 0.523 0.514 0.51 0.508 ...
```

C81 is a data frame. The first column contains the measure points (Cycle) and the consecutive columns contain the time stamps (e.g., "$ t.D1", in seconds according to the **chipPCR** manual) and the signal hight ("$ MFI.D1", as mean fluorescence intensity (MFI)). A brief look at the time stamps showed that the data are not uniformly spaced. The warning message by the CPP functions is just a reminder for to user to take care during the pre-processing. Methods like smoothing might cause artifacts (Spiess et al., 2015). Since we know that the HDA is a time-dependent reaction we do not have a use for the discrete measure points. Next we plotted and analysed the data.

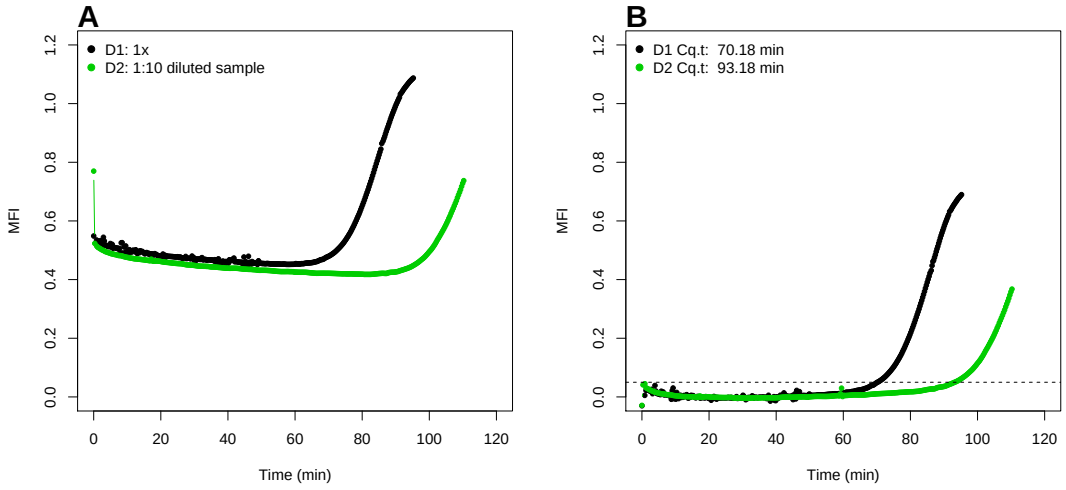

Figure 7: Quantitative isothermal amplification by Helicase Dependent Amplification (HDA). **(A)** The raw data of the HDA (D1, undiluted, D2 1 : 10 diluted) exhibit some outliers (detector artifacts), an off-set of circa 0.5 MFI and a slight negative trend in the baseline region (0 - 52 minutes). **(B)** First we used the CPP function to smooth the data with a spline function. Baselining was done with a linear regression model (robust MM-estimator). Finally, we used the th.cyc function (**chipPCR**) to calculate the cycle threshold time for samples D1 and D2. The threshold value was set to $r = 0.05$ ($--$, threshold line). Cq_t, required time to reach a defined threshold signal level. MFI, mean fluorescence intensity.

```
# Drawn in an 2-by-1 array on the device by two columns and one row.
par(mfrow = c(2, 1))

# Plot the raw data from the C81 dataset to the first array and add
# a legend. Note: The abscissa values (time in seconds) was divided
# by 60 (C81[, i] / 60) to convert to minutes.
plot(NA, NA, xlim = c(0, 120), ylim = c(0, 1.2), xlab = "Time (min)", ylab = "MFI")
mtext("A", cex = 2, side = 3, adj = 0, font = 2)
lapply(c(2, 4), function(i) {
  lines(C81[, i] / 60, C81[, i + 1], type = "b", pch = 20, col = i - 1)
})
legend(10, 0.8, c("D1: 1x", "D2: 1:10 diluted sample"), pch = 19, col = c(1,3), bty = "n")

# Prepare a plot on the second array for the pre-processed data.
plot(NA, NA, xlim = c(0, 120), ylim = c(0, 1.2), xlab = "Time (min)", ylab = "MFI")
mtext("B", cex = 2, side = 3, adj = 0, font = 2)
```

First, we used the CPP function to pre-process the raw data. Similar to the other case studies we baselined and smoothed the amplification curve data prior to the analysis of the of the Cq_t value. However, instead of the Savitzky-Golay smoother we used a cubic spline ($method = "spline"$) in the

CPP function. In addition, outliers were automatically removed in the baseline region (Figure 7A and B). The background range was defined by bare eye to be between the 1^{th} and 190^{th} data point (corresponds to baseline region between 0 and 52 minutes).

```
# Apply the CPP functions to pro-process the raw data.1) Baseline data to zero,
# 2) Smooth data with a spline, 3) Remove outliers in background range between
# entry 1 and 190. Assign the results of the analysis to the object 'res'.

res <- lapply(c(2, 4), function(i) {
  y.s <- CPP(C81[, i] / 60, C81[, i + 1],
             trans = TRUE,
             method = "spline",
             bg.outliers = TRUE,
             bg.range = c(1, 190))
  lines(C81[, i] / 60, y.s[["y.norm"]], type = "b", pch = 20, col = i - 1)
  # Use the th.cyc function to calculate the cycle threshold time (Cq.t).
  # The threshold signal level r was set to 0.05. NOTE: The function th.cyc
  # will give a warning in case data are not equidistant. This is intentional
  # to make the user aware of potential artificats due to pre-processing.
  paste(round(th.cyc(C81[, i] / 60, y.s[["y.norm"]], r = 0.05)[1], 2), "min")
})

# Add the cycle threshold time from the object 'res' to the plot.

abline(h = 0.05, lty = 2)
text(10, 0.55, "Cq.t:")
legend(10, 0.5, paste(c("D1: ", "D2: "), res), pch = 19, col = c(1, 3),
       bty = "n")
```

The pre-processed data were subjected to the analysis of the Cq_t values. It is important to note that the trend correction and proper baseline was a requirement for a sound calculation. We calculated Cq_t values of 70.18 minutes and 93.18 minutes for the stock (D1) and 1:10 (D2) diluted samples, receptively.

Case study four - digital PCR

We have developed the **dpcR** package for analysis and presentation of digital PCR experiments. The **dpcR** package can be used to build custom-made analysis pipelines and provides structures to be openly extended by the scientific community. Simulations and predictions of binomial and Poisson distributions, commonly used theoretical models of dPCR, statistical data analysis methods, plotting facilities and report generation tools are part of the package (Pabinger et al., 2014). Here, we show a case study for the **dpcR** package. Simulations are part in many educational curricula and greatly support teaching. In this case study, we mimicked an *in silico* experiment for a droplet digital PCR similar to Figure 1. The aim was to assess the concentration of the template sample. In the following we will use the expression partition as synonym for droplet. The number of positive partitions (k), total number of partitions (n) and the size of the partition are the only data required for the analysis. The estimate of the mean number of template molecules per partition ($\hat{\lambda}$) was calculated using following equation (Huggett et al., 2013):

$$\hat{\lambda} = -\ln(1 - \frac{k}{n}). \tag{1}$$

The average partition volume in our experiment was assumed to be 5 nL. We counted $n = 16800$ partitions in total from which $k = 4601$ partitions were positive. The binomial distribution of positive and negative partitions was used to determine $\hat{\lambda}$ (Figure 8). Our packages allows both easy estimation of a density of the parameter and calculation of confidence intervals using Wilson's method (Brown et al., 2001) at a confidence interval level of 0.95. The obtained mean number of template molecules per partition multiplied by the volume of the partitions ($16800 * 5\ nL$) constitutes the sample concentration.

```
# Load the dpcR package for the analysis of the digital PCR experiment.
require(dpcR)

# Analysis of a digital PCR experiment. The density estimation.
# In our in silico experiment we counted in total 16800 partitions (n).
# Thereof, 4601 were positive (k).
```

```
(dens <- dpcr_density(k = 4601, n = 16800, average = TRUE, methods = "wilson",
                      conf.level = 0.95))

# Let us assume, that every partition has roughly a volume of 5 nL.
# The total concentration (and its confidence intervals) in molecules/mL is
# (the factor 1e-6 is used for the conversion from nL to mL):
dens[4:6] / 5 * 1e-6
```

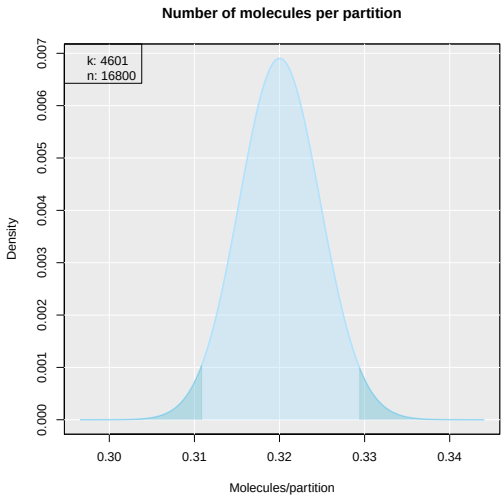

Figure 8: `dpcr_density` function from the **dpcR** package used for analysis of droplet digital PCR experiment. From 16800 counted partitions (n) 4601 were positive (k). The chart presents the distribution of mean number of template molecules per partition ($\hat{\lambda}$). n, number of total partitions; k, number of positive partitions.

Since we assumed a partition volume of 5 nL we have a total volume of 0.084 mL an 6.40^{-8} (95% CI: 6.2^{-8} - 6.6^{-8}) molecules/mL in the sample.

Selected functionality was implemented as interactive **shiny** GUI application to make the software accessible for users who are not fluent in R and for experts who wish to automatize routine tasks. Details and examples of the **shiny** web application framework for R can be found at `http://shiny.rstudio.com/`. We implemented flexible user interfaces, which run the analyses and graphical representation into interactive web applications either as service on a web server or on a local machine without knowledge of HTML or ECMAScript (see **dpcR** manual). The interface is designed in a cascade workflow approach (Data import → Analysis → Output → Export) with interactive users choice on input data, methods and parameters using typical GUI elements such as sliders, drop-downs and text fields. An example can be found at `https://michbur.shinyapps.io/dpcr_density/`. This approach enables the automatized output of R objects in combined plots, tables and summaries.

Case study five - digital PCR correction

There is an ongoing debate in the scientific community about the effect of the partition volume on the estimated copy numbers size (Huggett et al., 2014b; Corbisier et al., 2015; Majumdar et al., 2015). Corbisier et al. (2015) showed that the partition volume is a critical parameter for the measurement of copy number concentrations. Their study revealed that the average droplet volume defined in the QuantaSoft software (v. 1.3.2.0, BioRad QX100 Droplet Digital PCR System) is 8 % lower than the real volume. In consequence, results of quantifications were systematically biased between different dPCR platforms.

Case study four served as an introduction into the analysis of simulated dPCR experiments with the **dpcR** package. In the next case study, number five, we used the `pds_raw` dataset, which was generated by a BioRad QX100 Droplet Digital PCR System experiment. We re-analysed the data with the partition volume of 0.834 nL as proposed by Corbisier et al. (2015) and a volume of 0.90072 nL as used in the BioRad QX100 Droplet Digital PCR System.

Our experimental setup was as follows. A duplex assay was used to simultaneously detect a constant amount of genomic DNA (theoretically 10^2 copies/μL) and a variable amount of plasmid DNA (10-fold serial dilution, not shown). The genomic DNA was isolated from *Pseudomonas putida*

KT2440 and the plasmid was *pCOM10-StyA::EGFP StyB*. Template DNA was heat treated at 95°C for 5 min prior to PCR. Channel 1, primers for genomic DNA marker *ileS*, Taqman probes (FAM labelled). Channel 2, primers for plasmid DNA marker *styA*, Taqman probes (HEX labelled) (see (Jahn et al., 2013, 2014) and the manual of the **dpcR** package for details). Gating and partition clustering was taken without any modification from the data output of the BioRad QX100 Droplet Digital PCR System. Each partition is represented by a dot in Figure 9. First, we had a look at the data structure of the `pds_raw` dataset.

```
# Load the dpcR package and use the pds_raw dataset for the analysis of the
# digital PCR experiment.
require(dpcR)

# To get an overview of the dataset we used the head and summary R functions.
head(summary(pds_raw))

# The output shows that the dataset contains lists of different samples (A01 ...)
    Length Class        Mode
A01 "3"    "data.frame" "list"
A02 "3"    "data.frame" "list"
A03 "3"    "data.frame" "list"
A04 "3"    "data.frame" "list"
B01 "3"    "data.frame" "list"
B02 "3"    "data.frame" "list"

# Next we used str for the element A01. The element of the list contains a data frame
# with three columns. Two contains Amplitude values (fluorescence intensity) and one
# contains cluster resultes (interger values of 1 - 4).

str(pds_raw[["A01"]])
'data.frame':        11964 obs. of  3 variables:
 $ Assay1.Amplitude: num  397 399 402 416 417 ...
 $ Assay2.Amplitude: num  3732 3808 4007 3778 3685 ...
 $ Cluster         : int  4 4 4 4 4 4 4 4 4 4 ...
```

Since the structure of the dataset was known now we selected samples for the analysis. According to the **dpcR** manual contain the samples "A02", "B02", "C02" and "D02" the values for the replicates at a 1:100 dilution and "G04" the values for the non-template control.

```
# Select the wells for the analysis. A01 to D01 are four replicate dPCR reactions
# and G04 is the no template control (NTC).
wells <- c("A02", "B02", "C02", "D02", "G04")

# Set the arrangement for the plots. The first column contains the amplitude
# plots, column two the density functions and column three the concentration
# calculated on according to the droplet volume as defined in the QX100 system,
# or the method proposed by Corbisier et al. (2015).
par(mfrow = c(5,3))

# The function bioamp was used in a loop to extract the number of positive and negative
# partitions from the sample files. The results were assigned to the object 'res' and plotted.
# Horizontal and vertical lines show the threshold borders as defined by the QX100 system.

for (i in 1L:length(wells)) {
  cluster.info <- unique(pds_raw[wells[i]][[1]]["Cluster"])
  res <- bioamp(data = pds_raw[wells[i]][[1]], amp_x = 2, amp_y = 1,
                main = paste("Well", wells[i]), xlab = "Amplitude of ileS (FAM)",
                ylab = "Amplitude of styA (HEX)", xlim = c(500,4700),
                ylim = c(0,3300), pch = 19)

  legend("topright", as.character(cluster.info[, 1]), col = cluster.info[, 1], pch = 19)
```

Next we used the results from the object `res` to get the information about the number positive partitions for the plasmid DNA marker *styA*. This is to be found in the clusters 2 and 3.

```
  # Counts for the positive clusters 2 and 3 were assigned to new objects and further used by
  # the function dpcr_density to calculate the number of molecules per partition and the
```

```
  # confidence intervals. The results were plotted as density plot.
  k.tmp <- sum(res[1, "Cluster.2"], res[1, "Cluster.3"])
  # Counts for all clusters
  n.tmp <- sum(res[1, ])

  dens <- dpcr_density(k = k.tmp, n = n.tmp,
                       average = TRUE, methods = "wilson")
  legend("topright", paste("k:", k.tmp,"\nn:", n.tmp))

  # Finally, the concentration of the molecules was calculate with the volume used in
  # the QX100 system and as proposed by Corbisier et al. (2015).The results were added
  # as barplot with the confidence intervals.

  res.conc <- rbind(original = dens[4:6] /  0.90072 * 1e-6,
                    corrected = dens[4:6] / 0.834 * 1e-6)
  barplot(res.conc[, 1], col = c("white","grey"),
          names = c("Bio-Rad", "Corbisier"),
          main = "Influence of\nDroplet size", ylab = "molecules/ml",
          ylim = c(0,1.5*10E-8))
    arrows(c(0.7,1.9), res.conc[, 2], c(0.7,1.9), res.conc[, 3], angle = 90,
           code = 3, lwd = 2)
}
```

Our results show, that the replicates in Figure 9 A01 - D01 vary. However, the variation is within a similar range (Figure 9 column 2). The positive samples differ significantly from the negative control (Figure 9 G04). As proposed by Corbisier et al. (2015) the analysis with the non-corrected droplet volume leads to an underestimation of sample concentrations (Figure 9 column 3). However, our case studies shows, that the R environment can be used to circumvent problems of locked-in systems.

Discussion and Conclusion

This study gave a brief introduction to the analysis of qPCR, qIA, MCA and dPCR experiments with R. In addition to this, we briefly referenced to a vast collection of additional packages available from CRAN and Bioconductor. The packages may be considered as the building blocks (libraries) to create what users want and need. We showed that automatized research with R offers powerful means for statistical analysis and visualization. The software is not tied to a vendor or specific application (e.g., chamber or droplet based digital PCR, capillary or plate qPCR). It should be quite easy even for an inexperienced user to define a workflow and to set up an environment for specific needs in a broad range of technical settings (Figure 10). This environment enforces no monolithic integration. We claim that the modular structure allows the user to perform flexible data analysis, adjusted to their needs and to design frameworks for high-throughput analysis. Furthermore, R enables the user to access and reuse code for the creation of reports in various formats (e.g., HTML, PDF).

Most of the software is cross-platform open source software. Despite the fact that R is free of charge, it is quite possible to build commercial applications. The packages cover implementation of novel approaches and peer-reviewed analysis methods. R packages are an open environment to adopt to the growing knowledge in life sciences and medical sciences. Therefore, we argue that this environment may provide a structure for standardized nomenclature and serve as reference in qPCR and dPCR analysis. Speaking about openness, it needs to be emphasized that the main advantage of this software is its transparency at any time for anybody. Thus, it is possible to track numerical errors. A disadvantage of R is the lack of comprehensive GUIs for qPCR analysis. Graphical user interfaces (GUI) are key technologies to spread the use of R in bioanalytical sciences. Currently, we are establishing the "pcRuniveRsum" (http://michbur.github.io/pcRuniveRsum/) as an on-line resource for the interested users. The command-line structure makes R "inaccessible" for many novices. We try to solve this problem with easily accessible GUIs (Rödiger et al., 2012). However, the work on this additions has been recently started and is still in progress.

Acknowledgment

Part of this work was funded by the BMBF InnoProfile-Transfer-Projekt 03 IPT 611X, the European Regional Development Fund (ERDF/EFRE) on behalf of the European Union, the Sächsische Aufbaubank (Free State of Saxony, Germany) and the Russian Ministry of Education and Science (project No. RFMEFI62114X0003) and with usage of scientific equipment of Center for collective use "Biotech-

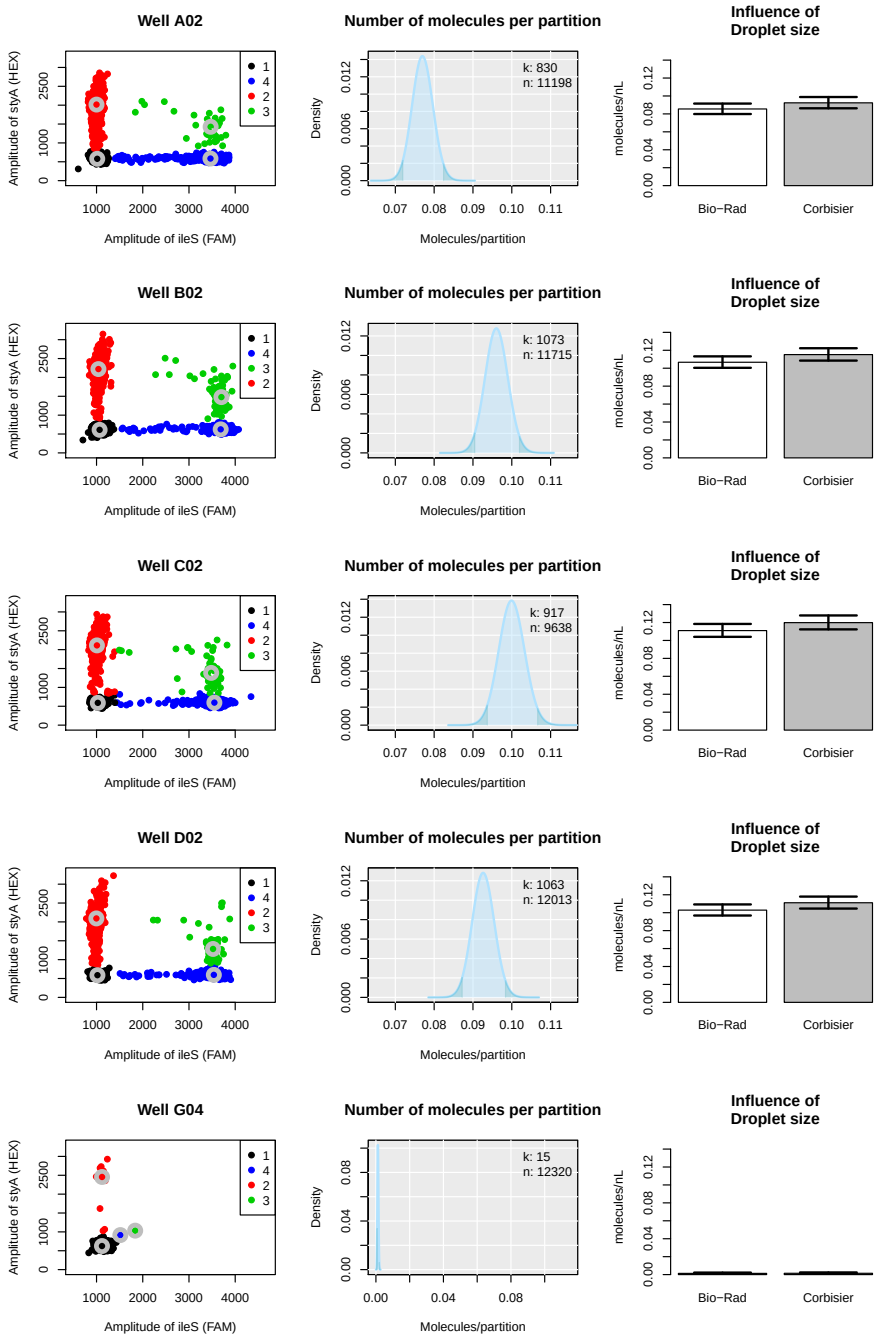

Figure 9: Analysis of a droplet digital PCR experiment. We used the `pds_raw` dataset from the **dpcR** package. All raw data were generated with a BioRad QX100 Droplet Digital PCR System. Column one: Amplitude plot of raw data shown by the `bioamp` function. Each dot represents a partition in a cluster. Negative partition are below and positive partitions above the dashed line (—). *Cluster* 1 •, *Cluster* 2 • = Target, *Cluster* 3 • = Target, *Cluster* 4 •. Well A01 - D04 are replicate measurements and Well G04 is the negative control; Column two: Density function with showing the number of molecules per partition. All replicates had similar numbers of counted positive partitions (k) and total number of partitions (n); Column three: Concentration of DNA molecules based on the volume used by the BioRad QX100 Droplet Digital PCR System and the volume proposed by Corbisier et al. (2015). n, number of total partitions; k, number of positive partitions; styA, styrene monooxygenase; ileS, isoleucyl-tRNA synthetase; FAM, Fluorescein channel; HEX, hexachlorofluorescein.

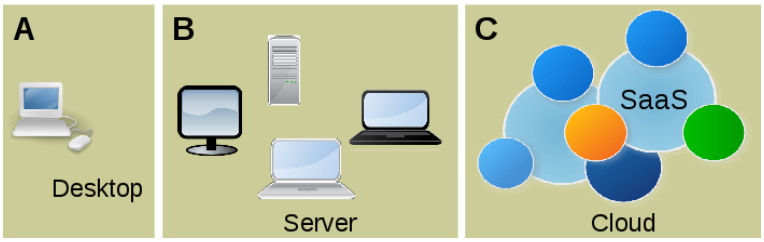

Figure 10: Deployment of R applications for the qPCR and dPCR experiments. **(A)** R is typically run from a desktop computer an operated by an GUI/IDE application such as **RStudio** or **RKWard**. This approach provides a flexible workflow for individuals. **(B)** Another approach is to run R with specific applications on a local server. Such scenarios are useful for the deployment within research departments or cooperate units (Nolan and Temple, 2014). **(C)** Cloud computing (CC) provides shared and scalable computing capacity (e.g., computing capacity, application software) and storage capacity (e.g., databases) as a service to an individual user or a community Service categories include: Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS) and Software-as-a-Service (SaaS) over a network. Providers of CC manage the infrastructure and resources to achieve coherence and economies of scale similar to a utility over a network (i. p., Internet) (Ohri, 2014).

nology" at All-Russia Research Institute of Agricultural Biotechnology. We would like to thank the R community. We would like to thank Mario Menschikowski (Technical University Dresden) for the droplet digital PCR samples.

Bibliography

M. G. Almiron, B. Lopes, A. L. C. Oliveira, A. C. Medeiros, and A. C. Frery. On the numerical accuracy of spreadsheets. *Journal of Statistical Software*, 34(4):1–29, 2010. ISSN 1548-7660. URL http://www.jstatsoft.org/v34/i04. [p2]

R. Bååth. The state of naming conventions in R. *The R Journal*, 4(2):74–75, 2012. ISSN 2073-4859. URL http://journal.r-project.org/archive/2012-2/RJournal_2012-2_Baaaath.pdf. [p2]

L. D. Brown, T. T. Cai, and A. DasGupta. Interval estimation for a binomial proportion. *Statistical Science*, 16(2):101–117, 2001. ISSN 0883-4237. URL http://www.jstor.org/stable/2676784. 01176. [p13]

P. Burns. Spreadsheet addiction. online, 2014. URL http://web.archive.org/web/20141009042532/http://www.burns-stat.com/documents/tutorials/spreadsheet-addiction/. [p2]

S. A. Bustin. The reproducibility of biomedical research: Sleepers awake! *Biomolecular Detection and Quantification*, 2:35–42, 2014. ISSN 2214-7535. doi: 10.1016/j.bdq.2015.01.002. URL http://www.sciencedirect.com/science/article/pii/S2214753515000030. [p1]

I. Castro-Conde and J. de Uña Álvarez. sgof: An R package for multiple testing problems. *The R Journal*, N(2):NN–NN, 2014. ISSN 2073-4859. URL http://journal.r-project.org/archive/accepted/conde-alvarez.pdf. [p1]

P. Corbisier, L. Pinheiro, S. Mazoua, A.-M. Kortekaas, P. Y. J. Chung, T. Gerganova, G. Roebben, H. Emons, and K. Emslie. DNA copy number concentration measured by digital and droplet digital quantitative PCR using certified reference materials. *Analytical and Bioanalytical Chemistry*, 407(7): 1831–1840, 2015. ISSN 1618-2642. doi: 10.1007/s00216-015-8458-z. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4336415/. [p14, 16, 17]

A. J. Durán, M. Pérez, and J. L. Varona. The misfortunes of a trio of mathematicians using computer algebra systems. Can we trust in them? *Notices of the American Mathematical Society*, 2014. ISSN 0002-9920. doi: dx.doi.org/10.1090/noti1173. URL http://www.ams.org/notices/201410/rnoti-p1249.pdf. [p2]

C. Gandrud. *Reproducible Research with R and RStudio*. Chapman and Hall/CRC, 2013. ISBN 9781466572843. [p3]

R. C. Gentleman, V. J. Carey, D. M. Bates, B. Bolstad, M. Dettling, S. Dudoit, B. Ellis, L. Gautier, Y. Ge, J. Gentry, K. Hornik, T. Hothorn, W. Huber, S. Iacus, R. Irizarry, F. Leisch, C. Li, M. Maechler, A. J. Rossini, G. Sawitzki, C. Smith, G. Smyth, L. Tierney, J. Y. Yang, and J. Zhang. Bioconductor: open

software development for computational biology and bioinformatics. *Genome Biology*, 5(10):R80, 2004. ISSN 1465-6906. doi: 10.1186/gb-2004-5-10-r80. URL http://genomebiology.com/2004/5/10/R80/abstract. 07026. [p2]

M. Guescini, D. Sisti, M. B. Rocchi, L. Stocchi, and V. Stocchi. A new real-time PCR method to overcome significant quantitative inaccuracy due to slight amplification inhibition. *BMC Bioinformatics*, 9(1): 326, 2008. ISSN 1471-2105. doi: 10.1186/1471-2105-9-326. URL http://www.biomedcentral.com/1471-2105/9/326/abstract. PMID: 18667053. [p4]

H. Hofmann, A. Unwin, and D. Cook. Let graphics tell the story - datasets in R. *The R Journal*, 5(1):117–130, 2013. ISSN 2073-4859. URL http://journal.r-project.org/archive/2013-1/RJournal_2013-1_hofmann-unwin-cook.pdf. [p3]

J. Huggett, J. O'Grady, and S. Bustin. How to make mathematics biology's next and better microscope. *Biomolecular Detection and Quantification*, 2014a. ISSN 2214-7535. doi: 10.1016/j.bdq.2014.09.001. URL http://www.sciencedirect.com/science/article/pii/S2214753514000060. [p1]

J. F. Huggett, C. A. Foy, V. Benes, K. Emslie, J. A. Garson, R. Haynes, J. Hellemans, M. Kubista, R. D. Mueller, T. Nolan, M. W. Pfaffl, G. L. Shipley, J. Vandesompele, C. T. Wittwer, and S. A. Bustin. The Digital MIQE Guidelines: Minimum Information for Publication of Quantitative Digital PCR Experiments. *Clinical Chemistry*, 59(6):892–902, 2013. ISSN 0009-9147, 1530-8561. doi: 10.1373/clinchem.2013.206375. URL http://www.clinchem.org/content/59/6/892. [p1, 13]

J. F. Huggett, S. Cowen, and C. A. Foy. Considerations for Digital PCR as an Accurate Molecular Diagnostic Tool. *Clinical Chemistry*, pages 79–88, 2014b. ISSN 0009-9147, 1530-8561. doi: 10.1373/clinchem.2014.221366. URL http://www.clinchem.org/content/early/2014/10/21/clinchem.2014.221366. [p14]

J. F. Huggett, J. O'Grady, and S. Bustin. qPCR, dPCR, NGS – A journey. *Biomolecular Detection and Quantification*, 2015. ISSN 2214-7535. doi: 10.1016/j.bdq.2015.01.001. URL http://www.sciencedirect.com/science/article/pii/S2214753515000029. [p1]

M. Jahn, J. Seifert, M. von Bergen, A. Schmid, B. Bühler, and S. Müller. Subpopulation-proteomics in prokaryotic populations. *Current Opinion in Biotechnology*, 24(1):79–87, 2013. ISSN 0958-1669. doi: 10.1016/j.copbio.2012.10.017. URL http://www.sciencedirect.com/science/article/pii/S0958166912001723. [p15]

M. Jahn, C. Vorpahl, D. Türkowsky, M. Lindmeyer, B. Bühler, H. Harms, and S. Müller. Accurate determination of pasmid copy number of flow-sorted cells using droplet digital PCR. *Analytical Chemistry*, 86(12):5969–5976, 2014. ISSN 0003-2700. doi: 10.1021/ac501118v. URL http://dx.doi.org/10.1021/ac501118v. [p15]

D. A. Khodakov and A. V. Ellis. Recent developments in nucleic acid identification using solid-phase enzymatic assays. *Microchimica Acta*, 181(13-14):1633–1646, 2014. ISSN 0026-3672, 1436-5073. doi: 10.1007/s00604-014-1167-z. URL http://link.springer.com/article/10.1007/s00604-014-1167-z. [p1]

M. Kuhn. CRAN task view: Reproducible research. 2014. URL http://CRAN.R-project.org/view=ReproducibleResearch. [p3]

T. J. Leeper. Archiving reproducible research with R and dataverse. *The R Journal*, 6(1):151–158, 2014. ISSN 2073-4859. URL http://journal.r-project.org/archive/2014-1/leeper.pdf. [p3]

S. Lefever, J. Hellemans, F. Pattyn, D. R. Przybylski, C. Taylor, R. Geurts, A. Untergasser, J. Vandesompele, and o. b. o. t. R. Consortium. RDML: structured language and reporting guidelines for real-time quantitative PCR data. *Nucleic Acids Research*, 37(7):2065–2069, 2009. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/gkp056. URL http://nar.oxfordjournals.org/content/37/7/2065. [p1, 8]

Z. Liu and S. Pounds. An R package that automatically collects and archives details for reproducible computing. *BMC Bioinformatics*, 15(1):138, 2014. ISSN 1471-2105. doi: 10.1186/1471-2105-15-138. URL http://www.biomedcentral.com/1471-2105/15/138/abstract. [p3]

T. Luo, L. Jiang, W. Sun, G. Fu, J. Mei, and Q. Gao. Multiplex real-time PCR melting curve assay to detect drug-resistant mutations of Mycobacterium tuberculosis. *Journal of Clinical Microbiology*, 49 (9):3132–3138, 2011. ISSN 0095-1137, 1098-660X. doi: 10.1128/JCM.02046-10. URL http://jcm.asm.org/content/49/9/3132. [p7]

N. Majumdar, T. Wessel, and J. Marks. Digital PCR modeling for maximal sensitivity, dynamic range and measurement precision. *PLoS ONE*, 10(3):e0118833, 2015. doi: 10.1371/journal.pone.0118833. URL http://dx.doi.org/10.1371/journal.pone.0118833. [p14]

B. D. McCullough and D. A. Heiser. On the accuracy of statistical procedures in microsoft excel 2007. *Computational Statistics & Data Analysis*, 52(10):4570–4578, 2008. ISSN 0167-9473. doi: 10.1016/j.csda.2008.03.004. URL http://www.sciencedirect.com/science/article/pii/S0167947308001606. [p2]

P. Michna and M. Woods. RNetCDF – A package for reading and writing NetCDF datasets. *The R Journal*, 5(2):29–37, 2013. ISSN 2073-4859. URL http://journal.r-project.org/archive/2013-2/michna-woods.pdf. [p3]

C. A. Milbury, Q. Zhong, J. Lin, M. Williams, J. Olson, D. R. Link, and B. Hutchison. Determining lower limits of detection of digital PCR assays for cancer-related gene mutations. *Biomolecular Detection and Quantification*, 1(1):8–22, 2014. ISSN 2214-7535. doi: 10.1016/j.bdq.2014.08.001. URL http://www.sciencedirect.com/science/article/pii/S2214753514000047. [p1]

A. A. Morley. Digital PCR: A brief history. *Biomolecular Detection and Quantification*, 1(1):1–2, 2014. ISSN 2214-7535. doi: 10.1016/j.bdq.2014.06.001. URL http://www.sciencedirect.com/science/article/pii/S2214753514000023. [p1]

P. Murrell. It's not what you draw, it's what you don't draw. *The R Journal*, 4(2):13–18, 2012. ISSN 2073-4859. URL http://journal.r-project.org/archive/2012-2/RJournal_2012-2_Murrell2.pdf. [p3]

G. J. Nixon, H. F. Svenstrup, C. E. Donald, C. Carder, J. M. Stephenson, S. Morris-Jones, J. F. Huggett, and C. A. Foy. A novel approach for evaluating the performance of real time quantitative loop-mediated isothermal amplification-based methods. *Biomolecular Detection and Quantification*, 2: 4–10, 2014. ISSN 2214-7535. doi: 10.1016/j.bdq.2014.11.001. URL http://www.sciencedirect.com/science/article/pii/S2214753514000096. [p1]

D. Nolan and D. L. Temple. *XML and Web Technologies for Data Sciences with R*. 2014. ISBN 978-1-4614-7900-0. URL http://www.springer.com/statistics/computational+statistics/book/978-1-4614-7899-7. [p18]

J. Oh. Automatic conversion of tables to longform dataframes. *The R Journal*, 6(2):16–26, 2014. ISSN 2073-4859. URL http://journal.r-project.org/archive/2014-2/oh.pdf. [p2]

A. Ohri. *R for Cloud Computing - An Approach for Data Scientists*. 2014. ISBN 978-1-4939-1701-3. URL http://www.springer.com/statistics/computational+statistics/book/978-1-4939-1701-3. [p18]

S. Pabinger, S. Rödiger, A. Kriegner, K. Vierlinger, and A. Weinhäusel. A survey of tools for the analysis of quantitative PCR (qPCR) data. *Biomolecular Detection and Quantification*, 1(1):23–33, 2014. ISSN 2214-7535. doi: 10.1016/j.bdq.2014.08.002. URL http://www.sciencedirect.com/science/article/pii/S2214753514000059. [p1, 2, 13]

R Development Core Team. *R Data Import/Export*. R Foundation for Statistical Computing, Vienna, Austria, 2012. URL http://www.R-project.org/. ISBN 3-900051-10-0. [p3]

C. Ritz and A.-N. Spiess. qpcR: an R package for sigmoidal model selection in quantitative real-time polymerase chain reaction analysis. *Bioinformatics*, 24(13):1549–1551, 2008. ISSN 1367-4803, 1460-2059. doi: 10.1093/bioinformatics/btn227. URL http://bioinformatics.oxfordjournals.org/content/24/13/1549. PMID: 18482995. [p3]

S. Rödiger, T. Friedrichsmeier, P. Kapat, and M. Michalke. RKWard: a comprehensive graphical user interface and integrated development environment for statistical analysis with R. *Journal of Statistical Software*, 49(9):1–34, 2012. ISSN 1548-7660. URL http://www.jstatsoft.org/v49/i09. [p3, 16]

S. Rödiger, A. Böhm, and I. Schimke. Surface melting curve analysis with R. *The R Journal*, 5(2):37–53, 2013a. ISSN 2073-4859. URL http://journal.r-project.org/archive/2013-2/roediger-bohm-schimke.pdf. [p2, 4, 7, 11]

S. Rödiger, P. Schierack, A. Böhm, J. Nitschke, I. Berger, U. Frömmel, C. Schmidt, M. Ruhland, I. Schimke, D. Roggenbuck, W. Lehmann, and C. Schröder. 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, 2013b. ISSN 0724-6145. doi: 10.1007/10_2011_132. URL http://link.springer.com/chapter/10.1007/10_2011_132. [p1, 2, 11]

S. Rödiger, C. Liebsch, C. Schmidt, W. Lehmann, U. Resch-Genger, U. Schedler, and P. Schierack. Nucleic acid detection based on the use of microbeads: a review. *Microchimica Acta*, pages 1–18, 2014. ISSN 0026-3672, 1436-5073. doi: 10.1007/s00604-014-1243-4. URL http://link.springer.com/article/10.1007/s00604-014-1243-4. [p1, 11]

S. Rödiger, M. Burdukiewicz, and P. Schierack. chipPCR: an R Package to Pre-Process Raw Data of Amplification Curves . *Bioinformatics*, –(–):NN–NN, 2015. ISSN 1367-4803, 1460-2059. [p3, 4]

RStudio Team. *RStudio: Integrated Development Environment for R*. RStudio, Inc., Boston, MA, 2012. URL http://www.rstudio.com/. [p3]

J. M. Ruijter, M. W. Pfaffl, S. Zhao, A. N. Spiess, G. Boggy, J. Blom, R. G. Rutledge, D. Sisti, A. Lievens, K. De Preter, S. Derveaux, J. Hellemans, and J. Vandesompele. Evaluation of qPCR curve analysis methods for reliable biomarker discovery: Bias, resolution, precision, and implications. *Methods*, 59 (1):32–46, 2013. ISSN 1046-2023. doi: 10.1016/j.ymeth.2012.08.011. URL http://www.sciencedirect.com/science/article/pii/S1046202312002290. [p1, 2, 4, 7]

J. M. Ruijter, P. Lorenz, J. M. Tuomi, M. Hecker, and M. J. B. v. d. Hoff. Fluorescent-increase kinetics of different fluorescent reporters used for qPCR depend on monitoring chemistry, targeted sequence, type of DNA input and PCR efficiency. *Microchimica Acta*, 181(13-14):1689–1696, 2014. ISSN 0026-3672, 1436-5073. doi: 10.1007/s00604-013-1155-8. URL http://link.springer.com/article/10.1007/s00604-013-1155-8. [p1]

D. A. Selck, M. A. Karymov, B. Sun, and R. 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, 2013. ISSN 0003-2700. doi: 10.1021/ac4030413. URL http://pubs.acs.org/doi/abs/10.1021/ac4030413. [p1]

A.-N. Spiess, C. Deutschmann, M. Burdukiewicz, R. Himmelreich, K. Klat, P. Schierack, and S. Rödiger. Impact of Smoothing on Parameter Estimation in Quantitative DNA Amplification Experiments. *Clinical Chemistry*, 61(2):379–388, 2015. ISSN 0009-9147, 1530-8561. doi: 10.1373/clinchem.2014.230656. URL http://www.clinchem.org/content/61/2/379. [p2, 12]

D. Svec, A. Tichopad, V. Novosadova, M. W. Pfaffl, and M. Kubista. How good is a PCR efficiency estimate: Recommendations for precise and robust qPCR efficiency assessments. *Biomolecular Detection and Quantification*. ISSN 2214-7535. doi: 10.1016/j.bdq.2015.01.005. URL http://www.sciencedirect.com/science/article/pii/S2214753515000169. [p4]

P. M. Valero-Mora and R. Ledesma. Graphical user interfaces for R. *Journal of Statistical Software*, 49(1): 1–8, 2012. ISSN 1548-7660. URL http://www.jstatsoft.org/v49/i01. [p3]

E. Viturro, C. Altenhofer, B. Zölch, A. Burgmaier, I. Riedmaier, and M. W. Pfaffl. Microfluidic high-throughput reverse-transcription quantitative PCR analysis of liver gene expression in lactating animals. *Microchimica Acta*, 181(13-14):1725–1732, 2014. ISSN 0026-3672, 1436-5073. doi: 10.1007/s00604-014-1205-x. URL http://link.springer.com/article/10.1007/s00604-014-1205-x. [p1]

J. Wu, R. Kodzius, W. Cao, and W. Wen. Extraction, amplification and detection of DNA in microfluidic chip-based assays. *Microchimica Acta*, 181(13-14):1611–1631, 2014. ISSN 0026-3672, 1436-5073. doi: 10.1007/s00604-013-1140-2. URL http://link.springer.com/article/10.1007/s00604-013-1140-2. [p1]

M. Zeller, W.-C. L. A. Guazzelli, and G. Williams. PMML: An open standard for sharing models. *The R Journal*, 1(1):60–65, 2009. ISSN 2073-4859. URL http://journal.r-project.org/archive/2009-1/RJournal_2009-1_Guazzelli+et+al.pdf. [p3]

Stefan Rödiger (corresponding author)
Faculty of Natural Sciences
Brandenburg University of Technology Cottbus–Senftenberg
Senftenberg
Germany Stefan.Roediger@b-tu.de

Michał Burdukiewicz
University of Wroclaw
Faculty of Biotechnoloy
Department of Genomics
Wroclaw
Poland michalburdukiewicz@gmail.com

Konstantin Blagodatskikh
All-Russia Research Institute of Agricultural Biotechnology
Center for collective use "Biotechnology"
Moscow
Russia k.blag@yandex.ru

Michael Jahn
Helmholtz Centre for Environmental Research - UFZ
Flow cytometry group / Environmental microbiology
Leipzig
Germany michael.jahn@ufz.de

Peter Schierack
Faculty of Natural Sciences
Brandenburg University of Technology Cottbus–Senftenberg
Senftenberg
Germany Peter.Schierack@hs-lausitz.de