



HAL
open science

SATQPCR: Website for statistical analysis of real-time quantitative pcr data.

Corinne Rancurel, van Trang Tran, Céline Elie, Frederique Hilliou

► **To cite this version:**

Corinne Rancurel, van Trang Tran, Céline Elie, Frederique Hilliou. SATQPCR: Website for statistical analysis of real-time quantitative pcr data.. Molecular and Cellular Probes, 2019, 46, 10.1016/j.mcp.2019.07.001 . hal-02264348

HAL Id: hal-02264348

<https://hal.science/hal-02264348v1>

Submitted on 25 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

SATQPCR: website for statistical analysis of real-time quantitative PCR data

Corinne Rancurel¹, Trang van Tran², Céline Elie² and Frédérique Hilliou^{1,*}

¹INRA/CNRS/UCA ISA 06903 400 route des Chappes, BP167 Sophia Antipolis France corinne.rancurel@inra.fr; frederique.hilliou@inra.fr, ²BIOASTER, 40 Avenue Tony Garnier 69007 Lyon 69007, France; trvtrang@yahoo.fr; celine.elie1@gmail.com

*To whom correspondence should be addressed (frederique.hilliou@inra.fr).

Abstract

SATQPCR is a web tool providing statistical analysis of real-time quantitative PCR data including all MIQE rules (gene efficiency, selection of reference genes and normalization with them).

Our application is a quick tool that provides to the biologist, graphs as well as statistical tables summarizing their results with the chosen methods (t-test or ANOVA with Tukey test).

The application is available at <http://satqpcr.sophia.inra.fr> with a demo dataset. Source code can be found at <https://framagit.org/>.

Supplementary information: Tutorials at <http://satqpcr.sophia.inra.fr/cgi/help.cgi>

Keywords: real-time quantitative PCR; statistics; web-application; MIQE; transcriptomic; data analysis

1 Introduction

Real-Time quantitative PCR (RT-qPCR) has been used in the last two decades by biologists in several fields of biology. The main application is to quantify relative gene expression [3]. The wide success of this technique can be attributed to low cost, speed, sensitivity and specificity of the measurements compared to techniques such as northern blot. It has also benefited from the increase of gene sequences in data banks. RT-qPCR can quantify low gene expression but also simultaneous multiple gene expressions in a single tube [11] with the use of specific primers. The accuracy of RT-qPCR techniques has contributed to the development of biomarkers in various fields from medical diagnostic to bioaggressors detection for agricultural crops as well as gene expression knowledge in academic research area.

From the beginning of RT-qPCR development, scientists have proposed rules to design and analyze these biological data through Minimum Information for Publication of Quantitative Real-Time PCR Experiments (MIQE) guidelines [3,4]. However, a single gene is often used to normalize expression of gene of interest (GOI) and primers efficiency are not given or used for this normalization. Student t-test is often misused in RT-qPCR analysis. Tools to implement MIQE rules exist such as geNorm or RefFinder (review in [7]) to evaluate reference genes and as HTqPCR for normalization, quantification and statistics (review in [8]). These tools are either performing only one step of MIQE rules or are implemented in programming languages such as R that necessitate coding knowledge.

Our web-application aims to facilitate the analytical work of biologists since we provide a ready to use tool to statistically analyze RT-qPCR data according to the MIQE guidelines. SATQPCR is a website that supports all quantitative PCR machine data as it uses Cq “parameters” usually provided by the machine. Cq value results from fluorescence measurement of a sample obtained and collected after each PCR cycle. It represents the accumulation of an amplicon in a sample. Our application needs a text file associating the basic Cq data and the efficiency calculated for each gene to quantify. The normalization of the data is based on the search of genes with the lowest gene expression variability using geNorm method [10]. The statistical analysis of our previous R script was improved [6] by including one-way and two-ways ANOVA. All graphs and tables resulting from the statistical analysis are visible on the website and can be downloaded using a single link.

2 Results

3.1 Description of files to provide

In the two needed files, all decimals must be written with a dot “.” and each file should be saved as a text file with a tabular separator.

3.2 Expression data file

You should have at least 3 biological replicates and 2 technical replicates for each sample of your experiments. In the text file you provide for the analysis (see example.txt on our web application and FAQ paragraphs “How to design an experiment”, “How to enter your data”) you have to give gene names in the first line, gene efficiency in the 2nd line. Each column corresponds to a gene. Each following line displays Cq values for all genes analyzed. Each cell represents the value for one technical replicate of one biological replicate. All technical replicates of the same biological sample have exactly the same name in the first column. The biological replicates are organized in blocks of lines. Each biological replicate has the same number of row (balanced experimental design). The missing values must be replaced by 'NA' value.

3.3 Factors file for ANOVA

You should give the ANOVA factors associated to your experimental design in a text file. An example of this is the “factor.txt” file on the website. The first row describes all samples with one line per sample. The following rows correspond to the different factor of the ANOVA that will explain the experiments. In each cell the modality corresponding to this column factor and line sample is given.

3.4 Selection of reference genes (REF)

REFs are the most stably expressed genes across all tested samples and all biological replicates of one data set. Their identification is based on the stability parameter and coefficient of variation. SATQPCR uses the algorithm described in [10] to identify the reference genes. SATQPCR will provide a list of reference genes for you by finding the most stables genes (at least 2). The user can also provide a list of reference genes however it is not recommended. The REFs are specific to each experiment and should be validated for each data analyzed. These reference genes will be used to compute normalization factors and to normalize and compute relative quantities for each gene in each sample [6].

3.5 Rescaling normalized expression for each gene of interest (GOI)

By default, for each gene, your data are rescaled using the sample with the lowest expression. Our web-application allows to rescale your expression data with a sample chosen by the user. User will have to choose the sample name in the box “Sample name used to rescale”.

3.6 Web-application SATQPCR

User should provide two files “sample file” and “factor file” to the web-application. The “sample file” is mandatory. If an ANOVA is computed, the web-application is asking to upload the “factor file”. When you upload the files, the application is automatically checking the consistency of your files and it returns the number of biological and technical replicates found in the “sample file”. It automatically attributes a unique ID for your “job”. The web-form also gives an option on the number reference genes to choose as well as if the application or the user should choose them. The form also asks if a rescaling of the gene expression has to be done using a specific sample condition. By default, the rescaling will be done with the lowest expressed sample. Finally, the form offers to compute either standard error or standard deviation. Once these 5 parameters are submitted in the form, all normalization, rescaling and statistical analyses are run

3.7 Visualization

All results are given in a single web-page where a link is given to download a zipped folder containing all files (graphical files, statistical tables) as well as the two input files (sample and factor) and a parameter file summarizing all options chosen in the form provided by the application. The graphical files are in jpeg format and all others files are text files.

3 Conclusion

Even if new statistical analysis methods have been proposed for RT-qPCR such as the Unified Censored Normal Regression Model [9], the procedure using references genes as performed in our web-application SATQPCR is adequate for small datasets which are the most often used. Our web application is free compare to commercial packages such as

qBASE+ Biogazelle NV [5] and GenEx (MultiD Analyses AB) and the biologist does not need to learn R language as in the open source tool DAG Expression [1]. Our web-application is following MIQE rules with normalization steps including gene efficiency and the use of at least two "reference" genes. The application is very fast and provides graphs as well as statistical tables summarizing the results of the chosen methods (t-test or ANOVA with Tukey test). Our web-application will provide a quick tool (few seconds) to the biologists in order to analyze RT-qPCR data for any set of samples including technical and biological replicates using MIQE rules and statistical tests.

Acknowledgements

The authors would like to thank E Wanjberg, S Touzeau and A Paquet for fruitful statistical discussion and J. Gouzy for help in installing a Virtual Machine. FH would like to thank MC Arias for helpful discussion.

Funding

This work was supported by the Institut National de Recherche Agronomique-Département de Santé des Plantes et Environnement [CATI-BBRIC to FH and CR].

Conflict of Interest: none declared.

References

- [1] Ballester, M *et al.* (2013) Dag expression: high-throughput gene expression analysis of real-time PCR data using standard curves for relative quantification. *Plos One* 8 (11): E80385. DOI 10.1371/journal.pone.0080385
- [2] Bustin, S.A. *et al.* (2013) The need for transparency and good practices in the qPCR literature. *Nat Methods* 10 (11): 1063–67. <https://doi.org/10.1038/nmeth.2697>
- [3] Bustin, S.A. (2004) A-Z of quantitative PCR. Publisher: International University Line. ISBN 0963681788.
- [4] Bustin, S.A. *et al.* (2010) MIQE precis: practical implementation of minimum standard guidelines for fluorescence-based quantitative real-time PCR experiments. *BMC Molecular Biology* 11 (1): 74. doi:10.1186/1471-2199-11-74
- [5] Hellemans, J. *et al.* (2007) QBASE relative quantification framework and software for management and automated analysis of real-time quantitative PCR Data. *Genome Biology* 8 (2): R19. <https://doi.org/10.1186/gb-2007-8-2-r19>
- [6] Hilliou, F. and Tran, T. (2013). RqPCR analysis: Analysis of quantitative Real-Time PCR Data. In *International Conference On Bioinformatics Models, Methods And Algorithms (Bioinformatics 2013)*, Edited By Scitepress Digital Library, 202–11. Barcelona, Spain. DOI 10.5220/0004312002020211
- [7] Kuang, J. *et al.* (2018) An overview of technical considerations when using quantitative Real-Time PCR analysis of gene expression in human exercise. *Research Plos One* 13 (5): E0196438. DOI 10.1371/journal.pone.0196438
- [8] Pabinger, S. *et al.* (2014) A survey of tools for the analysis of quantitative PCR (qPCR) data. *Biomolecular Detection And Quantification* 1 (1): 23–33. DOI 10.1016/J.BDQ.2014.08.002
- [9] Pipelers, P. *et al.* (2017) A Unified Censored Normal Regression Model For QPCR differential gene expression analysis. Edited By Alfred S Lewin. *Plos One* 12 (8): E0182832. DOI 10.1371/journal.pone.0182832
- [10] Vandesompele, J. *et al.* (2002) Accurate Normalization Of Real-Time Quantitative RT-PCR Data By Geometric Averaging Of Multiple Internal Control Genes. *Genome Biology* 3 (7): Research0034.1-Research0034.11. <https://doi.org/10.1186/gb-2002-3-7-research0034>
- [11] Wittwer, C.T. *et al.* (2001) Real-Time multiplex PCR assays. *Methods* 25 (4): 430–42. DOI 10.1006/meth.2001.1265