

QPCRinR: A powerful and interactive web application for the analysis of RT-qPCR data with linear mixed models

Ioannis Iliopoulos and Ioannis Theologidis

Step-by-step User's Guide

Chapter 1: Introduction

QPCRinR is a web tool for the analysis of RT-qPCR data with Linear Mixed Models (LLMs). It is accessible at <https://itheo.shinyapps.io/QPCRinR/>. The interface contains multiple tabs at the top of the webpage that are used by means of a linear progression from data submission to result collection.

On the *Welcome* page, there is this guidance brochure for an easy and comprehensive usage of QPCRinR, which opens by clicking the button [*QPCRinR User's Guide*]. This supplementary material comprises a step-by-step tutorial for the analysis of the incorporated demo data set which involves partial data for the expression of four genes (3 targets + 1 reference) at three different conditions (2 treatments + 1 control) at three distinct time-points. The demo data set, more guides and resources can be found at <https://github.com/theojohn/QPCRinR>.

Start by clicking *Load Data* tab or [*Launch!*].

Chapter 2: Data preparation and submission

2.1. Data preparation

QPCRinR requires one single text file that can be either comma, tab or semicolon delimited (i.e. .txt, .csv). The application accepts single sheet excel files (*.xlsx format), as well. Data should be organized as a table with the first row comprising the header of each column. The structure of the table should be in long format and contain at least a column with threshold cycle values (C_T) and a column with gene names. Single-letter names, single numbers and special characters should be avoided as headers. The same holds for all other information included in the body of the data table, except for the C_T column about single numbers. In this tutorial, the demo data set is a file in tab delimited txt format.

2.2 Data submission

Data submission is done in the *Load Data* tab of the application. The left panel of the tab contains all the tools and buttons for proper data input and manipulation. [*Header*] and [*StingAsFactors*] boxes are ticked by default and should remain as such. To upload a data set, clicking [*Browse...*] under the **Input Data** label is required. The delimiter type can be adjusted either before or after data uploading by clicking the appropriate radio button.

For this tutorial, [Demo Set] should be clicked. As soon as this happens, an interactive data table and a summary frame are displayed in the main panel of the page. In the table, the first 10 entries appear. At the same time, the left panel is expanded, and the Variable Definition menu appears. If the selection of the delimiter at the **Separator** menu is appropriate for this particular data set (i.e., “Tab”), then the columns in the table of the main panel will appear well-defined and clearly separated. Furthermore, on the left panel, under the title **Variable Name**, the scrollable menu should only read “experiment”. The summary frame underneath will look as shown below:

```
'data.frame':      174 obs. of  8 variables:
 $ experiment : int  9 9 9 9 9 9 9 9 9 ...
 $ Calibrators: Factor w/ 2 levels "", "Calib": 1 1 1 1 1 1 2 2 1 1 ...
 $ samples    : int  19 20 25 26 31 32 13 13 19 20 ...
 $ replicates : Factor w/ 48 levels "13_1","13_2",...: 13 15 25 27 37 39 1 3 13 15 ...
 $ gene       : Factor w/ 4 levels "AC01","PP00",...: 4 4 4 4 4 4 4 4 3 3 ...
 $ times      : Factor w/ 3 levels "02h","06h","24h": 1 1 2 2 3 3 3 3 1 1 ...
 $ treatment  : Factor w/ 3 levels "Control","FLG22",...: 1 1 1 1 1 1 1 1 1 1 ...
 $ Ct         : num  18.2 17.3 19.2 18.1 18.6 ...
```

Otherwise, the data in the main table will look disorganized, the summary panel will contain a long, weird string of letters, and the content of the **Variable Name** scrollable menu will contain collapsed variable names.

After confirming that “Tab” is the **Separator** selection, it is the moment for Variable Definition. All but C_T variables in the summary frame should be *factors*. To achieve this, **Variable Name** and **Variable Class Type** should be adjusted. Select “experiment” in the **Variable Name** drop-down list and then “factor” in the **Variable Class Type** drop-down lists. Press [Submit Class Change]. Repeat the same for the variable “samples”. The summary frame should then look like this:

```
'data.frame':      174 obs. of  8 variables:
 $ experiment : Factor w/ 2 levels "9","10": 1 1 1 1 1 1 1 1 1 1 ...
 $ Calibrators: Factor w/ 2 levels "", "Calib": 1 1 1 1 1 1 2 2 1 1 ...
 $ samples    : Factor w/ 21 levels "13","15","17",...: 4 5 10 11 16 17 1 1 4 5 ...
 $ replicates : Factor w/ 48 levels "13_1","13_2",...: 13 15 25 27 37 39 1 3 13 15 ...
 $ gene       : Factor w/ 4 levels "AC01","PP00",...: 4 4 4 4 4 4 4 4 3 3 ...
 $ times      : Factor w/ 3 levels "02h","06h","24h": 1 1 2 2 3 3 3 3 1 1 ...
 $ treatment  : Factor w/ 3 levels "Control","FLG22",...: 1 1 1 1 1 1 1 1 1 1 ...
 $ Ct         : num  18.2 17.3 19.2 18.1 18.6 ...
```

It is crucial that [Submit Class Change] is clicked **for each variable**, otherwise the changes will not occur. At this point, the user can proceed to the next step by clicking either the right arrow at the bottom-right of the page, or the *Model Specification* tab at the top. It is also possible for the user to further inspect and explore the input data set in the main panel.

Chapter 3: Model Specification

Modeling RT-qPCR data by Linear Mixed Models (LMMs) is the core feature of QPCRinR. The demo data set is a part of a big experiment that has been conducted at stages, defined as “experiment” runs. For each run, calibrator samples have been used to make the combination of the information plausible. Calibrator samples are particular RNA preps that are stocked and used as spikes in consecutive assays, so that multiple experiments can be normalized and combined (see the main article and references therein). When calibrator samples are included in an analysis, an additional column with a specific header is required in the data set. In this column, only rows corresponding to calibrators should be filled in with a word-tag so as to distinguish this kind of samples from all others.

For this tutorial, tick the box next to [*Inter-Run Calibration*]. Then, the left panel is expanded, and two text boxes appear. Type “Calib” in the first (**Please insert Calibrators' Tag**) and “experiment” in the second (**Please insert Plate Column Name**). The inputs in these boxes should be strictly the same as the word-tag and header of the calibration column.

After calibrator samples are set, the actual model of the analysis should be defined. LMMs, that constitute the core of QPCRinR, are specified by a formula indicating the dependent variable of the model (response), as well as the independent and the random variables. This can be achieved by dragging and dropping the names of variables in visually defined buckets, in the main panel of the *Model Specification* tab. Alternatively, for users that are more familiar with formulas in R, the model formula can be explicitly typed in the provided text box.

In the main panel, click variable “Ct” from the **Model Components** bucket and drag it to the **Response – Drag CT value here**. This way, the dependent variable is specified. Only one variable is allowed in the **Response** bucket. To formulate the independent variables of the model, i.e., the fixed part, drag and drop “gene”, “treatment” and “times” from the **Model Components** bucket to the **Fixed Effects**. Multiple objects are allowed, and when doing so, their main effects and interactions will be modeled. In the same way, drag and drop “samples” to the **Random Effects** bucket. The equivalent formula in R to this process, if the manual option was selected (by ticking the “*or input model formula manually*” box) would be the following:

$$Ct \sim \text{gene} * \text{treatment} * \text{times} + (1|\text{samples})$$

In this model, the part after the “+” sign represents a random sample-specific intercept. To run this model, click [*Run Model!*].

Shortly, the output of the analysis will appear under the clicked button. The first table contains the model evaluation metrics that inform the user about the performance of the model.

Underneath, two grey panels will be displayed; 1) the summary of the model and 2) the factor combination estimates with their confidence intervals (CI).

At this stage, the user can test and compare different model expressions and select the best before proceeding to the next stage. Below, other model formulas that can be inputted manually are provided:

- i. $Ct \sim \text{gene} * \text{treatment} * \text{times}$
- ii. $Ct \sim \text{gene} * \text{treatment} * \text{times} + (1|\text{gene})$
- iii. $Ct \sim \text{gene} * \text{treatment} * \text{times} + (1|\text{samples})$
- iv. $Ct \sim \text{gene} * \text{treatment} * \text{times} + (1|\text{samples}) + (1|\text{gene})$
- v. $Ct \sim \text{gene} * \text{treatment} * \text{times} + (1|\text{samples}) + (1|\text{samples:gene})$

Some of them produce worse or better metrics, while others produce warning balloons. The user can choose the best model, which in this case is the model (v); a model that includes sample-specific and sample-by-gene-specific random intercepts. Note that for the loglikelihood (logLik) metric, the higher value the better, while the opposite holds for both AIC and BIC.

Case	logLik*	AIC**	BIC**
i.	-185.42	444.85	554.73
ii.***	-179.56	435.11	547.97
iii.	-150.85	377.71	490.56
iv.***	-150.85	379.71	495.53
v.	-20.86	119.71	235.54

* The higher, the better

** The lower, the better

*** Warning messages for (potential) failure of the models to converge

Chapter 4: Contrast Assignment

When model selection and run are concluded, the definition of reference gene(s) and control treatment follow, so that contrast specification and, thus, log2 fold-changes (log2FC) per factor combination are calculated. Note that the calculation of the marginal means of factor combinations is possible only when interactions among those factors are included in the model formula at the previous stage – the default case for the drag-and-drop input. In the manual input, this is signified by the “*” character between the names of main effects.

Click “UBI3” box to choose this gene as reference. Then the left panel is expanded downwards. If the selection under *Control in...* is “treatment” click [... *continue*...]; otherwise, click the drop-down list and choose “treatment” and then click [... *continue*...]. The left panel is expanded even more and *Choose Control* option appears. In this drop-down menu, the levels of factor “treatment” appear. Choose *Control* and click [*Confirm Selection*]. The log2FC values (“estimate”) appear in the main panel (consult main text for methodological background of the log2FC calculations), as well as their standard errors, degrees of freedom, t-test ratios and p-values.

contrast	estimate	SE	df	t.ratio	p.value
AC01_FLG22_02h	4.903	1.32	27	3.710	0.0009
PP00_FLG22_02h	4.025	1.32	27	3.045	0.0051
TD20_FLG22_02h	1.928	1.32	27	1.458	0.1563
AC01_S2_02h	2.318	1.32	27	1.754	0.0908
PP00_S2_02h	1.730	1.32	27	1.309	0.2015
TD20_S2_02h	-0.145	1.32	27	-0.110	0.9135
AC01_FLG22_06h	3.922	1.32	27	2.968	0.0062
PP00_FLG22_06h	4.892	1.32	27	3.702	0.0010
TD20_FLG22_06h	2.932	1.32	27	2.219	0.0351
AC01_S2_06h	2.122	1.32	27	1.606	0.1199
PP00_S2_06h	2.213	1.32	27	1.674	0.1057
TD20_S2_06h	1.115	1.32	27	0.843	0.4064
AC01_FLG22_24h	2.950	1.32	27	2.232	0.0341
PP00_FLG22_24h	2.662	1.32	27	2.014	0.0540
TD20_FLG22_24h	9.752	1.32	27	7.379	<.0001
AC01_S2_24h	3.362	1.32	27	2.544	0.0170
PP00_S2_24h	2.197	1.32	27	1.663	0.1080
TD20_S2_24h	3.354	1.32	27	2.538	0.0172

Degrees-of-freedom method: kenward-roger

These estimates correspond to the log2 of fold change of a target gene in the specified conditions described by the symbols between the underscore characters (“_”), compared to the control condition and normalized by the expression of the reference gene.

Chapter 5: Result table and visualization

As soon as log2FC estimates are displayed, the user can proceed to the *Results* tab. No calculations occur here, only extraction and visualization of the outcome of the analysis.

The result table appears in the main panel of the *Results* tab. It contains the same information as the output in the *Contrast Assignment* tab, but now it is searchable and interactive. Moreover, its contents can be saved as a csv file, by clicking [*Download Result Table*].

QPCRinR provides the possibility to create and download highly customizable bar plots of the output displayed in the main panel of the *Results* tab.

In the left panel click and select “times” in the drop-down menu under the title *Define x-axis* to define the x-axis of the plot. The y-axis is predefined by default to correspond to the log2FC estimates. Select “treatment” under the title *Fill by* to give each bar of the plot a different color, according to the corresponding treatment and then select “gene” under the title *Split by* to create facets (sub-panels) by each gene. Click [*Create Plot*] to generate the plot in the main panel, under the results table. (Additionally, in case there are more than three main factors to be displayed, the *Merge Factors* functionality can combine multiple factors for visualization purposes).

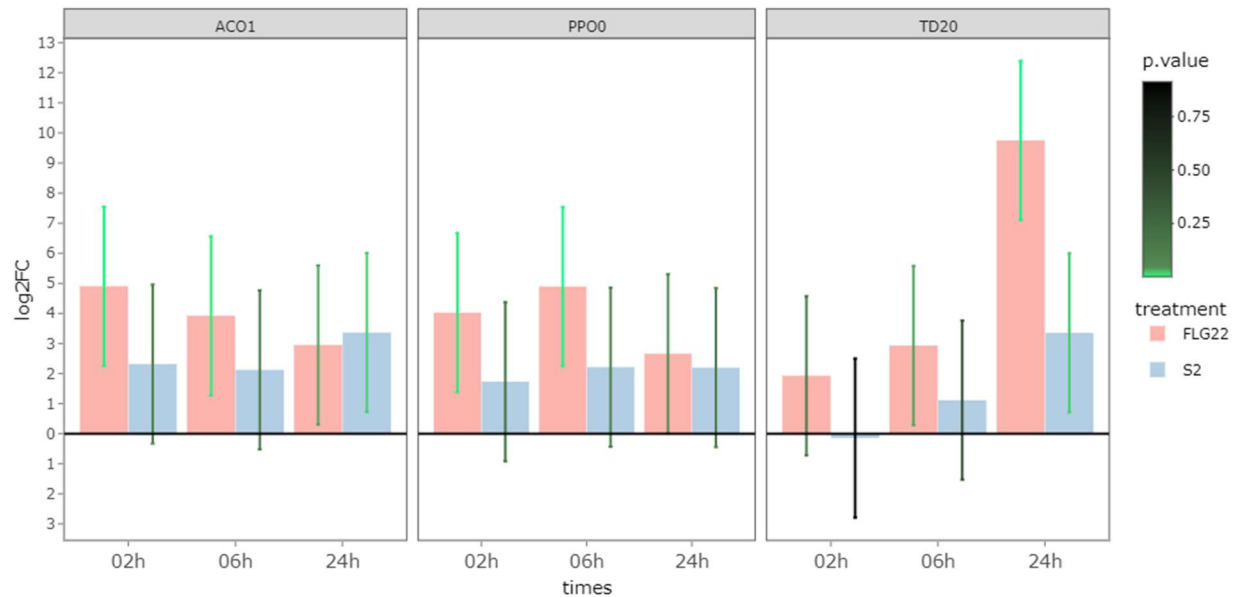


Figure. The results bar plot in “theme_test” and “ColPast_16” format. Timepoints define the x-axis while different colors correspond to the different treatments (FLG22 and S2). The third, -Control- defines the baseline together with the reference gene(s). Subpanels (facets) display different target genes (ACO1, PPO0 and TD20). Each bar represents the relative log2 fold-change (log2FC) estimate of each contrast, while the error bars represent the 2*SE of the estimate. Hence, comparisons whose bars cross the x-axis are not significant at $\alpha = 5\%$.

Note that the scale in the y-axis is always positive; a bar over the baseline designates upregulation, while a bar under the baseline means downregulation of the target gene at any specific condition compared to the control and normalized by the reference(s) gene(s). The interval bars depict 2x the standard error (2*SE) of each estimate so that a visual inspection of significance is feasible.

The format of the plot can be further customized in the *Select properties of plot* menu, in the left panel. The sliders under *Height* and *Width* control the relevant properties in pixel dimensions. There are also options for the rotation of the x-axis labels for the cases they do not fit in horizontal configuration. Furthermore, the general appearance (theme) and the color palette of the plot can be edited in the last two drop-down menus. There are even more options that appear when hovering the cursor over the plot, the most important being the save as function that is available after clicking the camera icon.

TROUBLESHOOTING

Problem/Message	Possible reason	Suggested solution
V1, V2, ... Vn appear in the column names of the input data table [Load Data]	<i>Header</i> button is not selected	Tick <i>Header</i> button in the left panel
V1, V2, ... Vn appear in the <i>Variable Name</i> drop-down menu [Load Data]	<i>Header</i> button is not selected	Tick <i>Header</i> button in the left panel
Weird string of letters in the <i>Variable Name</i> drop-down menu [Load Data]	Data file column separator not properly defined	Select the appropriate <i>Separator</i> in the left panel
There's an error Invalid [...] string at “ [Load Data]	Selected data file not in appropriate format	Select a txt, csv, or xlsx file
There's an error Contrasts can be applied only to factors with 2 or more levels [Run Model]	Improper input of calibrator tag or plate column name	Rewrite the tag and name carefully in the text box. Pay attention to “hidden” space characters
There's an error Objective in x0 returns NA [Run Model]	Improper header (column name) in one or more factors	Replace header names that start with numbers or contain special characters (/ , _ . , * , +) or are duplicated in other columns
There is an error :1:4: unexpected symbol 1: v. Ct ^	Improper characters before “Ct” in the beginning of model formula (here “v.”)	Remove “v.” or any other improper characters in the beginning of the formula
Boundary (singular) fit: [...][Run Model]	Possible overfitting of the model by complex random effects structure	Try to fit the most complex model consistent with the experimental design, removing terms required to allow a non-singular fit
NaN or Inf [Contrast Assignment]	Interactions of fixed factors not modeled	Replace “+” with “*” in the manual model formula
There's an error Object ‘x_y_z_NA’ not found [Contrast Assignment]	Improper level names in one or more factors	Replace level names that start with numbers or contain special characters (/ , _ . , * , +) or are single letters or are duplicated in other columns
Multiple error bars per category in plots [Results]	Two or more factors are overlapping	Use <i>Fill by</i> and <i>Split by</i> for different factors. Use <i>Merge Factors</i> when number of fixed factors > 3