BIOINFORMATICS APPLICATIONS NOTE Vol. 24 no. 13 2008, pages 1549–1551
doi:10.1093/bioinformatics/btn227

Gene expression

qpcR: an R package for sigmoidal model selection in quantitative real-time polymerase chain reaction analysis

Christian Ritz[1] and Andrej-Nikolai Spiess[2,*]

[1]Statistics Group, Department of Natural Sciences, Faculty of Life Sciences, University of Copenhagen, Copenhagen, Denmark and [2]Department of Andrology, University Hospital Hamburg-Eppendorf, Hamburg, Germany

Received on October 22, 2007; revised on February 7, 2008; accepted on May 7, 2008

Advance Access publication May 14, 2008

Associate Editor: Trey Ideker

ABSTRACT

Summary: The *qpcR* library is an add-on to the free *R* statistical environment performing sigmoidal model selection in real-time quantitative polymerase chain reaction (PCR) data analysis. Additionally, the package implements the most commonly used algorithms for real-time PCR data analysis and is capable of extensive statistical comparison for the selection and evaluation of the different models based on several measures of goodness of fit.

Availability: www.dr-spiess.de/qpcR.html.

Contact: a.spiess@uke.uni-hamburg.de

Supplementary Information: Statistical evaluations of the implemented methods can be found at www.dr-spiess.de under 'Supplemental Data'.

1 INTRODUCTION

When investigating differential gene expression, quantitative real-time PCR (qPCR) data of two or more different conditions (such as control/treatment or healthy/pathological) are compared by using the fluorescence data acquired by the hardware. The basic approach for quantification between samples relies on the calculation of fold differences by means of the PCR efficiency and the threshold cycle (also often termed 'crossing point').

Many different methods have been developed that can mainly be divided into the following categories according to how they estimate the PCR efficiency and/or the threshold cycles:

Linear methods:

(1) calculation from the slope of a calibration curve (Pfaffl *et al.*, 2001);

(2) based directly on the threshold values (Pfaffl, 2001);

(3) by mid-value-point regression (Peirson *et al.*, 2003); and

(4) with the window-of-linearity method (Ramakers *et al.*, 2003).

Non-linear methods:

(1) fitting four-parameter sigmoidal models onto the fluorescence raw data (Liu *et al.*, 2002; Rutledge, 2004);

(2) fitting an exponential model after identification of the exponential (log-linear) amplification phase (Tichopad *et al.*, 2003; Zhao *et al.*, 2005).

The list above is not exhaustive, as methods for real-time PCR data analysis are in constant development, such as stochastic methods based on branching process theory (Jagers *et al.*, 2003).

The major drawback of the four-parameter models is that they imply symmetry of the lower and upper part of the curve, inducing the same curvature on either side of the inflection point. Thus, we have also included a novel five-parameter sigmoidal model with an additional parameter for asymmetric data. Five-parameter models have only very recently found their way into the dose–response analysis of immunological data (Gottschalk *et al.*, 2005).

Unfortunately it is a fact that different methods in qPCR analysis can yield very different values in respect to PCR efficiency, threshold cycle or estimation of the exponential phase (Karlen *et al.*, 2007; Skern *et al.*, 2005). The best model can severely depend on the used enzymatic system, detection technology or hardware platform. The *qpcR* package we present here will provide valuable assistance to the researcher in the selection of the best model.

2 SOFTWARE FEATURES

The *qpcR* package has been developed for the free statistical *R* environment (www.r-project.org) and will run under the major operating systems. The non-linear fitting engine is based on another add-on package (*drc*; Ritz *et al.*, 2005).

The advantage is that all implemented methods are housed in one software package, and due to the open source nature of the software they can also easily be accessed for inspection, validation or modification purposes.

At present the following methods are available, but the package is in constant development to include novel and interesting approaches:

- Fitting of three-, four- and five-parameter logistic (*Boltzmann*) or log-logistic models onto qPCR data. The novel five-parameter models were included as these have shown to exhibit better performance for asymmetric data (Van der Graaf *et al.*, 1999).
- Selection of the best model by performing nested *F*-tests essentially comparing residual sums of squares for competing models involving three, four or five parameters (Fig. 1A). This is common practise for non-linear fitting regimes (Bates *et al.*, 1988).
- Evaluation of essential PCR parameters (efficiency, threshold cycles) from the sigmoidal fit of the best performing model

*To whom correspondence should be addressed.

```
> m1 <- multdrc(F1.1 ~ Cycles, data=reps, fct=l3())          A
> m1
A 'drc' model.
Call:
multdrc(formula = F1.1 ~ Cycles, data = reps, fct = l3())
Coefficients:
b:(Intercept)  d:(Intercept)  e:(Intercept)
    -11.91          11.33          17.57
> mselect(m1, fctList=list(l4(),l5()), nested=T)
      logLik      AIC          Res var      Nested F test
l3() 48.13457  -88.26914   0.008744114
l4() 51.92993  -93.85985   0.007655675   0.0086435245
l5() 58.29216 -104.58431   0.006038992   0.0007751713
```

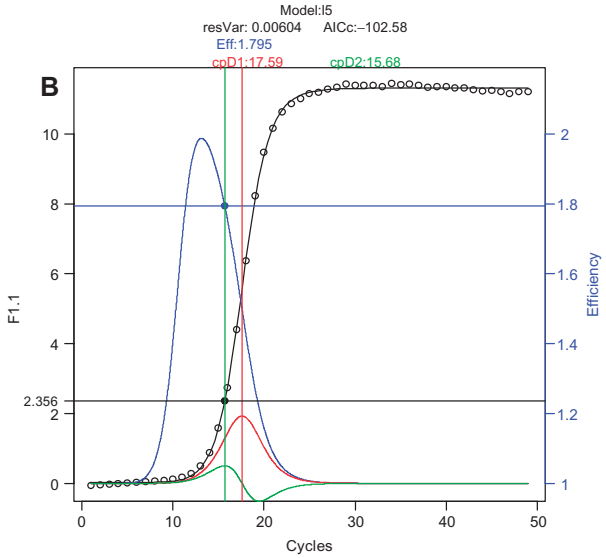

Fig. 1. The *qpcR* package for sigmoidal model selection. (**A**) Console output from the *qpcR* package (example). A three-parameter log-logistic model was fitted to qPCR data obtained from the amplification of the S27a transcript using human testicular RNA and a Lightcycler ™ (Roche) instrument. The best fitting model was selected by the 'mselect' function that chooses the best model according to the nested *F*-tests on the residual variance [here: the five-parameter model (short: 'l5') is compared to the three- and four-parameter models]. (**B**) A fitted five-parameter log-logistic model ('l5') on single qPCR data. The graph displays various parameters from the model fitting process, such as model name, the residual variance, the efficiency curve (blue), the first and second derivatives curves (red and green, respectively) and their maxima (i.e. threshold cycles).

and display as comprehensive graphics (Fig. 1B). The PCR efficiency can be calculated from any point of the sigmoidal fit, but is taken by default at the second derivative maximum of the curve as $E = F(n)/F(n-1)$, with n being the cycle number at second derivative maximum.

- Further optimization of the sigmoidal fitting process by eliminating cycles in the plateau phase (Rutledge, 2004).
- Derivation of values using classical quantitation methods, such as the window-of-linearity method (Ramakers *et al.*, 2003), exponential fitting (Tichopad *et al.*, 2003) or calibration curves (Pfaffl *et al.*, 2001).
- Model assessment and comparison through many measures of goodness of fit such as the residual variance, R^2, the Akaike Information Criterion (AIC) and the root-mean-squared-error (RMSE).
- Batch analysis of many runs with all methods and summarization of these as tabular data.
- Prediction of either fluorescence or cycle values from data and calculation of confidence intervals. The confidence intervals are based on assuming asymptotical normality of the parameter estimates. For the sample sizes that we consider, we believe this approximation will yield a satisfactory coverage.

3 CONCLUSIONS

Currently, several methods for the analysis of qPCR data exist that unfortunately can lead to very different results in the estimation of the essential qPCR parameters, such as the PCR efficiency or the threshold cycle values. We have developed the *qPCR* package for the statistical *R* environment in order to deliver a software tool for the researcher such that the main algorithms are provided in one software. We have also implemented several new functions previously unavailable, such as finding the exponential region by minimizing residual variance or conducting a statistical model selection process for sigmoidal fitting of real-time PCR data. The latter includes the novel five-parameter sigmoidal models that are a viable alternative to the existing four-parameter models due to the inherent feature of taking asymmetrical data into account. Finally we have also added sensitive measures for assessing goodness of fit and for model selection (i.e. the AIC).

In summary, the *qpcR* package provides the researcher with a unifying software tool that is capable of state-of-the-art methods for sigmoidal model fitting in qPCR data analysis.

ACKNOWLEDGEMENTS

We thank C. Feig and H. Capallo-Obermann for fruitful discussions.

Funding: This work was supported by grant Sp721/1-1 of the German Research Foundation (DFG) to A.N.S.

Conflict of Interest: none declared.

REFERENCES

Bates,D.M. *et al.* (1988) *Nonlinear Regression Analysis and its Applications*. John Wiley & Sons, Hoboken, NJ, USA.

Gottschalk,P.G. *et al.* (2005) The five-parameter logistic: a characterization and comparison with the four-parameter logistic. *Anal. Biochem.*, **343**, 54–65.

Jagers,P. *et al.* (2003) Random variation and concentration effects in PCR. *J. Theor. Biol.*, **224**, 299–304.

Karlen,Y. *et al.* (2007) Statistical significance of quantitative PCR. *BMC Bioinformatics*, **8**, 131.

Liu,W. *et al.* (2002) Validation of a quantitative method for real time PCR kinetics. *Biochem. Biophys. Res. Commun.*, **294**, 347–353.

Peirson,S.N. *et al.* (2003) Experimental validation of novel and conventional approaches to quantitative real-time PCR data analysis. *Nucleic Acids Res.*, **31**, e73.

Pfaffl,M.W. (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.*, **29**, e45.

Pfaffl,M.W. *et al.* (2001) Tissue-specific expression pattern of estrogen receptors (ER): quantification of ER alpha and ER beta mRNA with real-time RT-PCR. *Apmis*, **109**, 345–355.

Ramakers,C. *et al.* (2003) Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neurosci. Lett.*, **339**, 62–66

Ritz,C. *et al.* (2005) Bioassay analysis using R. *J. Stat. Soft.*, **12**, 1–22.
Rutledge,R.G. (2004) Sigmoidal curve-fitting redefines quantitative real-time PCR with the prospective of developing automated high-throughput applications. *Nucleic Acids Res.*, **32**, e178.
Skern,R. *et al.* (2005) Relative transcript quantification by quantitative PCR: roughly right or precisely wrong? *BMC Mol. Biol.*, **6**, 10.
Tichopad,A. *et al.* (2003) Standardized determination of real-time PCR efficiency from a single reaction set-up. *Nucleic Acids Res.*, **31**, e122.
Van der Graaf,P.H. *et al.* (1999) Analysis of asymmetry of agonist concentration-effect curves. *J. Pharmacol. Toxicol. Methods*, **41**, 107–115.
Zhao,S. *et al.* (2005) Comprehensive algorithm for quantitative real-time polymerase chain reaction. *J. Comput. Biol.*, **12**, 1047–1064.