Genomics 94 (2009) 146–152

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Methods

A powerful and flexible linear mixed model framework for the analysis of relative quantification RT-PCR data

Juan Pedro Steibel [a,b,*], Rosangela Poletto [c], Paul M. Coussens [a], Guilherme J.M. Rosa [d]

[a] Department of Animal Science, Michigan State University, East Lansing, MI 48824, USA
[b] Department of Fisheries and Wildlife, Michigan State University, East Lansing, MI 48824, USA
[c] Department of Animal Sciences, Purdue University, West Lafayette, IN 47907, USA
[d] Department of Dairy Science, University of Wisconsin – Madison, Madison, WI 53706 USA

ARTICLE INFO

Article history:
Received 24 January 2009
Accepted 28 April 2009
Available online 5 May 2009

Keywords:
Gene expression
qRT-PCR
Linear mixed models
Hypothesis testing

ABSTRACT

Quantitative reverse transcription polymerase chain reaction (qRT-PCR) is currently viewed as the most precise technique to quantify levels of messenger RNA. Relative quantification compares the expression of a target gene under two or more experimental conditions normalized to the measured expression of a control gene. The statistical methods and software currently available for the analysis of relative quantification of RT-PCR data lack the flexibility and statistical properties to produce valid inferences in a wide range of experimental situations. In this paper we present a novel method for the analysis of relative quantification of qRT-PCR data, which consists of the analysis of cycles to threshold values (C_T) for a target and a control gene using a general linear mixed model methodology. Our method allows testing of a broader class of hypotheses than traditional analyses such as the classical comparative C_T. Moreover, a simulation study using plasmode datasets indicated that the estimated fold-change in pairwise comparisons was the same using either linear mixed models or a comparative C_T method, but the linear mixed model approach was more powerful. In summary, the method presented in this paper is more accurate, powerful and flexible than the traditional methods for analysis of qRT-PCR data. This new method is especially useful for studies involving multiple experimental factors and complex designs.

Introduction

Reverse transcription (RT), followed by quantitative polymerase chain reaction (qPCR), is currently the method of choice to quantify levels of messenger (m)RNA [1]. At present, there are several instrumentations and chemistries available for implementation of this technique, all of which rely on the same fundamental principle [2]. This principle consists of the specific amplification of cDNA from a target transcript in several cycles of PCR, coupled with measurement of a fluorescence intensity that is assumed to be directly proportional to the amount of product in each cycle [3]. This methodology has been extensively validated, and its accuracy and specificity have been proved for the different chemistries available [4].

The quantitative output of the qRT-PCR consists of an amplification curve, which is composed of a set of cycle numbers and associated fluorescence intensities that are ulteriorly summarized in a single value called the cycles to threshold (C_T). The C_T is a unitless value defined as the fractional cycle number at which the sample fluorescence signal passes a fixed threshold above the baseline. Because the threshold is set within the exponential amplification phase, the C_T is proportional to the (negative) log of the initial transcript copy number (or log-transcript concentration) of the assayed sample. The constant of proportionality of the C_T to the log-concentration is the amplification efficiency (E).

Absolute and relative quantification strategies can be applied to measure mRNA abundance using qRT-PCR [3,5]. Relative quantification compares the expression of a target gene under various conditions (treatments) normalized to the measured expression of an internal control [6] (assumed to be constantly expressed across samples). In general, the numerous mathematical expressions available for such calculation [7–16] may be summarized by the equation below [6]:

$$\mathrm{FC}_{\mathrm{trt}_1:\mathrm{trt}_2} = \frac{\left(E_{\mathrm{Target}}\right)^{\Delta C_{\mathrm{T(target)}}(\mathrm{trt}_2 - \mathrm{trt}_1)}}{\left(E_{\mathrm{Control}}\right)^{\Delta C_{\mathrm{T(control)}}(\mathrm{trt}_2 - \mathrm{trt}_1)}}, \quad (1)$$

where, $\mathrm{FC}_{\mathrm{trt}_1:\mathrm{trt}_2}$ is the relative expression (fold-change) of the target gene in a sample from treatment 1 compared to a sample from treatment 2, E_{Target} and E_{Control} are the amplification efficiencies of the target and the control genes, respectively, and $\Delta C_{\mathrm{T(target)}}(\mathrm{trt}_2 - \mathrm{trt}_1)$

* Corresponding author. Department of Animal Science, Michigan State University, East Lansing, MI 48824, USA. Fax: +1 517 353 1699.
E-mail address: steibelj@msu.edu (J.P. Steibel).

0888-7543/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2009.04.008

and $\Delta C_{T(Control)}(trt_2 - trt_1)$ correspond to the C_T of the treatment 2 minus the C_T of the treatment 1, for the target and control genes, respectively. If both amplification efficiencies take the maximum possible value ($E=2$), expression (1) becomes the familiar $2^{-\Delta\Delta CT}$ expression [9]. Moreover, almost any other mathematical expression or method available in the literature to calculate fold-change is a variant of Eq. (1). The differences among variants of Eq. (1) refer mainly to estimation of the efficiency either from a relative standard curve [5] or from individual amplification curves [7,13].

The methods based on expression (1) are mathematical equations devised to calculate fold-change between two samples. In some cases, these equations, however, lack the statistical formalism needed to draw valid inferences, especially when multiple levels of biological replicates from each experimental group are assayed [7,10,17]. Moreover, many ad-hoc approaches associated with formulas similar to Eq. (1) have been used with the objective of generating a set of "companion" *P*-values or standard errors [9,11]. However, few of them are valid in the presence of both biological and technical replication. Currently, the REST® software [17] is one of the few programs that implements a valid statistical analysis to test hypotheses and estimate the fold-changes using Eq. (1). However, such software is limited to the analysis of pairwise comparisons with respect to a control group, under a completely randomized design.

Even if a valid statistical test can be implemented for pairwise contrasts based on Eq. (1), comparing two treatments at a time in the context of a large experiment may be inefficient and lead to reduced power. In such case, a linear model could be used to analyze data from all treatment groups simultaneously in the same fashion of classical analysis of variance (ANOVA) techniques.

A linear mixed model [18] was recently proposed for the implementation of the so-called analytical method [10]. Such a model is potentially more flexible than the existing alternatives, but it makes the strong assumption that there is a common random effect for the control and test genes in each biological replicate. Assuming no gene-specific biological effects is not realistic, as it is expected that constitutively expressed genes will have more similar expression pattern across biological replicates than regulated genes. Violation of such assumption, in turn, may lead to wrong inferences. Consequently, there is a need for a formal statistical method for analysis of the relative quantification RT-PCR data that allows accommodation of more complex experimental designs (such as blocking factors) and testing of general hypotheses (including interactions, pairwise and group contrasts).

The objective of this paper is to present a novel, flexible method for analysis of relative quantification RT-PCR data using linear mixed models. The main advantage of the model is that it can be used to compute valid *P*-values associated with any general linear hypothesis of interest. Additionally, the model allows proper accounting of all sources of variation and it is expected to be more powerful than methods based on individual pairwise comparisons. In this paper we use a variety of approaches to validate the proposed methodology, to compare it with existing methods, and to illustrate its flexibility. First, our model is compared to other alternatives using a real dataset. Second, a model-free simulation based on the same dataset is used for comparative validation of the methodology. Lastly, several datasets are analyzed and different linear models are compared.

Results

Motivating example

Quantitative RT-PCR was used to study expression of the gene diazepam binding inhibitor (DBI) in the brain of piglets subject to weaning and social isolation treatments [19]. The experimental layout followed a randomized complete block design ($n=3$ litters) and the treatments consisted of a 2×2 factorial combination of weaning (early-weaned or non-weaned) and social isolation (isolated or control).

Preliminary assays indicated that *Sus scrofa* 18S ribosomal RNA (18S) was suitable for use as an endogenous control gene and that the amplification efficiency for primers of the two genes (18S and DBI) was close to two [19]. All reactions were performed in triplicate but some observations were excluded from the analysis because of evidence of non-specific amplifications (as revealed by dissociation curve analyses) [20]. The following model (denoted as Model I) was used for the analysis of the joint expression of DBI and 18S:

$$y_{gijkr} = TG^*_{gi} + l_{gj} + B_{gijk} + D_{ijk} + e_{gijkr},$$

where y_{gijkr} is the C_T obtained from the thermocycler software for the *g*th gene (18S or DBI) from the *r*th well, corresponding to the *k*th animal in the *j*th litter subjected to the *i*th treatment, TG^*_{gi} is the effect of treatment *i* in the expression of gene *g*, $l_{gj}\sim N(0,\sigma^2_{lg})$ is a gene-specific random effect of the *j*th litter, $B_{gijk}\sim N(0,\sigma^2_{Bg})$ is a gene-specific random effect of the *k*th piglet in the *j*th litter, $D_{ijk}\sim N(0,\sigma^2_D)$ is a random sample-specific effect (common to both genes), and $e_{gijkr}\sim N(0,\sigma^2_e)$ is a residual term. The sample-specific effect, D_{ijk}, captures differences among samples that are common to both genes, particularly those that affect total mRNA concentration, such as differential extraction or amplification efficiencies among samples. The treatments consisted of the combination of two factors, and the sub-index $i=1, 2, 3, 4$ corresponds to: early weaning + control (EWC), early weaning + isolation (EWI), non-weaning + control (NWC), and non-weaning + isolation (NWI), respectively.

Model I was fit to the data using the SAS mixed procedure [21] and a residual analysis was performed to check model assumptions. Tests of differential expression among groups were performed for the interaction of weaning by isolation and for pairwise treatment differences (simple effects). Point and interval estimates of fold-changes were approximated from the linear contrasts (in the log scale) by back transformation. The fold-changes were also estimated with the $2^{-\Delta\Delta CT}$ method [9] (ΔC_T) using a procedure presented in the original work [9].

In addition, an alternative linear model (denoted as Model II) was also used to analyze the data:

$$y_{gijkr} = TG^*_{gi} + l_j + D_{ijk} + \varepsilon_{gijkr}.$$

Model II is a simplified version of Model I, without the gene-specific sample and litter effects, and is equivalent to a previously published model for analysis of amplification curve data [18]. We anticipate that Model II is under parameterized and it may lead to wrong inferences.

Testing and estimating differential expression

Contrarily to the ΔCT procedure, Model I yielded a formal test for the interaction between isolation and weaning. There was no evidence of interaction effect between isolation and weaning on the expression of DBI ($P=0.829$), but there was a significant three-fold decrease in DBI gene expression due to isolation ($P=0.003$). As mentioned, the traditional analysis method (ΔCT) does not allow testing of this interaction, but it may be still used to estimate the fold-change of pairwise comparisons (Fig. 1).

While the estimates of fold-change were similar using Models I, II and ΔCT, the confidence intervals for the fold-changes based on ΔCT were wider than those based on Models I and II, and the general conclusions were not equivalent. For example, Models I and II indicated a significant decrease in the expression of DBI in response to social isolation in both early-weaned and non-weaned animals ($P=0.013$ and $P=0.019$ respectively from Model I), while ΔCT only detected the contrast EWI – EWC as significant ($P=0.03$). At a

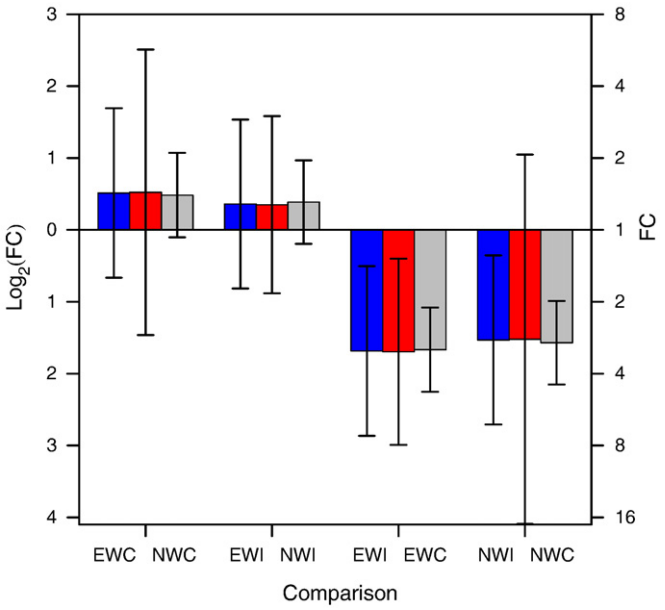

Fig. 1. Fold-change estimates. The $\log_2$ fold-changes for four contrasts (abscissa) are presented. Fold-change scale is included on the right axis. Segments indicate the 95% confidence interval. Comparisons whose confidence interval include the value 0 (1 in fold-change scale) are not significant at $\alpha = 5\%$. Bar color indicates the analysis method. ■: Model I, ■: ΔCT and ■: Model II.

significance level of $\alpha = 5\%$, Models I and II yielded the same conclusions, but the confidence intervals were narrower for Model II. Thus, depending on the significance level adopted, the conclusions might differ between these two approaches.

Validation through simulation using plasmode datasets

A simulation experiment was used to validate and compare alternative analysis methods. A fair simulation study, however, precludes the use of any of the analysis models as the data generation process. Alternatively, we permuted the real data and added constant values to generate a population of 1000 datasets with known fold-changes (Table 1), while keeping the original data structure, distribution and variability unchanged. The resulting dataset is called a plasmode [22] and similar techniques have been used to assess, for example, the validity of analysis methods in microarrays [23]. The plasmode datasets were analyzed using three methods: Models I, II, and ΔCT.

Table 1 summarizes the point estimates of the fold-change obtained for each pairwise contrast using the three analysis methods. All three methods yielded unbiased estimates of the fold-change, and they had roughly the same mean square error (MSE).

Confidence intervals (95%) for the fold-changes are presented on Table 2. The narrowest confidence intervals corresponded to Model II, followed by Model I and ΔCT. Nevertheless, a further analysis of the real coverage of these "nominal" 95% confidence intervals (values within parentheses in Table 2) revealed that Model II yielded intervals

Table 1
Average and mean squared error of the empirical distribution of point estimates of the fold-change derived from plasmode datasets.

Contrast	*Actual FC*	$FC^{a}_{[I]}$	$FC^{b}_{[\Delta CT2]}$	$FC^{c}_{[II]}$
EWI – EWC	2.000	2.059 (0.3922)	2.059 (0.3959)	2.054 (0.3720)
EWI – NWI	1.414	1.442 (0.1588)	1.443 (01602)	1.440 (0.1503)
EWC – NWC	0.707	0.744 (0.0470)	0.744 (0.0474)	0.742 (0.0443)
NWI – NWC	1.000	1.054 (0.0925)	1.054 (0.0937)	1.051 (0.0875)

Columns present the mean estimates with their mean squared errors (in parenthesis) from 1000 simulations. [a]Model I; [b]ΔCT; and [c]Model II.

Table 2
Width and coverage (in parenthesis) of nominal 95% confidence intervals of fold-change derived from plasmode datasets.

Contrast	$CI_{[I]}{}^{a}$	$CI_{[\Delta CT]}{}^{b}$	$CI_{[II]}{}^{c}$
EWI – EWC	3.03 (93.7%)	6.03 (91.0%)	1.63 (79.8%)
EWI – NWI	2.12 (95.4%)	4.34 (94.1%)	1.14 (84.2%)
EWC – NWC	1.09 (94.7%)	2.24 (93.0%)	0.58 (81.4%)
NWI – NWC	1.55 (94.7%)	3.22 (92.9%)	0.83 (82.0%)

Each column presents the mean width of the 95% confidence interval for the fold-change with the actual coverage of the interval in parenthesis. Desirable properties are coverage close to 95% and small interval width. [a]Model I; [b]ΔCT and [c]Model II.

with significantly less coverage than the other two methods. The real coverage of confidence intervals obtained from Model II was well below the nominal 95% confidence level. Confidence intervals calculated from Model I exhibited the closest coverage to the nominal level. The coverage obtained from ΔCT-derived confidence intervals was also close to 95%, but the width of the confidence intervals was sensibly larger.

Models I and II allowed testing general linear hypotheses related to the interaction and main effects (including pairwise comparisons of treatments). Conversely, ΔCT allowed only pairwise contrasts between pairs of treatments (Table 3).

Under the null hypothesis (contrast NWI – NWC in Table 3), Model I and ΔCT yielded a type I error rate very close to the nominal 5% test value, and the discrepancies observed in Table 3 for the ΔCT method were within the expected simulation error (based on extensive simulations not shown in this paper). The realized type I error rate of tests from Model II was clearly above the nominal error level (0.07 for $\alpha = 0.01$, and 0.18 for $\alpha = 0.05$).

Under the alternative hypothesis (Table 3, all comparisons except NWI – NWC), Model II showed the highest probability of declaring significant a fold-change larger than 1, but part of this apparent power comes from an inflated type I error rate as shown before. Model I was more powerful than ΔCT and, in absolute terms, the increase in power was more evident for larger fold-changes. This indicates that Model I is the best model for analysis because it provides adequate control of type I error rate, but considerably more power than ΔCT.

Model checking in experimental datasets

Although Model I is suitable only for the analysis of the described dataset, an equivalent model can be elicited for any specific data structure or design layout. The main components of Model I are the random sample effects and the random interaction between sample and gene factors. Moreover, gene-specific variances are assumed for the sample-gene interaction. The measurement error term (residual effects) is assumed homoskedastic with respect to genes.

To assess the adequacy of these assumptions in a broader set of experimental data [19,24–28], six different datasets where analyzed. Details of the datasets are presented in Table 4. The datasets included

Table 3
Proportion of rejected hypotheses estimated from plasmode datasets.

Contrast	$\alpha = 1\%$			$\alpha = 5\%$		
	$p^{a}_{[I]}$	$p^{b}_{[\Delta CT]}$	$p^{c}_{[II]}$	$p^{a}_{[I]}$	$P^{b}_{[\Delta CT]}$	$p^{c}_{[II]}$
EWI – EWC	0.157	0.058	0.700	0.439	0.257	0.872
EWI – NWI	0.030	0.026	0.246	0.131	0.098	0.415
EWC – NWC	0.031	0.026	0.276	0.152	0.104	0.429
NWI – NWC	0.010	0.029	0.072	0.053	0.071	0.180
Wean. × isol.	0.075	–	0.416	0.249	–	0.584

p: proportion of rejected tests from the 1000 simulated datasets in each contrast at two significance levels (α). The NWI–NWC contrast corresponded to the null hypothesis (no differential expression) and the expected value is $p = \alpha$. For contrasts different from NWI–NWC, a larger value of *p* implies more power. The sub-index indicates the analysis method. [a]Model I; [b]ΔCT and [c]Model II.

Table 4
Description of experimental datasets used for model checking and result comparison.

Dataset name[a]	Published reference	Number of genes	Experimental design	Biological samples	Assay replicates
SPLIT	[25]	2	Split plot design. Main plot factor: disease status. Subplot factor: infection	12	2
TC	[30]	5	Longitudinal time course experiment. Timepoints: 0, 2, 4, 8, 16 h after infection.	4	2
PFC	[19]	5	2×2 Factorial in a randomized complete block design. Factors: weaning; isolation	12	3
MRD	[27]	2	Completely randomized design with two groups.	8	3
TLD	[24]	64	Completely randomized design with two groups.	9	4
SHK	[28]	6	Completely randomized design with one group.	80	1

[a] Abbreviations used in the text to refer to each dataset.

one to 63 test genes and four to 80 biological samples. All datasets but one included technical replicates (assay replicates). Twelve alternative models were compared using the Akaike information criteria (AIC) and the Bayesian information criteria (BIC) [29]. The models represented the combinations of different assumptions: sample-specific random effect (included or not), sample-gene random interaction (homoskedastic, heteroskedastic, or not included in the model), and residual variance (homogeneous or heterogeneous across genes).

The effects included in the best-fit model for each dataset are shown in Table 5. A random sample effect was present in all models. Similarly, the gene by sample interaction with heterogeneous variances among genes was selected for almost all datasets. The only exception was the MRD [27] dataset where the model with homogeneous variances was preferred. Gene-specific residual variances were also generally favored by the model selection criteria. In the TLD [24] dataset, heterogeneous residual variances could not be fit due to convergence problems. In the SHK [28] dataset, the residual term included both the residual and sample-gene interaction effects of the other models because this dataset lacks technical replicates. Consequently, the model with heterogeneous residuals indicates the presence of a gene by sample interaction with heterogeneous variances, heterogeneity of variance in the measurement errors, or both. In the remaining datasets, the heterogeneity of residual variances was caused by different (gene-specific) precisions for measurement of gene expression. Disentangling the sources of such heterogeneity is beyond the scope of this paper, but we anticipate that differential amplification efficiencies may be one of such causes.

Table 5
Effects included in best-fit models corresponding to different datasets.

Dataset	Fixed effects[a]	Random effects		
		Sample[b]	Sample-gene[c]	Residual[d]
SPLIT	Gene * Status* infection	Yes	Yes. Gene-specific variance	Gene-specific variance
TC	Gene*Time	Yes	Yes. Gene-specific variance	Gene-specific variance
PFC	Gene * Weaning * Isolation	Yes	Yes. Gene-specific variance	Gene-specific variance
MRD	Gene*strain	Yes	No/Yes[e]	Gene-specific variance
TLD	Gene*mutation	Yes	Yes. Gene-specific variance	Homogeneous variance
SHK	Gene	Yes	Gene-specific variance	

[a] Specification for fixed effects.
[b] Random sample effect included (yes) or not (no).
[c] Random sample by gene interaction included (yes/no) and variance of the effect (Homogeneous or Gene-specific).
[d] Specification of the residual variance (Homogeneous or Gene-specific).
[e] The model without sample-gene random effect and the model with a sample-gene interaction (with homogeneous variances) yielded the same values for both selection criteria.

Table 6
Contrasts of interest in each dataset.

Dataset	Comparisons tested with ΔCT and LMM[a]	Comparisons tested with LMM only[b]
SPLIT	Simple effect of Infection Simple effect of status	Interaction infection by status
TC	Every time versus baseline	Linear and quadratic trend
PFC	Simple effect of weaning Simple effect of isolation	Interaction weaning by isolation
Mrd	Wild type versus mutant	–
TLD	Mutated versus unmutated	–

[a] Contrasts that can be obtained with both methodologies.
[b] Contrasts that are calculated with the linear model only.

Except for a few subtle differences among the models selected for each dataset, the general model including sample and gene-sample random effects was always preferred. The inclusion of heterogeneous residual variances had only a marginal effect on the tests for differential expression (results not shown). Contrarily, omitting sample-gene effects from the models increased the type I error rates over the nominal value (as shown in the previous sections).

Linear mixed models versus ΔCT to test general hypothesis

Linear mixed models allowed testing general contrasts (Table 6). In experiments consisting of two treatments in a completely randomized design [24,27] the contrast of interest could be tested using either ΔCT or linear mixed models. However, in datasets involving more complex treatment structures and sampling schemes, only the linear mixed models allowed testing complex hypotheses. For example, through the use of a linear mixed model, we could test for interactions in factorial experiments [19,25] and for trend contrasts in a longitudinal study.

Discussion

Linear mixed models versus comparative C_T analysis methods

In this paper we presented a novel method (Model I) for the analysis of relative quantification qRT-PCR data. Our approach consists of the analysis of (raw or efficiency corrected) C_T values for a target and a control gene using a general linear mixed model. Currently, the use of qRT-PCR is pervasive in functional genomics studies and the complexity of experimental designs or sampling schemes have increased considerably [30–32]. However, the statistical and mathematical approaches available for analysis of such data lack flexibility and statistical properties needed to produce useful and valid inferences in complex experimental layouts [33].

A general advantage of using our linear mixed model method is that it is possible to test any general linear hypothesis; for instance, in the first real data example presented in this paper, we could test the hypothesis of interaction between social isolation and early weaning in the expression of DBI in the brain of the piglets. Contrarily, the traditional analysis method (comparative C_T) could not test the same hypothesis and its application was restricted to pairwise comparisons of treatments. With other datasets, we used linear mixed models to test for linear and quadratic trends in time course experiments and for interactions and main effects in a factorial experiment within a split-plot design. These contrasts could not be implemented correctly using the comparative C_T methodology.

Our linear mixed model is more flexible than other methodologies proposed in the literature for the analysis of qRT-PCR data. Yuan et al. [33] proposed different linear models for the analysis of qRT-PCR data with only one level of replication. But in practice, most qRT-PCR

experiments include both technical and biological replicates. Even when technical replicates can be collapsed by taking their average, a hierarchical model may perform better in certain cases [34]. Fu et al. [34] proposed a generalized estimating equation model that can fit data with both technical and biological replication, but they assumed that the technical replicates of control and test genes can be paired. In practice this is seldom the case. Conversely, our method is more flexible and allows the incorporation of an arbitrarily complex experimental protocol in both treatment structure (factorial, time courses, etc.) and sampling scheme (blocks, split-plots, etc.). Moreover, our linear mixed model is suitable for any pattern of technical and biological replication.

Relative performance for pairwise contrasts

For pair contrasts, our method (Model I), as well as Model II and ΔCT approaches produced similar estimates of fold-change. For hypothesis testing and interval estimation the methods yielded divergent results: Model I and ΔCT outperformed Model II in terms of controlling the type I error rate at the nominal level. Comparing the simulation results from Model I and ΔCT, it is evident that the simultaneous analysis of all groups with a linear model (Model I) provides more power than independent pairwise comparisons (ΔCT). This is not surprising because a pair contrast within Model I had six degrees of freedom while the paired *t*-test associated with ΔCT had only two degrees of freedom. Moreover, the advantage of the mixed model methodology over the ΔCT could be greater if more treatments were included. For instance, in a completely randomized design with 10 treatments or groups and three biological replicates in each, any *t*-test between a pair of treatments will have four degrees of freedom, while an ANOVA-based *F*-test will have 20 degrees of freedom. On the other hand, if a certain experiment is restricted to two treatments or groups, both methods (Model I and ΔCT) will yield identical results.

Importance of correct model elicitation

An important result from this paper is the importance of properly modelling all sources of variation in order to draw valid inferences. Our model assumes a Gaussian distribution of the log expression. It also assumes heterogeneous variances in expression of target and control genes, and the presence of sample-specific effects related to the measurement protocol. A priori, all these assumptions are plausible. The assumption of normally distributed log-expression levels has been extensively used [28,35,36]. Also, the heteroskedastic models for analysis of several candidate control genes presented better fit than homoskedastic alternatives [28]. Finally, it is sensible to include a sample-specific effect (common to both genes) that represents the total mRNA level in the sample. The inclusion of such effect is a unique feature of our modelling approach.

The aforementioned assumptions were reasonable in this experiment; however, we also conducted a simulation study to evaluate the performance of our model. We avoided the use of a parametric model for the simulation by creating a population of datasets (plasmode) using re-sampling methods. In other words, neither the original distribution of the C_T values nor the relative technical and biological variabilities were altered. From this simulation, Model I emerged as the best model for the analysis, showing a correct type I error rate and confidence intervals coverage. Contrarily, Model II overstated the significance of the comparisons and the coverage of the confidence intervals. While these results are specific for these data, we believe that the inclusion of a gene-sample-specific effect separated from a sample-specific effect (as in Model I) is more plausible than solely the inclusion of a sample effect common to both genes (as it is implicit in Model II).

We validated these assumptions by analyzing a set of experimental data that included different number of genes and biological replications. The importance of including sample and gene-specific random effects was confirmed by model selection in every dataset. Additionally, we found heterogeneous residual variances in most of the datasets.

In summary, we have shown the importance of proper modelling of qRT-PCR data to correctly control the type I error, and we have provided a general linear model framework for such analyses. The most important feature of our modelling approach is the use of (raw or efficiency corrected) C_T data as response variables to conduct a joint analysis of target and control gene expressions, modelling simultaneously the biological and technical variation. Furthermore, Model I represents a set-up of the general linear mixed model approach (the most appropriate for our real data example), but as shown, it can be easily expanded to fit data from other designs. Finally, our method is more accurate, powerful and flexible than existing alternatives for qRT-PCR data analysis, and it is especially useful in studies involving more than two treatments or time points and multiple experimental factors.

Methods

Materials and RT-PCR reactions

Sample collection, mRNA extraction, cDNA synthesis and PCR protocols for the motivating example are described in detail in the original paper [19].

Model derivation

We assume that the expression z_{gijk} (copy number or concentration of mRNA) of gene g in sample k of litter (block) j and experimental group i can be described by:

$$-log\left(z_{gijk}\right) = \mathrm{TG}_{gi} + l_{gj} + B_{gijk}, \tag{2}$$

where TG_{ig} is the effect of the i^{th} treatment on the expression of gene g, l_{gj} is the random effect of litter on each gene [$l_{gj} \sim N(0, \sigma^2_{lg})$] and B_{gijk} is the gene and sample-specific effect [$B_{gijk} \sim \mathrm{N}(0,\sigma^2_{Bg})$].

If mRNA is isolated, cDNA is synthesized and qRT-PCR is conduced in several independent wells for each sample with primers for each of the genes, the generated data may be analyzed with the following model:

$$y_{gijkr} = \mathrm{TG}_{gi} + l_{gj} + B_{gijk} + D_{ijk} + e_{gijkr}, \tag{3}$$

where y_{gijkr} is a measured expression level in the log scale (for example: C_T), D_{ijk} is a sample-specific effect introduced by the experimental protocol, and e_{gijkr} is a well-specific measurement error.

In the model above (Eq. (3)), D_{ijk} represents a measurement artefact that is sample-specific, and it is assumed to be $D_{ijk} \sim N(d_i,\sigma^2_D)$. This implies that the experimental protocol affects the measurement on the sample for all assayed genes in the same way, but it may generate a treatment bias (d_i). These assumptions (apart from the specific Gaussian distribution) are standard in relative quantification analyses. Moreover, the existence of the D_{ijk} effects is supposedly the reason to include a control gene in such assays.

If we assume that the $TG_{gi} = \mu + \tau_{gi}$ and that $\tau_{gi} = 0$ for the control gene, and we fit this model assuming $D_{ijk} \sim N(0,\sigma^2_D)$ (i.e. Model I), the following values for the TG^*_{gi} effects are expected:

$$\begin{cases} \mathrm{TG}^*_{gi} = \mu + \tau_{gi} + d_i, & \text{for } g = \text{target} \\ \mathrm{TG}^*_{gi} = \mu + d_i, & \text{for } g = \text{control} \end{cases} \tag{4}$$

Hypothesis testing and estimation

Suppose that the interest is to estimate the fold-change between EWI ($i = 2$) and EWC ($i = 1$) for the target gene ($g = 2$) normalized to

the control gene ($g=1$). This is equivalent to estimate the log-difference (or log of the fold-change) using:

$$\mathrm{dif}_{(\mathrm{EWI}-\mathrm{EWC})} = (\mathrm{TG}^*_{22} - TG^*_{21}) - (TG^*_{12} - TG^*_{11}). \quad (5)$$

It is clear that if Eq. (4) holds (i.e., if there is no differential expression of the control gene), the expectation of Eq. (5) is:

$$E\left[\mathrm{dif}_{(\mathrm{EWI}-\mathrm{EWC})}\right] = \mathrm{TG}_{22} + d_2 - \mathrm{TG}_{21} - d_1 - (d_2 - d_1) = \mathrm{TG}_{22} - \mathrm{TG}_{21}, \quad (6)$$

which is the quantity of interest. Furthermore, point estimates, hypothesis tests and confidence intervals of Eq. (5) are readily available, and the fold-change estimates may be approximated by transforming point estimates and confidence interval limits to the original scale. For example, if the data $\boldsymbol{z}$ are C_T values, the fold-change estimation formulae would be:

$$\mathrm{FC}_{(\mathrm{EWI}-\mathrm{EWC})} = 2^{-\mathrm{diff}_{(\mathrm{EWI}-\mathrm{EWC})}}. \quad (7)$$

The general linear contrast approach presented here is very flexible and multiple control genes can be easily incorporated into this testing scheme. For example, if the average of two control genes ($g=1,2$) is used to normalize the expression of the test gene ($g=3$), the model is fit to the expression of the three genes simultaneously and the linear contrast in Eq. (5) is replaced with $(TG^*_{32} - 0.5(TG^*_{22} + TG^*_{12})) - (TG^*_{31} - 0.5(TG^*_{21} + TG^*_{11}))$.

Data and response variable

The response variable in the model given by Eq. (3) may be any measure proportional to the log-mRNA concentration in the samples. In our particular example, the amplification efficiency was close to the optimal value ($E=2$), and consequently the C_T values constituted a suitable response. The Supplementary material includes a detailed explanation of an alternative response variable when the amplification efficiency is smaller than two.

ΔCT procedure

The ΔCT method of this paper used Eq. (1) assuming $E=2$. The general procedure was proposed in the original work of Livak and Schmittgen [9], and we adapted such protocol to the case of littermates as follows.

Suppose again that the interest is to estimate the fold-change between EWI ($i=2$) and EWC ($i=1$) for the target gene ($g=2$) normalized to the control gene ($g=1$) using ΔCT methods. Let $\overline{C}_{\mathrm{T}_{gijk}}$ be the average C_T value across technical replicates from the quantification of gene g in samples k of litter j and experimental group i. The corrected log-ratio ($-\Delta\Delta$CT) comparing EWI to EWC in a pair of littermates (indexed by jk and jk') is defined as:

$$-\Delta\Delta\mathrm{CT}_{\mathrm{EWI}(k):\mathrm{EWI}(k)} = -\left[\left(\overline{C}_{\mathrm{T}_{22jk}} - \overline{C}_{\mathrm{T}_{21jk}}\right) - \left(\overline{C}_{\mathrm{T}_{12jk'}} - \overline{C}_{\mathrm{T}_{11jk'}}\right)\right]. \quad (8)$$

Averaging the result of Eq. (8) over litters yields $-\Delta\Delta\mathrm{CT}_{\mathrm{EWI:EWC}}$ and its standard error ($\mathrm{SE}_{\mathrm{EWI:EWC}}$). A test statistic is computed as follows:

$$T = \frac{-\Delta\Delta\mathrm{CT}_{\mathrm{EWI:EWC}}}{\mathrm{SE}_{\mathrm{EWI:EWC}}}. \quad (9)$$

Eq. (10) can be used to test the hypothesis of differential expression by comparing t to the quantiles of a Student-t distribution with 2 degrees of freedom (number of litters minus one). To construct a 95% confidence interval of the log-ratio between EWI and EWC we used the following expression:

$$\mathrm{CI}_{\log-\mathrm{ratio}(\mathrm{EWI:EWC})} = -\Delta\Delta\mathrm{CT}_{\mathrm{EWI:EWC}} \pm t_{(2,0.975)} \cdot SE_{\mathrm{EWI:EWC}}, \quad (10)$$

where $t_{(2,0.975)}$ is the 97.5 quantile of the Student-t distribution with 2 degrees of freedom. A 95% confidence interval of fold-change is obtained by back-transforming the result of Eq. (10) in similar fashion to Eq. (7).

Programs for analysis

The models implemented in this paper can be readily fit using mixed model software. As an illustration, SAS code to fit the motivating example is presented in Fig. 2. In the Supplementary material, we present a detailed tutorial on how to generalize this code to other experimental situations. We also provide a SAS macro that can generate the Proc Mixed statements needed for some of the most common experimental designs encountered in the qPCR literature. The macro is publicly available at https://www.msu.edu/~steibelj/JP_files/QPCR.html.

Simulation study

From the expression data (raw C_T) from the target (DBI) and the control (18S) genes, we computed the arithmetic mean of each treatment (averaging out all the available biological and technical replicates), subtracted the corresponding average treatment value from each individual observation and added the general mean. The result of this procedure is a dataset that keeps the original variability among litters and among technical replicates, but that has a common mean for all treatments. Subsequently, the data were reshuffled to create 1000 datasets. Within each litter, the treatment memberships were permuted among the four treatments, but the technical replicates were kept together. Then, the observations corresponding to EWC animals were increased by the value 0.5 and the observations

Basic Code:	Details:
1 **proc mixed** data=ratio; 2 class gene trt sample litter; 3 model y=gene|trt; 4 random sample; 5 random litter litter*trt/group=gene; **6 run;**	1) proc mixed invoked to fit model. 2) gene, treatment, sample and litter are factors used in the model. 3) Fixed effects are: treatment, gene and their interaction (these components expand TG in model I). 4) Random effect of sample (D in model I). 5) Random effects of litter (l in model I) and litter*treatment (B in model I) are fit with gene specific variances (group statement).

Fig. 2. Code for implementation of mixed model analysis with model I. SAS proc mixed code (left) with an explanation of each line of code (right).

corresponding to EWI animals were decreased by the value 0.5. Consequently, the resulting population of trials had roughly the same biological and technical variability of the original data, but a known fold-change for each treatment pair (second column in Table 1).

The analysis of this population of datasets provided a set of 1000 *P*-values for each of the hypothesis tested. The type I error rate was estimated by the rate of rejections (for certain nominal α) in the comparison EWC–NWC. Conversely, power was estimated counting the number of rejections in any non-null hypothesis. The coverage of the confidence intervals was estimated from the proportion of intervals that contained the true fold-change value.

Model selection

A detailed model description used for each dataset of Table 4 for model selection and results comparison is presented in the Supplementary material.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ygeno.2009.04.008.

References

[1] T. Nolan, R.E. Hands, S.A. Bustin, Quantification of mRNA using real-time RT-PCR, Nat. Protoc. 1 (2006) 1559–1582.
[2] S.A. Bustin, Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays, J. Mol. Endocrinol. 25 (2000) 169–193.
[3] A. Giulietti, L. Overbergh, D. Valckx, B. Decallonne, R. Bouillon, C. Mathieu, An overview of real-time quantitative PCR: applications to quantify cytokine gene expression, Methods 25 (2001) 386–401.
[4] J. Winer, C.K. Jung, I. Shackel, P.M. Williams, Development and validation of real-time quantitative reverse transcriptase-polymerase chain reaction for monitoring gene expression in cardiac myocytes in vitro, Anal. Biochem. 270 (1999) 41–49.
[5] M.R. Johnson, K. Wang, J.B. Smith, M.J. Heslin, R.B. Diasio, Quantitation of dihydropyrimidine dehydrogenase expression by real-time reverse transcription polymerase chain reaction, Anal. Biochem. 278 (2000) 175–184.
[6] M.W. Pfaffl, A new mathematical model for relative quantification in real-time RT-PCR, Nucleic Acids Res. 29 (2001) e45.
[7] A. Gentle, F. Anastasopoulos, N.A. McBrien, High-resolution semi-quantitative real-time PCR without the use of a standard curve, Biotechniques 31 (2001) 502–508.
[8] W. Liu, D.A. Saint, A new quantitative method of real time reverse transcription polymerase chain reaction assay based on simulation of polymerase chain reaction kinetics, Anal. Biochem. 302 (2002) 52–59.
[9] K.J. Livak, T.D. Schmittgen, Analysis of relative gene expression data using real-time quantitative PCR and the 2(− Delta Delta C(T)) Method, Methods 25 (2001) 402–408.
[10] J.H. Marino, P. Cook, K.S. Miller, Accurate and statistically verified quantification of relative mRNA abundances using SYBR Green I and real-time RT-PCR, J. Immunol. Methods 283 (2003) 291–306.
[11] P.Y. Muller, H. Janovjak, A.R. Miserez, Z. Dobbie, Processing of gene expression data generated by quantitative real-time RT-PCR, Biotechniques 32 (2002) 1372–1374, 1376, 1378–1379.
[12] C. Ramakers, J.M. Ruijter, R.H. Deprez, A.F. Moorman, Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data, Neurosci. Lett. 339 (2003) 62–66.
[13] S. Swillens, J.C. Goffard, Y. Marechal, A. de Kerchove d'Exaerde, H. El Housni, Instant evaluation of the absolute initial number of cDNA copies from a single real-time PCR curve, Nucleic Acids Res. 32 (2004) e56.
[14] A. Tichopad, A. Didier, M.W. Pfaffl, Inhibition of real-time RT-PCR quantification due to tissue-specific contaminants, Mol. Cell Probes 18 (2004) 45–50.
[15] A. Tichopad, M. Dilger, G. Schwarz, M.W. Pfaffl, Standardized determination of real-time PCR efficiency from a single reaction set-up, Nucleic Acids Res. 31 (2003) e122.
[16] A. Tichopad, A. Dzidic, M.W. Pfaffl, Improving quantitative real-time RT-PCR reproducibility by boosting primer-linked amplification efficiency, Biotechnol. Lett. 24 (2002) 2053–2056.
[17] M.W. Pfaffl, G.W. Horgan, L. Dempfle, Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR, Nucleic Acids Res. 30 (2002) e36.
[18] P. Cook, C. Fu, M. Hickey, E.S. Han, K.S. Miller, SAS programs for real-time RT-PCR having multiple independent samples, Biotechniques 37 (2004) 990–995.
[19] R. Poletto, J.M. Siegford, J.P. Steibel, P.M. Coussens, A.J. Zanella, Investigation of changes in global gene expression in the frontal cortex of early-weaned and socially isolated piglets using microarray and quantitative real-time RT-PCR, Brain Res. 1068 (2006) 7–15.
[20] S. Giglio, P.T. Monis, C.P. Saint, Demonstration of preferential binding of SYBR Green I to specific DNA fragments in real-time multiplex PCR, Nucleic Acids Res. 31 (2003) e136.
[21] R.C. Littell, G.A. Milliken, W.W. Stroup, R.D. Wolfinger, SAS for Mixed Models, SAS Institute, Inc., Cary, NC, 2006.
[22] D.B. Allison, X. Cui, G.P. Page, M. Sabripour, Microarray data analysis: from disarray to consolidation and consensus, Nat. Rev. Genet. 7 (2006) 55–65.
[23] G.L. Gadbury, Q. Xiang, L. Yang, S. Barnes, G.P. Page, D.B. Allison, Evaluating statistical methods using plasmode data sets in the age of massive public databases: an illustration using false discovery rates, PLoS Genet. 4 (2008) e1000098.
[24] L.V. Abruzzo, K.Y. Lee, A. Fuller, A. Silverman, M.J. Keating, L.J. Medeiros, K.R. Coombes, Validation of oligonucleotide microarray data using microfluidic low-density arrays: a new statistical method to normalize real-time RT-PCR data, Biotechniques 38 (2005) 785–792.
[25] P.M. Coussens, C.J. Colvin, G.J. Rosa, J. Perez Laspiur, M.D. Elftman, Evidence for a novel gene expression program in peripheral blood mononuclear cells from *Mycobacterium avium* subsp. paratuberculosis-infected cattle, Infect. Immun. 71 (2003) 6487–6498.
[26] P.M. Coussens, A. Jeffers, C. Colvin, Rapid and transient activation of gene expression in peripheral blood mononuclear cells from Johne's disease positive cows exposed to *Mycobacterium paratuberculosis* in vitro, Microb. Pathog. 36 (2004) 93–108.
[27] S.N. Peirson, J.N. Butler, R.G. Foster, Experimental validation of novel and conventional approaches to quantitative real-time PCR data analysis, Nucleic Acids Res. 31 (2003) e73.
[28] A. Szabo, C.M. Perou, M. Karaca, L. Perreard, J.F. Quackenbush, P.S. Bernard, Statistical modeling for selecting housekeeper genes, Genome Biol. 5 (2004) R59.
[29] G. Schwarz, Estimating dimension of a model, Ann. Stat. 6 (1978) 461–464.
[30] H. Hamalainen, H. Zhou, W. Chou, H. Hashizume, R. Heller, R. Lahesmaa, Distinct gene expression profiles of human type 1 and type 2 T helper cells, Genome Biol. 2 (2001) R022.
[31] M.J. Martinez, S. Roy, A.B. Archuletta, P.D. Wentzell, S.S. Anna-Arriola, A.L. Rodriguez, A.D. Aragon, G.A. Quinones, C. Allen, M. Werner-Washburne, Genomic analysis of stationary-phase and exit in *Saccharomyces cerevisiae*: gene expression and identification of novel essential genes, Mol. Biol. Cell 15 (2004) 5295–5305.
[32] A. Recinos III, B.K. Carr, D.B. Bartos, I. Boldogh, J.R. Carmical, L.M. Belalcazar, A.R. Brasier, Liver gene expression associated with diet and lesion development in atherosclerosis-prone mice: induction of components of alternative complement pathway, Physiol. Genomics 19 (2004) 131–142.
[33] J.S. Yuan, A. Reed, F. Chen, C.N. Stewart Jr., Statistical analysis of real-time PCR data, BMC Bioinformatics 7 (2006) 85.
[34] W.J.J. Fu, J.B. Hu, T. Spencer, R. Carroll, G.Y. Wu, Statistical models in assessing fold change of gene expression in real-time RT-PCR experiments, Comput. Biol. Chem. 30 (2006) 21–26.
[35] C.L. Andersen, J.L. Jensen, T.F. Orntoft, Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets, Cancer Res. 64 (2004) 5245–5250.
[36] A.M. Brunner, I.A. Yakovlev, S.H. Strauss, Validating internal controls for quantitative plant gene expression studies, BMC Plant Biol. 4 (2004) 14.