Presubmission Inquiry

*Nature Protocols*

Robert Clark\*1, Matt Brousil2, Saumik Basu1, Clare Casteel3, Alan Goodman4, David W. Crowder1

1 Department of Entomology, Washington State University, Pullman, WA, USA

2 Center for Environmental Research, Education and Outreach, Washington State University, Pullman, WA, USA

3 Department of Plant Pathology, University of California Davis, Davis, CA, USA

4 School of Molecular Biosciences, Washington State University, Pullman, WA, USA

\* Corresponding author: [robert.e.clark@wsu.edu](mailto:robert.e.clark@wsu.edu)

Title:

Statistical analysis of experimental real-time PCR data (DDCT) using multivariate and generalized linear models in R.

Overview of technique:

The statistical protocol outlined in this manuscript builds on established statistical theory from design of experiments (DOE) and widely accepted molecular approaches for quantifying gene expression. While many tools are available for calculating DDCT values and troubleshooting rtPCR, there is currently few open-source techniques for rigorous statistical analysis that can evaluate the outcomes of experiments. Using a fully annotated and open source R-software pipeline, this manuscript will carry experimenters from raw thermal cycler data to hypothesis-testing models, to visualization of results with publication quality figures. Statistical arguments for this new approach compared to current techniques are also demonstrated using simulations. Examples are provided from multiple types of experiments including gene expression in response to environmental stress and mRNA transcripts generated by pathogens (viral titer).

Applications and target audience:

This method is robust and can be used to evaluate any rtPCR data including delta-Ct (absolute quantification) and delta-delta ct (relative expression compared to an experimental control). Effectively, any researcher that wishes to analyze transcript accumulation data will be able to test hypotheses quickly and accurately. Importantly, our target audience is not limited to bioinformaticists or statisticians; the R-pipeline presented here can be utilized by anyone with a basic familiarity with ANOVA in R. The results of the pipeline, mainly plots of means and standard errors, follow a familiar format already widely presented in molecular biology publications.

Advantages, limitations, and adaptations:

The method is straightforward from a statistical point of view, relying on tried-and-true univariate (GLM) and multivariate (MANOVA) models. Since R is open source software, the approach will also be widely available, and the script is modifiable. The primary limitation of this method is that not all rtPCR data may not always fit assumptions of linear models, and that it will be up to the experimenter to establish if this is the case. To meet this challenge, we have provided simple diagnostic tools to aid this issue. Future adaptations to the method may be necessary if delta-delta CT method is deprecated and different mathematical assumptions are made about rtPCR data in the future. However, even gene expression data structure changes, the statistical approach of using parameter estimates from MANOVA/GLM will be retained.

Example Abstract:

Quantification of gene expression using rtPCR and Delta-Ct transformation has become a fundamental tool for molecular biologists. Despite the ubiquity of this technique, accessible statistical approaches to evaluate data generated by rtPCR are still not available. Since the results of complex laboratory experiments hinges on effective statistical analysis, molecular biologists should have a toolkit available that applies widely accepted modeling approaches. This protocol provides an R-software based pipeline for modeling relative gene expression for multiple loci using multivariate methods (MANOVA) and examples of how to test complex hypotheses using generalized linear models (GLMs and GLMMs). Additionally, simulation scenarios are provided to justify how parameters are estimated, such as means and standard errors, from these statistical models.