J. Dairy Sci. 94:4950–4961
doi:10.3168/jds.2011-4147

A simple analytical and experimental procedure for selection of reference genes for reverse-transcription quantitative PCR normalization data

R. Manjarin,* N. L. Trottier,* P. S. Weber,† J. S. Liesman,* N. P. Taylor,* and J. P. Steibel*‡[1]
*Department of Animal Science,
†Department of Large Animal Clinical Sciences, and
‡Department of Fisheries and Wildlife, Michigan State University, East Lansing 48824

ABSTRACT

Variation in cellular activity in a tissue induces changes in RNA concentration, which affects the validity of gene mRNA abundance analyzed by reverse transcription quantitative PCR (RT-qPCR). A common way of accounting for such variation consists of the use of reference genes for normalization. Programs such as geNorm may be used to select suitable reference genes, although a large set of genes that are not coregulated must be analyzed to obtain accurate results. The objective of this study was to propose an alternative experimental and analytical protocol to assess the invariance of reference genes in porcine mammary tissue using mammary RNA and DNA concentrations as correction factors. Mammary glands were biopsied from 4 sows on d 110 of gestation (prepartum), on d 5 (early) and 17 (peak) of lactation, and on d 5 after weaning (postweaning). Relative expression of 7 potential reference genes, *API5*, *MRPL39*, *VAPB*, *ACTB*, *GAPDH*, *RPS23*, and *MTG1*, and one candidate gene, *SLC7A1*, was quantified by RT-qPCR using a relative standard curve approach. Variation in gene expression levels, measured as cycles to threshold at each stage of mammary physiological activity, was tested using a linear mixed model fitting RNA and DNA concentrations as covariates. Results were compared with those obtained with geNorm analysis, and genes selected by each method were used to normalize *SLC7A1*. Quantified relative mRNA abundance of *GAPDH* and *MRPL39* remained unchanged across stages of mammary physiological activity after accounting for changes in tissue RNA and DNA concentration. In contrast, geNorm analysis selected *MTG1*, *MRPL39*, and *VAPB* as the best reference genes. However, when target gene *SLC7A1* was normalized with genes selected either based on our proposed protocol or by geNorm, fold changes in mRNA abundance did not differ. In conclusion, the proposed analytical protocol assesses expression invariance of potential reference genes by accounting for variation in tissue RNA and DNA concentrations and thus represents an alternative method to select suitable reference genes for RT-qPCR analysis.
Key words: geNorm, normalization, quantitative PCR, sow

INTRODUCTION

Reverse transcription-quantitative PCR (**RT-qPCR**) is the most common and reliable method for rapid quantification of mRNA (Pfaffl, 2001; Wong and Medrano, 2005). However, variations in RNA extraction yield, reverse-transcription yield, loading cDNA, and efficiency of RT-qPCR amplification make the analysis prone to errors (Bustin et al., 2009, 2010). The reliability of the RT-qPCR analysis can be improved by amplifying, simultaneously with the target gene, a reference gene that serves as an internal control against which other mRNA values can be normalized (Bustin, 2000, 2002). However, the identification of a reference gene whose mRNA copy number per cell remains constant across treatments or samples represents a challenge, as the expression of all genes is regulated during the cell cycle to control diverse cellular functions. A feasible method for selection of suitable reference genes is to statistically test for differences in relative gene mRNA abundance between treatments. However, testing differential expression of the reference gene requires quantification of mRNA abundance of the gene in question, which in itself poses a circular problem (Pfaffl et al., 2004). The search for a suitable reference gene becomes more difficult when the expression of target genes is compared between tissues with different cell growth and proliferation, such as the lactating mammary gland (Tucker, 1987; Capuco et al., 2001). In bovine and porcine mammary tissue, Bionaz and Loor (2007) and Tramontana et al. (2008) have shown that an increase in total cellular RNA between pregnancy and lactation caused an artifactual dilution of mRNA across all mammary genes tested, including that of reference genes. As a result, the same authors proposed the use of geNorm software (http://medgen.

Received January 4, 2011.
Accepted June 8, 2011.
[1]Corresponding author: steibelj@msu.edu

ugent.be/~jvdesomp/genorm/; Vandesompele et al., 2002) to perform pairwise ratio-stability comparisons between multiple unrelated genes for selection of the best reference genes. However, such statistical method relies on the assumption that the genes tested are not co-regulated (Vandesompele et al., 2002; Derveaux et al., 2010). An alternative approach would be to correct gene expression by accounting for the variation in mRNA abundance resulting from total cellular RNA changes. However, expression of reference gene mRNA abundance relative to an increasing amount of RNA would be subject to dilution, as previously acknowledged by Bustin (2000, 2002). Therefore, the objectives of the present study were (1) to propose an analytical protocol to assess invariance of reference genes by accounting for the artifactual dilution effect on gene expression, and (2) to validate the proposed protocol by comparing expression results of a target gene normalized using the reference genes selected either with the proposal analytical protocol or with geNorm.

MATERIALS AND METHODS

Animal and Tissue Collection

The study was performed in accordance with the Institutional Animal Care and Use Committee at Michigan State University. Four sows (all parity 5, Landrace × Yorkshire) were selected at 107 d of pregnancy and moved to farrowing crates in a room maintained at 20°C. Animals were fed a corn and soybean meal-based diet that met all nutrient requirements for lactating sows nursing 10 piglets with a predicted piglet ADG of 250 g (NRC, 1998). Sows were fed a maximum of 5 kg/d to ensure equal DM intake. Litters were equalized to 10 piglets within 48 h after farrowing and piglets weaned at d 21 of lactation. Mammary parenchymal tissue was biopsied in the morning, following a 12-h overnight fast, from the first and second thoracic glands according to the procedure described by Kirkwood et al. (2007). Mammary tissue was sampled at 4 different physiological stages: d 110 of gestation (prepartum), d 5 (early) and d 17 (peak) of lactation, and d 5 after weaning (postweaning). Mammary tissue was flash-frozen in liquid N_2 and stored at −80°C. For mammary tissue collection, piglets were isolated in an adjacent pen equipped with a heat lamp. Three hours after biopsy, piglets were returned to sows and allowed to nurse.

Sample Analysis

RNA Extraction and cDNA Synthesis. Ribonucleic acid was extracted from mammary tissue using the PerfectPure RNA Cell and Tissue Kit according to the manufacturer's instructions (5 PRIME, Gaithersburg, MD). Isolated RNA was tested for purity by spectrophotometry (NanoDrop 1000, Thermo Scientific, Wilmington, DE) and for quantity and integrity using the Agilent Bioanalyzer 2100 with the RNA 6000 Nano Labchip (Agilent Technologies, Palo Alto, CA). The A_{260}/A_{280} ratio ranged from 2.06 to 2.11, whereas the RNA integrity number (RIN) values ranged from 8.7 to 10. Complementary DNA was synthesized using 2 μg of total RNA from each sample as template in reverse transcription reactions using Superscript III reverse transcriptase and oligo$(dT)_{15–18}$ primer (Invitrogen, Carlsbad, CA) as recommended by the manufacturer. Final concentration of cDNA was quantified by spectrophotometry (NanoDrop 1000), and the cDNA was diluted to a working stock containing 10 ng/μL and stored at −20°C.

Primer Design. Primer sequences for 7 potential reference genes and candidate gene *SLC7A1* are presented in Table 1. Potential reference genes were selected based on previous studies (Bionaz and Loor, 2007; Tramontana et al., 2008), but the primers used were different from those published to optimize the efficiency of the RT-qPCR reaction in our samples. Primers were designed based on publicly available swine cDNA and expressed sequence tag sequences deposited in the National Center for Biotechnology Information database using Primer Express software (v. 3.0, Applied Biosystems, Foster City, CA) with default settings. Designed primers were blasted against published swine (*Sus scrofa*), human (*Homo sapiens*), bovine (*Bos taurus*), and rat (*Rattus norvegicus*) genome sequences; pairs that showed significant alignment (i.e., high query coverage) with nucleotide sequences other than the protein of interest in any of the species mentioned were discarded. Amplicons from the primer pair were not sequenced. Evaluation of primer-dimer formation was performed by the presence of a single peak in the dissociation curve after the RT-qPCR reaction, indicative of a single amplification product. Primers were not designed to span exon–exon junctions; however, the method used to extract RNA provided a step for DNA digestion by performing on-column DNase treatment (PerfectPure RNADNase, 5 PRIME) to eliminate genomic DNA.

Primer pairs were optimized for concentration using a primer optimization matrix (Mikeska and Dobrovic, 2009), and a relative standard curve was used to determine the efficiency (Yuan et al., 2006). The standard curve was constructed using cDNA synthesized from an RNA pool made of all samples using the following amounts of cDNA (in duplicate): 40, 20, 10, 5, and 2.5 ng. Efficiency of the RT-qPCR reaction for each gene was calculated from the slope of the standard curve

Table 1. Primer information for reverse-transcription quantitative real-time (RT-qPCR) assays

Accession No.[1]	Gene	Protein	Primer[2]	Primer (5′ to 3′)	E[3] (%)
CV872150.1	*API5*	Apoptosis inhibitor 5	F. 502	CTGGAGTGGTGGCAATAATCTCT	99.4
			R. 568	CCAAGGGAGCTCAGGTTTAGC	
AY610067.1	*MRPL39*	Mitochondrial ribosomal protein L39	F. 540	TCGCTGGAGCTTTCTGCTATG	103.5
			R. 601	TGTTGGCATCCACTCATCAAG	
NM_001123213.1	*VAPB*	Vesicle-associated membrane protein-associated protein B/C	F. 1012	TGGCGCTGGTGGTTTTG	101.9
			R. 1072	CCTACAAGGCGATCTTCCCTATG	
DQ452569.1	*ACTB*	β-Actin	F. 746	TGCGGGACATCAAGGAGAA	111.8
			R. 803	GCCATCTCCTGCTCGAAGTC	
AF017079.1	*GAPDH*	Glyceraldehyde-3-phosphate dehydrogenase	F. 376	CGTCCCTGAGACACGATGGT	97.6
			R. 429	CCCGATGCGGCCAAAT	
XM_001927465.1	*RPS23*	Ribosomal protein 23	F. 52	CCACCGACGGGACCATAA	99.6
			R.115	CAGGGCTGTGCCCAAATG	
XM_001927648.1	*MTG1*	Mitochondrial GTPase 1	F. 463	GGCAAGTCCTCGCTCATCAA	102.7
			R. 525	CTTGGTGGCTTTTCCTTTCCT	
NM_001012613.1	*SLC7A1*	CAT-1	F. 1172	GGGCTGCTGTTTAAGTTTTTGG	100.4
			R. 1234	CGTGGCGATTATTGGTGTTTT	

[1]Accession number corresponds to the cDNA or the expressed sequence tag (EST) sequence deposited in the National Center for Biotechnology Information database from which the primers were designed.

[2]Direction (F = forward; R = reverse) and hybridization position for each primer (5′–3′) within the nucleotide sequence from which the primers were designed.

[3]Primer pair efficiency (E) was calculated as follow: $E = -1 + 10^{(-1/slope)} \times 100$. The R^2 values for all standard curves for reference and candidate genes were >0.98, indicating excellent linear relationships between quantities of serially diluted cDNA and cycle threshold (Ct) when RT-qPCR was performed.

using the following formula: $[10^{(-1/slope)} - 1] \times 100$, as described by Yuan et al. (2006). Specific hybridization of the primers was validated by the presence of a unique peak in the dissociation curve at the end of the RT-qPCR amplification. Nontemplate controls were included in all RT-qPCR plates to validate that primers were not amplifying contaminating DNA.

RT-qPCR Assay. A total of 3 μL (30 ng) of template cDNA, 12.5 μL of SYBR Green master mix (Applied Biosystems), 6 μL each of 10 μ*M* forward and reverse primers, and 3.5 μL of diethylpyrocarbonate (DEPC)-treated and nuclease-free water (Fisher Scientific, Fair Lawn, NJ) were used in each RT-qPCR reaction. Reactions were performed in MicroAmp Optical 96-Well Reaction Plates (Applied Biosystems) that were sealed with sealing foil, centrifuged at 400 × *g* for 1 min, and loaded into an ABI Prism 7500 Sequence Detection System (Applied Biosystems). Cycle to threshold values were obtained in duplicate for each sample on the analyzer. Coefficient of variation was calculated for all duplicates and 0.1 used as cut-off value. Conditions for amplification and quantification included 2 initial steps (50°C for 2 min and 95°C for 10 min) followed by an amplification program (step 3) repeated for 40 cycles (95°C for 15 s and 60°C for 1 min). Data was calculated with the 7000 RQ Sequence Detection Systems Software (version 2.2.1, Applied Biosystems). A relative standard curve was used as the RT-qPCR method to correct for differences on RT-qPCR reaction efficiency between plates (Larionov et al., 2005). Normality of the residuals was tested using the Shapiro-Wilk test under the UNIVARIATE procedure (SAS Institute Inc., Cary, NC).

Total RNA and DNA Quantification. Total DNA and RNA were extracted from mammary tissue and measured as described before (Labarca and Paigen, 1980; Capuco et al., 2001). Values for each sample were $\log_2$-transformed for further analysis as described below.

Analytical Protocol for Reference Gene Selection

Three consecutive steps were used (Figure 1): (1) log-transformation of the relative amounts from the standard curve for each gene (Steibel et al., 2009); (2) a lack-of-fit test to assess a linear relationship between expression of each reference gene and tissue RNA and DNA concentrations; and (3) a linear model to assess invariance of each reference gene expression across samples after correction with RNA and DNA concentrations.

Log-Transformation of the Relative Amounts from the Standard Curve. Duplicate relative RNA amounts from the standard curve for each sample were $\log_2$-transformed and averaged:

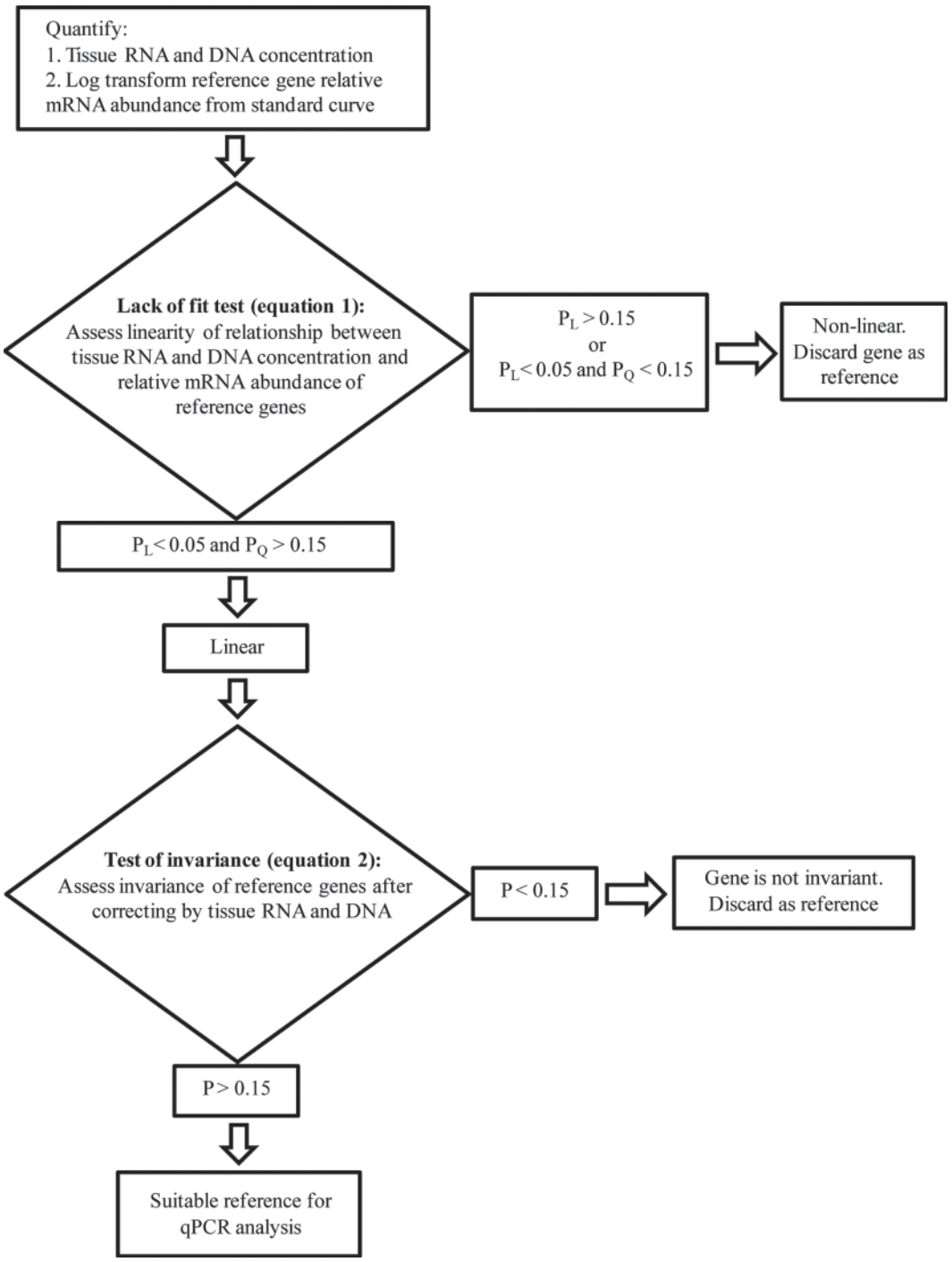

Figure 1. Flowchart for selection of suitable reference genes for reverse-transcription quantitative PCR (RT-qPCR) reaction using tissue RNA and DNA concentration to account for artifactual dilution effect in gene expression. P_L = *P*-value for testing linear term in equation 1; P_Q = *P*-value for testing quadratic term in equation 1; and P = *P*-value for stage factor model with DNA and RNA concentrations in mammary tissue in equation 2.

$$\mathrm{Log}_2 Y_{ij} = \frac{\mathrm{Log}\left(Q_{1ij}\right) + \mathrm{Log}\left(Q_{2ij}\right)}{2}, \qquad [1]$$

where Y_{ij} is the value of each potential reference gene corrected by the standard curve and $\log_2$-transformed for each sow (j) at each stage of lactation (i), and Q_{1ij} and Q_{2ij} are the quantities from the thermocycler

software corrected by the standard curve for each gene (i.e., duplicates).

Lack-of-Fit Test. Test for a linear relationship between each potential reference gene and mammary RNA and DNA concentrations was performed by a lack-of-fit sums of squares test using a linear mixed model procedure (PROC MIXED) in SAS (Figure 2a). The model included the random effect of sow and the continuous effect of tissue RNA and DNA concentration. The statistical model is as follows:

$$\mathrm{Log_2Y_{ij}} = \mu + \beta_1 x_{1ij} + \beta_2 {x_{1ij}}^2 + \beta_3 x_{2ij} + \beta_4 {x_{2ij}}^2 + \mathrm{b_j} + \mathrm{e_{ij}}, \quad [2]$$

where $\mathrm{Log_2Y_{ij}}$ is the value of potential reference genes corrected by the standard curve and $\log_2$-transformed for each sow (j) at each stage of lactation (i), μ is the overall mRNA mean, β_1 and β_2 are the regression coefficients relating tissue RNA concentration to the potential reference gene for each sow (j), x_{1ij} is the tissue RNA concentration for the jth sow within the ith stage of lactation, β_3 and β_4 are the regression coefficients relating tissue DNA concentration to the potential reference gene, x_{2ij} is the total DNA for the jth sow within the ith stage of lactation, $\mathrm{b_j}$ is the random effect of the sow, and $\mathrm{e_{ij}}$ is the experimental error.

Linear Model. A test for invariance expression of reference genes with linear relationship between gene mRNA abundance and tissue RNA and DNA concentrations was performed using a linear mixed model in SAS (Figure 2b). The model included stage of lactation as a fixed effect, sow as random effect, and DNA and RNA concentrations in mammary tissue as continuous variables. The statistical model is as follows:

$$\mathrm{Log_2Y_{ij}} = \mu + \alpha_\mathrm{i} + \beta_1 x_\mathrm{ij} + \beta_2 x_{2ij} + \mathrm{b_j} + \mathrm{e_{ij}}, \quad [3]$$

where $\mathrm{Log_2Y_{ij}}$ is the value of potential reference genes corrected by the standard curve and log base 2 transformed for each sow (j) at each stage of lactation (i), μ is the overall mRNA mean, α_i is the fixed effect of the ith level of stage of lactation, β_i is the regression coefficient relating tissue RNA concentration to the potential reference gene, x_{1ij} is the tissue RNA concentration for the jth sow within the ith stage of lactation, β_2 is the regression coefficient relating tissue DNA concentration to the potential reference gene, x_{2ij} is the total DNA for the jth sow within the ith stage of lactation, $\mathrm{b_j}$ is the random effect of the sow, and $\mathrm{e_{ij}}$ is the experimental error.

a)

```
proc mixed data=exp1;
  by gene;
  class sow stage;
  model logamt= logRNA logDNA logRNA2 logDNA2/htype=1;
  random sow;
run;
```

b)

```
proc mixed data=exp1;
  by gene;
  class sow stage;
  model logamt= stage logRNA logDNA;
  random sow;
  lsmeans stage/pdiff;
run;
```

Figure 2. Software code (PROC MIXED; SAS Institute Inc., Cary, NC) to assess (a) the relationship between potential reference gene expression and mammary RNA and DNA concentration, and (b) the invariance of each reference gene expression across samples after correction with mammary RNA and DNA concentrations. Color version available in the online PDF.

Validation of Protocol for Reference Gene Selection

Validation of the analytical protocol was performed by comparing expression results of a target gene normalized using the reference genes selected either with the proposal analytical protocol or with geNorm.

Selection of Best Reference Genes with geNorm. To determine the most stable genes among the set of potential reference genes, relative mRNA amounts from the standard curve were entered directly into the geNorm software, as described previously by Vandesompele et al. (2002). The gene expression stability (**M**) value for each gene and the numbers of reference genes that should be used as normalization factors were then calculated by geNorm. Briefly, M was determined as the average pairwise variation of each gene with all other reference genes, whereas the number of reference genes that should be used was calculated by analysis of the pairwise variation (Vn/Vn+1) between 2 sequential normalization factors (NFn and NFn+1). Such normalization factors (NFn) are based on the geometric mean of the expression levels of n and n+1 best reference genes (Vandesompele et al., 2002).

Target Gene Normalization and Statistical Analysis. Normalization of relative gene expression values from a standard curve was made by dividing the relative mRNA amounts of each target gene by the geometric mean of relative mRNA amounts of several reference genes:

$$Y_{ij} = \frac{Q_{Cij}}{\sqrt[3]{Q_{R1ij} \times Q_{R2ij} \times Q_{R3ij}}}, \quad [4]$$

where Y_{ij} are target gene normalized relative mRNA amounts for the *j*th sow within the *i*th stage of lactation, and Q_{Cij} and Q_{Rij} are target and reference gene quantities from the thermocycler software corrected by the standard curve, respectively. Then, the quotient between target and reference gene quantities (Equation [4]) was $\log_2$-transformed:

$$\begin{aligned} Log_2Y_{ij} &= Log_2\left[\frac{Q_C}{\sqrt[3]{Q_{R1} \times Q_{R2} \times Q_{R3}}}\right] \\ &= Log_2\left(Q_C\right) - Log_2\left(Q_{R1} \times Q_{R2} \times Q_{R3}\right)^{\frac{1}{3}} \\ &= Log_2\left(Q_C\right) - \frac{1}{3}\left(Log_2Q_{R1} + Log_2Q_{R2} + Log_2Q_{R3}\right), \end{aligned} \quad [5]$$

where Log_2Y_{ij} is the value of target gene corrected by the standard curve, $\log_2$-transformed and normalized for the jth sow within the ith stage of lactation, Q_{Cij} and Q_{Rij} are target and reference gene quantities from the thermocycler software corrected by the standard curve, respectively. Both normalization methods (Equations [4] and [5]) are mathematically equivalent. Thus, target gene fold changes calculated with data obtained from either Equation [4] or [5] are identical. However, log-transformed quantities are additive, improving normality of the residuals and homogeneity of the variance (Steibel et al., 2009).

Changes in target gene expression in response to physiological phases of mammary activity were assessed using a linear mixed model procedure of SAS that included the fixed effect of stage of lactation and the random effect of sow:

$$Log_2Y_{ij} = \mu + \alpha_i + b_j + e_{ij}, \quad [6]$$

where Log_2Y_{ij} is the expression value of target gene corrected by the standard curve, $\log_2$-transformed and normalized for the jth sow within the ith stage of lactation, μ is the overall mean of target gene, α_i is the fixed effect of the ith stage of lactation, b_j is the random effect of the jth sow, and e_{ij} is the experimental error.

Results were presented as target gene fold changes. Briefly, least squares means at prepartum were subtracted from least squares means at early lactation, peak lactation, and postweaning, and then 2 was raised to the power of the differences:

$$Y = 2^{(\Delta LogY\ Early\ lactation - \Delta LogY\ Prepartum)}, \quad [7]$$

where Y is the target gene fold change. *P*-values associated with the fold changes were the same as those from least squares means differences, and therefore were taken directly from the SAS output. Intervals of standard error bars for each fold change were calculated using the following formula:

$$Z = 2^{\mu \pm SE}, \quad [8]$$

where *Z* are the upper and lower limits of standard error bars for fold change, μ is the least squares means difference, and SE is the standard error of the least squares means difference obtained from the SAS output. Multiple comparisons were accounted for with Bonferroni adjustment, and $P < 0.05$ was used for determining significance.

RESULTS

Analysis Protocol for Reference Gene Selection

The lack-of-fit test showed that a linear regression provided a reasonable fit ($P < 0.05$ for linear and P

> 0.15 for quadratic) to the regression between tissue RNA and DNA concentration and the relative mRNA abundance of *AP15*, *GAPDH*, *ACTB*, and *MRPL39* (Table 2). Tissue DNA concentration was included in the lack-of-fit test to correct for changes in cell number; however, the *P*-value associated with the relationship between DNA concentration and the relative mRNA abundance of genes cannot be used to select reference genes, as not all tissues undergo changes in cell number (see Appendix Table A1). The linear mixed model fitting tissue RNA and DNA concentrations as regression variables to assess invariance of each reference gene expression across stages of mammary physiological activity showed that relative mRNA abundance of *AP15*, *MRPL39* and *GAPDH* remained unchanged ($P > 0.15$), whereas that of *ACTB* was significantly different ($P \leq 0.15$) (Table 2).

Validation of Protocol for Reference Gene Selection

Analysis by geNorm showed that *MTG1*, *MRPL39*, and *VAPB* had the lowest average expression stability parameter (M), and therefore were selected as the most stable set in porcine mammary gland (Figure 3a). Analysis of the pairwise variation between sequential normalization factors showed that the optimal number of reference genes was 3, as their pairwise variation was below the 0.15 cut-off value proposed by Vandesompele et al. (2002), and inclusion of a fourth gene increased (V2/3 = 0.11 and V3/4 = 0.12) the pairwise variation (Figure 3b).

Expression of target gene *SLC7A1* was significantly different across stages of mammary physiological activity when normalized with the geometric mean of the relative mRNA abundance of reference genes selected with the proposed protocol (i.e., *AP15*, *MRPL39* and *GAPDH*; $P < 0.0001$) or that of geNorm (i.e., *MRPL39*, *MTG1*, and *VAPB*; $P = 0.0001$). Fold change ± SE and *P*-values for the different comparisons are presented in Figure 4. Compared with prepartum, *SLC7A1* fold change increased at early ($P < 0.01$) and peak lactation ($P < 0.001$), irrespective of the normalizing gene set used.

DISCUSSION

Previous studies have shown that upregulation of genes involved in milk synthesis causes a large increase in transcript abundance of those genes, resulting in an artifactual dilution of mRNA abundance of reference genes that actually maintain a stable amount of mRNA per cell during lactation (Bionaz and Loor, 2007; Tramontana et al., 2008). In support of these authors' work, the current results indicated a significant increase ($P < 0.05$; Figure 5) in RNA per milligram of mammary tissue from prepartum to peak lactation, whereas relative mRNA abundance of all potential reference genes decreased over the same period ($P < 0.05$; Figure 6a, b and c; see statistical analysis of tissue RNA and DNA concentrations, and relative mRNA abundance of reference genes in the Appendix). As an extreme example of this paradox, data analysis showed an amino acid transporter, known to be upregulated during lactation (Manjarin et al., 2011), as the only gene whose expression remained unchanged from prepartum to postweaning ($P > 0.1$; Figure 6c). Consequently, validation of reference gene expression across different stage of mammary physiological activity depends on an approach that accounts for artifactual changes in expression due to variation in tissue RNA concentration.

Mammary tissue undergoes continuous changes in cell proliferation during lactation, which can be quantified

Table 2. Lack-of-fit test and linear mixed model (fitting tissue RNA and DNA concentration) used to assess invariance of *API5*, *MRPL39*, *VAPB*, *GAPDH*, *ACTB*, *MTG1*, *RPS23*, and *SLC7A1* in porcine mammary tissue over the prepartum to postweaning periods

	Linear[1]	Quadratic[2]	
Gene	RNA	RNA	Stage: *P*-value[3]
API5	<0.001	0.28	0.36
GAPDH	0.01	0.83	0.22
MRPL39	0.02	0.43	0.16
ACTB	<0.001	0.15	0.03
MTG1	0.06	0.51	NA[4]
VAPB	0.35	0.25	NA
RPS23	0.07	0.06	NA
SLC7A1	0.47	0.10	NA

[1] *P*-value for testing linear term in equation [1].

[2] *P*-value for testing quadratic term in equation [1].

[3] *P*-value for stage factor model with RNA concentration in mammary gland in equation [2].

[4] Not applicable.

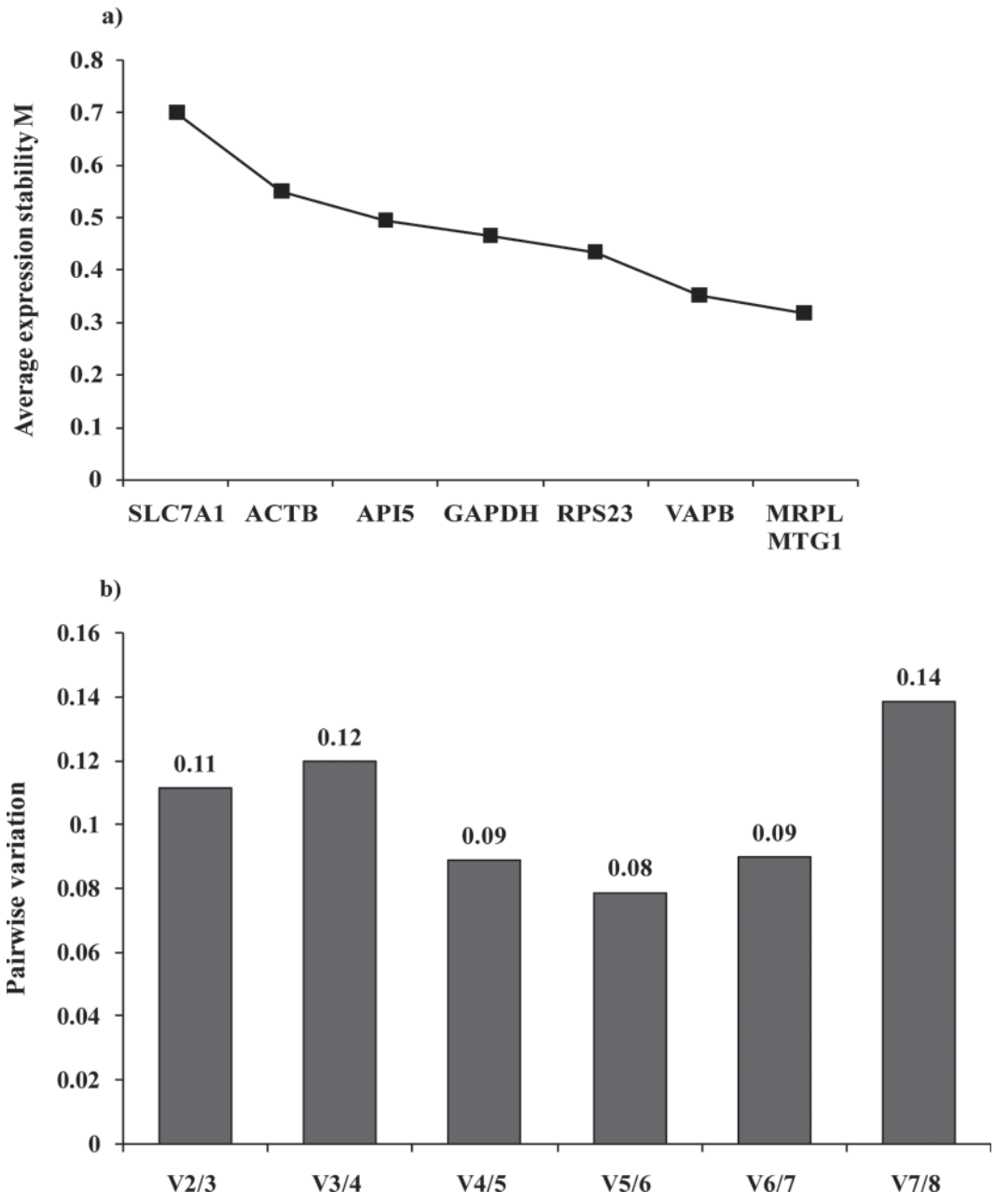

Figure 3. (a) Average expression stability values (M) of reference genes plotted from least stable (left) to most stable (right), and (b) pairwise variation (Vn/Vn + 1) between the normalization factors NFn and NFn+1 used for determination of optimal number of reference genes for normalization.

by analysis of mammary DNA concentration (Tucker, 1969, 1987; Capuco et al., 2001). Accordingly, Kim et al. (1999) showed a linear increase in the amount of DNA in porcine mammary gland during lactation ($P < 0.05$), suggesting cellular hyperplasia in the tissue. In the present study, a significant decrease ($P < 0.05$) was observed in DNA concentration postweaning (Figure 5), probably due to apoptosis of mammary cells (Ford et al., 2003). Variation in cell number is likely accompanied by changes in tissue RNA concentration, which does not cause an artifactual dilution of mRNA. Thus, if mammary DNA concentration differed between samples, the use of mammary RNA concentration alone as a correction factor for reference gene selection would lead to misleading results (Appendix Table A1). Conversely, utilization of both RNA and DNA concentrations would provide the exact amount of RNA causing the artifactual dilution effect. Consequently, to protect

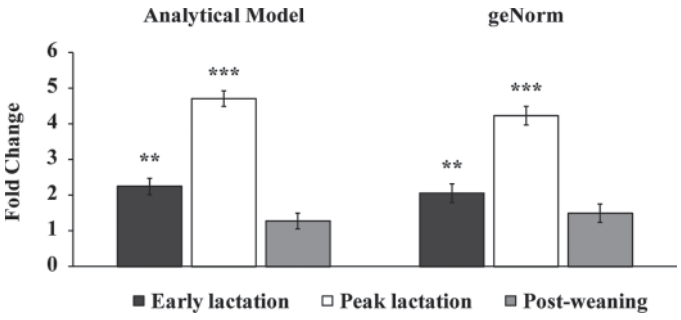

Figure 4. Fold changes of target gene *SLC7A1* in porcine mammary tissue at early lactation, peak lactation, and postweaning compared with prepartum. Target gene was normalized using the geometric mean of reference genes selected either with the proposed analytical protocol or with geNorm. $^{**}P < 0.01$; $^{***}P < 0.001$.

against tissue RNA changes unrelated to gene expression, mammary DNA concentration is also included as covariate in the proposed analytical method.

The proposed analytical protocol and the geNorm process selected different reference genes, but *SLC7A1* fold changes did not differ when using genes obtained under either method. These results indicate that, among the analyzed genes, more than one set of suitable reference genes exists. Differences between analyses are due to the statistical approach used. geNorm selected as the best reference genes those whose changes in mRNA abundance were more related to each other, whereas the proposed analytical protocol selected those reference genes whose expression levels were not different across stages of mammary physiological activity, after accounting for variation in tissue RNA and DNA concentrations. It is noteworthy that the results from geNorm analysis in the present study coincided only partially with those previously reported by Tramontana et al. (2008), despite the fact that mammary tissue from lactating animals was used in both studies. As such, *MTG1* was previously discarded as one of the least stable genes, whereas in the present study, *MTG1* was selected as the most stable gene, along with *MRPL39*. Variation between analyses is likely due to the inclusion of few different genes and one additional stage of mammary physiological activity (i.e., postweaning) in the present study. These results indicate the importance of testing for invariance of potential reference genes in every experiment, rather than relying on previously published reference genes.

The proposed analytical protocol and geNorm differ in several ways. In contrast to geNorm, the proposed approach does not account for analytical errors inherent to the RT-qPCR reaction. However, the artifactual dilution effect is not an analytical error. The method presented herein uses tissue RNA and DNA concentrations, which are directly related to reference gene quantities, to account for the artifactual dilution effect. The same dilution effect applies to all genes, so normalizing the target gene with suitable reference genes cancels out the dilution effect. As a result, the proposed method is a tool to uncover appropriate reference genes, but normalization of RT-qPCR data with reference gene expression corrected by tissue RNA and DNA is not recommended. On the other hand, geNorm

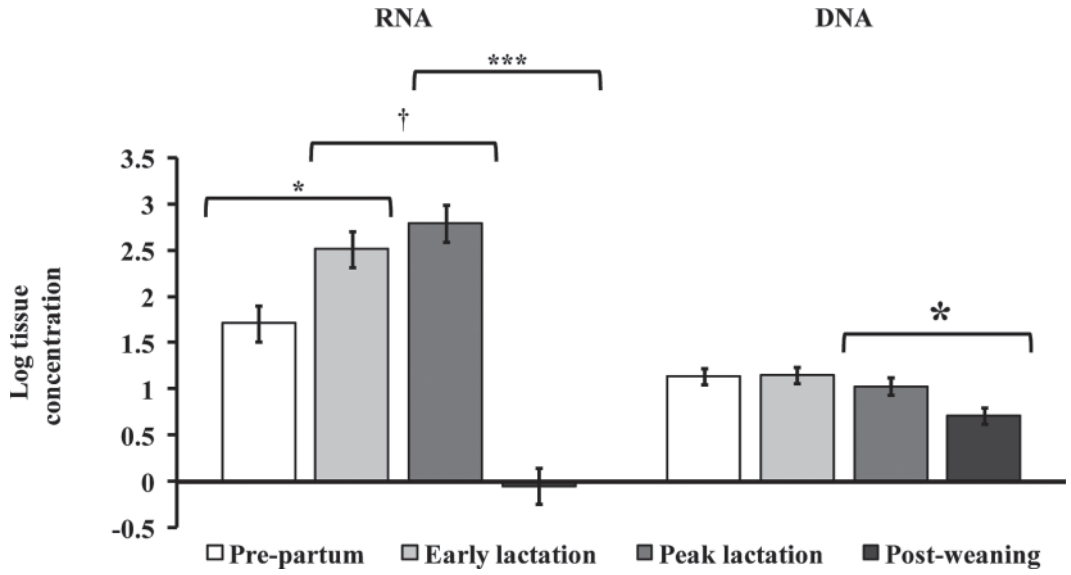

Figure 5. Log-transformed RNA and DNA concentrations in sow mammary tissue. Data were compared between d 5 of lactation (early) and d 110 of gestation (prepartum), d 17 of lactation (peak) and early lactation, and d 5 after weaning (postweaning) and peak lactation. Statistical analysis is shown in Appendix. $\dagger P < 0.1$; $^{*}P < 0.05$; $^{***}P < 0.001$.

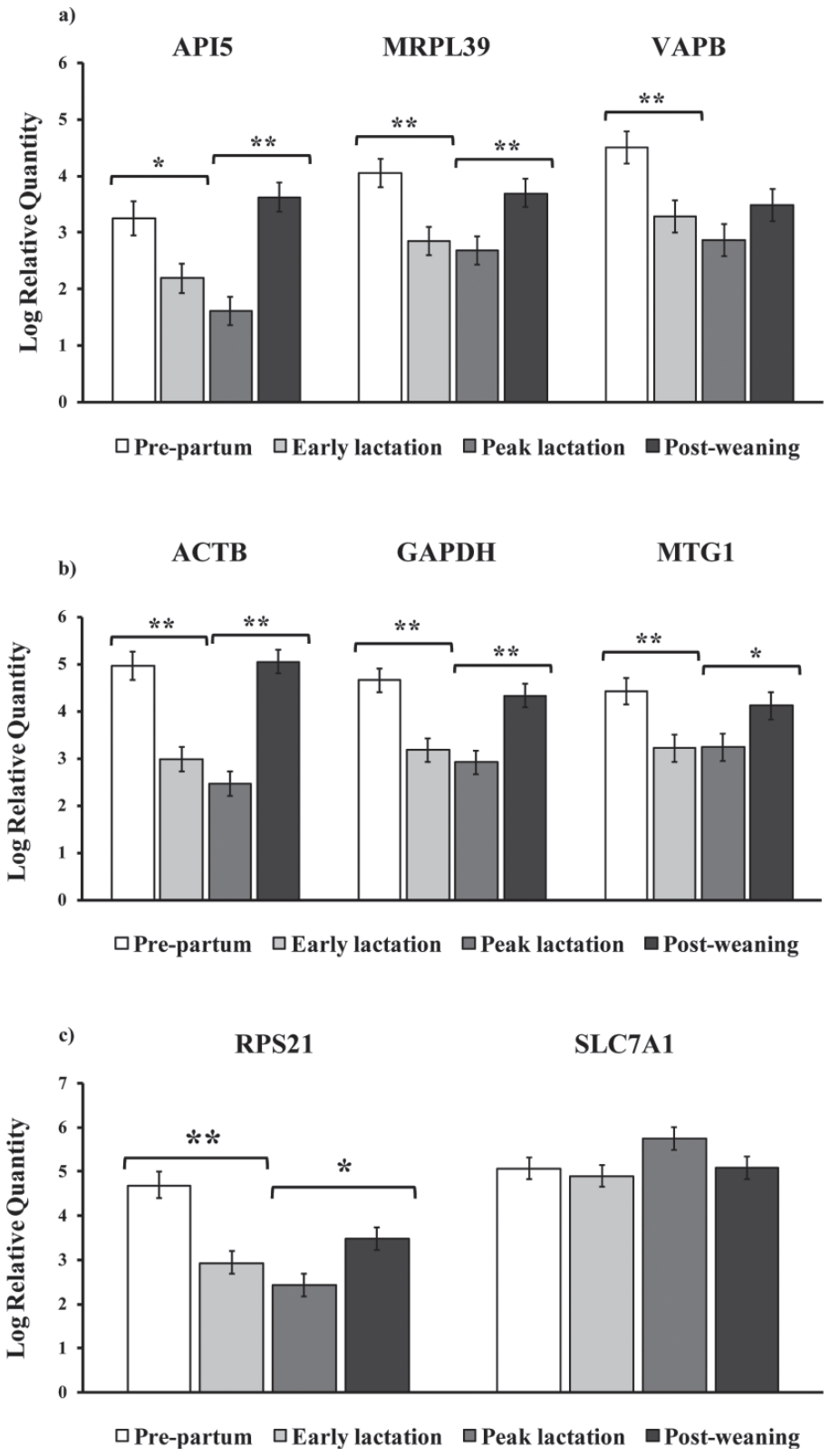

Figure 6. Log-transformed relative mRNA abundance of (a) *API5*, *MRPL39*, and *VAPB*; (b) *ACTB*, *GAPDH*, and *MTG1*; and (c) *RPS21* and *SLC7A1* in sow mammary tissue measured by reverse-transcription quantitative PCR (RT-qPCR). Data were compared between d 5 of lactation (early) and d 110 of gestation (prepartum), d 17 of lactation (peak) and early lactation, and d 5 after weaning (postweaning) and peak lactation. Statistical analysis is shown in Appendix. $^{*}P < 0.05$; $^{**}P < 0.01$.

analysis relies on current knowledge of interactions between the selected genes, which to date are only partially understood. Conversely, the proposed analytical model does not assume lack of co-regulation between potential reference genes because invariance analysis of each gene tested is independent of expression values of every other gene in the data set.

Finally, the proposed analytical method quantifies the exact number of genes that remain stable. Thus, all reference genes in the set will be discarded if significant differences in relative mRNA remain after accounting for tissue DNA and RNA variation. Likewise, a single reference gene can also be selected. However, normalization against a single reference gene is not acceptable. The reference gene corrects for all other transcripts and thus, even though its mRNA abundance remains invariant under the experimental conditions described, an error in reference gene expression analysis will drastically affect the results. In fact,, Vandesompele et al. (2002) demonstrated an error associated with the use of only 1 or even 2 reference genes; thus, to account for inherent variation in expression of reference genes, a minimum of 3 control genes is recommended (Bustin et al., 2009).

CONCLUSIONS

In summary, the proposed analytical protocol involves 3 consecutive steps to accurately test for reference gene invariance (Figure 1). First, it requires the log-transformation of relative quantities for analysis of gene expression values in a linear model. Second, it assesses the relationship between tissue RNA and DNA concentrations and reference gene expression with a lack-of-fit test, as only a linear relationship between mRNA abundance of reference genes and the external parameters to the RT-qPCR reaction support a dilution effect on gene expression. Finally, it challenges the invariance of reference genes using a linear model that accounts for variation in tissue RNA and DNA concentrations between samples. In conclusion, the proposed analytical protocol assesses expression invariance of potential reference genes by accounting for the artifactual dilution effect in mRNA abundance. As a result, it represents a valid alternative to select suitable reference genes for RT-qPCR analysis.

ACKNOWLEDGMENTS

The execution of this work was made possible by funding from the National Pork Board (Des Moines, IA). The authors thank the Michigan State University Swine Teaching and Research Center (East Lansing) staff, Al Snedegar and Kevin Turner, for assistance in animal care and handling. Valuable thoughts from anonymous reviewers are also appreciated.

REFERENCES

Bionaz, M., and J. J. Loor. 2007. Identification of reference genes for quantitative real-time PCR in the bovine mammary gland during the lactation cycle. Physiol. Genomics 29:312–319.

Bustin, S. A. 2000. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. J. Mol. Endocrinol. 25:169–193.

Bustin, S. A. 2002. Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): Trends and problems. J. Mol. Endocrinol. 29:23–39.

Bustin, S. A., J. F. Beaulieu, J. Huggett, R. Jaggi, F. S. Kibenge, P. A. Olsvik, L. C. Penning, and S. Toegel. 2010. MIQE précis: Practical implementation of minimum standard guidelines for fluorescence-based quantitative real-time PCR experiments. BMC Mol. Biol. 11:74–79.

Bustin, S. A., V. Benes, J. A. Garson, J. Hellemans, J. Huggett, M. Kubista, R. Mueller, T. Nolan, M. W. Pfaffl, G. L. Shipley, J. Vandesompele, and C. T. Wittwer. 2009. The MIQE guidelines: Minimum information for publication of quantitative real-time PCR experiments. Clin. Chem. 55:611–622.

Capuco, A. V., D. L. Wood, R. Baldwin, K. Mcleod, and M. J. Paape. 2001. Mammary cell number, proliferation, and apoptosis during a bovine lactation: Relation to milk production and effect of bST. J. Dairy Sci. 84:2177–2187.

Derveaux, S., J. Vandesompele, and J. Hellemans. 2010. How to do successful gene expression analysis using real-time PCR. Methods 50:227–230.

Ford, J. A., S. W. Kim, S. L. Rodriquez-Zas, and W. L. Hurley. 2003. Quantification of mammary gland tissue size and composition changes after weaning in sows. J. Anim. Sci. 81:2583–2589.

Kim, S. W., W. L. Hurley, I. K. Han, and R. A. Easter. 1999. Changes in tissue composition associated with mammary gland growth during lactation in sows. J. Anim. Sci. 77:2510–2516.

Kirkwood, R. N., J. Pérez-Laspiur, N. K. Ames, J. B. Moore, A. Cegielski, and N. L. Trottier. 2007. Mammary gland biopsy does not affect lactation performance in sows. Can. J. Anim. Sci. 87:281–284.

Labarca, C., and K. Paigen. 1980. A simple, rapid and sensitive DNA assay procedure. Anal. Biochem. 102:344–352.

Larionov, A., A. Krause, and W. Miller. 2005. A standard curve based method for relative real time PCR data processing. BMC Bioinformatics 6:62–77.

Manjarin, R., J. P. Steibel, V. Zamora, N. Am-in, R. N. Kirkwood, C. W. Ernst, P. S. Weber, N. P. Taylor, and N. L. Trottier. 2011. Transcript abundance of amino acid transporters, β-casein and α-lactalbumin in mammary tissue of peri-parturient, lactating and post-weaned sows. J. Dairy Sci. 94:3467–3476.

Mikeska, T., and A. Dobrovic. 2009. Validation of a primer optimisation matrix to improve the performance of reverse transcription -quantitative real-time PCR assays. BMC Res. Notes 2:112–117.

NRC. 1998. Nutrient Requirements of Swine. 10th rev. ed. National Academy Press, Washington, DC.

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

Pfaffl, M. W., A. Tichopad, C. Prgomet, and T. P. Neuvians. 2004. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper–Excel-based tool using pair-wise correlations. Biotechnol. Lett. 26:509–515.

Steibel, J. P., R. Poletto, P. M. Coussens, and G. J. Rosa. 2009. A powerful and flexible linear mixed model framework for the analysis of relative quantification RT-PCR data. Genomics 94:146–152.

Tramontana, S., M. Bionaz, A. Sharma, D. E. Graugnard, E. A. Cutler, P. Ajmone-Marsan, W. L. Hurley, and J. J. Loor. 2008. Internal controls for quantitative polymerase chain reaction of swine mammary glands during pregnancy and lactation. J. Dairy Sci. 91:3057–3066.

Tucker, H. A. 1969. Factors affecting mammary gland cell numbers. J. Dairy Sci. 52:720–729.
Tucker, H. A. 1987. Quantitative estimates of mammary growth during various physiological states: A review. J. Dairy Sci. 70:1958–1966.
Vandesompele, J., K. De Preter, F. Pattyn, B. Poppe, N. Van Roy, A. De Paepe, and F. Speleman. 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol. 3:research0034.1–research0034.11.
Wong, M. L., and J. F. Medrano. 2005. Real-time PCR for mRNA quantitation. Biotechniques 39:75–85.
Yuan, J. S., A. Reed, F. Chen, and C. N. Stewart. 2006. Statistical analysis of real-time PCR data. BMC Bioinformatics 7:85–97.

APPENDIX

Statistical Analysis for Figures 5 and 6

Test for changes of tissue RNA and DNA concentrations (Figure 5) was performed using a linear mixed model in SAS (Table 2). The model included stage of lactation as a fixed effect and sow as random effect. The statistical model is as follows:

$$Z_{ij} = \mu + \alpha_i + b_j + e_{ij},$$

where Z_{ij} is the tissue RNA or DNA concentration for each sow (j) at each stage of lactation (i), μ is the overall tissue RNA or DNA concentration mean, α_i is the fixed effect of the ith level of stage of lactation, b_j is the random effect of the sow, and e_{ij} is the experimental error.

Test for invariance expression of reference genes without accounting for tissue RNA and DNA concentration (Figure 6a, b, and c) was performed using a linear mixed model in SAS. The model included stage of lactation as a fixed effect and sow as random effect. The statistical model is as follows:

$$Y_{ij} = \mu + \alpha_i + b_j + e_{ij},$$

where Y_{ij} is the value of potential reference genes corrected by the standard curve and log base 2 transformed for each sow (j) at each stage of lactation (i), μ is the overall mean, α_i is the fixed effect of the ith level of stage of lactation, b_j is the random effect of the sow, and e_{ij} is the experimental error.

Table A1. Example regarding the utilization of both RNA and DNA concentration as covariates for selection of reference genes in porcine mammary tissue[1]

Sample	Stage	RNA	DNA	RNA:DNA
Sample 1	1	10	10	1
	2	20	20	1
	3	30	30	1
	4	10	10	1
Sample 2	1	10	10	1
	2	20	10	2
	3	30	10	3
	4	10	10	1
Sample 3	1	10	10	1
	2	20	12	1.67
	3	30	14	2.14
	4	1	7	0.14

[1]In sample 1, changes in tissue RNA concentration are only due to an increase in cell proliferation. Therefore, utilization of these RNA concentration values as covariates to correct for reference gene expression would be incorrect. Utilization of the RNA:DNA ratio is necessary to protect against an increase in RNA concentration unrelated to gene expression. In sample 2, changes in tissue RNA concentration are due only to an increase in cell activity. Therefore, utilization of these RNA concentration values as covariates to correct for reference gene expression is necessary, whereas the use of RNA:DNA instead would not modify the results. Finally, sample 3 represents the mammary gland during lactation. A clear increase in cell activity occurs, indicated by an increase in RNA:DNA. However, changes also occur in cell proliferation, indicated by variation in DNA concentration. Therefore, utilization of RNA concentration alone as covariate to correct for reference gene expression would be incorrect, because RNA changes are associated with both cell activity and cell proliferation (which does not cause any dilution effect on the reference genes). Utilization of RNA:DNA instead provides the exact amount of RNA causing the dilution effect.