Published in final edited form as:

J Med Entomol. 2013 January ; 50(1): 79–84.

Validation of Internal Reference Genes for Real-Time Quantitative Polymerase Chain Reaction Studies in the Tick, *Ixodes scapularis* (Acari: Ixodidae)

Juraj Koči, Ladislav Šimo, and Yoonseong Park¹

Department of Entomology, Kansas State University, 123 Waters Hall, Manhattan, KS 66506

Abstract

Obtaining reliable gene expression data using real-time quantitative polymerase chain reaction (qPCR) is highly dependent on the choice of normalization method. We tested the expression stability of multiple candidate genes in the salivary glands (SG) and synganglia (SYN) of female *Ixodes scapularis* (Say) ticks in multiple blood-feeding phases. We found that the amount of total RNA in both the SG and SYN increases dramatically during tick feeding, with 34× and 5.8× increases from 62 and 7.1 ng of unfed tick, respectively. We tested candidate genes that were predicted from *I. scapularis* genome data to encode glyceraldehyde 3-phosphate dehydrogenase (*gapdh*), ribosomal protein L13A (*l13a*), TATA box-binding protein (*tbp*), ribosomal protein S4 (*rps4*), glucose 6-phosphate dehydrogenase (*gpdh*), and beta-glucuronidase (*gusb*). The geNorm and NormFinder algorithms were used to analyze data from different feeding phases (i.e., daily samples from unfed to fully engorged females over a 7-d period in three replicate experiments). We found that the *rps4* and *l13a* genes showed highly stable expression patterns over the feeding duration in both the SG and SYN. Furthermore, the highly expressed *rps4* gene makes it useful as a normalization factor when we perform studies using minute amounts of dissected tissue for qPCR. We conclude that *rps4* and *l13a*, whether individually or as a pair, serve as suitable internal reference genes for qRT-PCR studies in the SG and SYN of *I. scapularis*.

Keywords

tick; reference gene; qPCR; geNorm; NormFinder

Real-time quantitative polymerase chain reaction (qPCR) is one of the most precise and sensitive high-throughput technologies for gene expression analysis. Obtaining accurate qPCR results requires normalization of the data for the gene of interest to specific reference genes to compensate for such variables as the amount of starting material, the amplification efficiency, and other differences among the samples (Livak and Schmittgen, 2001, Pfaffl 2001, Bustin 2002). Ideal internal reference genes for normalization should show stable expression levels in spatiotemporal samplings of the biological material that are independent of experimental conditions (Thellin et al. 1999, Lee et al. 2002). Some commonly used reference genes in many organisms include glyceraldehyde-3-phosphate dehydrogenase (*gapdh*), beta-glucuronidase (*gusb*), TATA-box binding protein (*tbp*), ribosomal proteins, and β -actin (de Jonge et al. 2007, Valente et al. 2009), whereas the use of multiple reference genes is also suggested for enhanced accuracy (Vandesompele et al. 2002). Theoretically,

the simultaneous use of a large number of reference genes would be the most proper approach to normalization. However, to simplify the experimental procedure when resources are limited, use of only a few genes (or even a single reference gene) has been the most frequent approach used in many studies.

Studies of the blacklegged tick *Ixodes scapularis* (Say), an obligate hematophagous ectoparasite that transmits tick-borne diseases such as Lyme borreliosis (CDC 2011), were recently expanded to include the use of various molecular tools after the genome sequence became available in VectorBase (Megy et al. 2012). These include studies that analyze the target gene(s) spatiotemporal expression patterns and that use RNA interference for the suppression of target gene expression; such methodologies require techniques for accurate quantification of target mRNA for the validation of experimental results.

In this study, the transcripts of six commonly used reference genes glyceraldehyde-3-phosphate dehydrogenase (*gapdh*), ribosomal protein L13A (*l13a*), TATA box-binding protein (*tbp*), ribosomal protein S4 (*rps4*), glucose-6-phosphate dehydrogenase (*gpdh*), and beta-glucuronidase (*gusb*) were tested to identify the most suitable reference genes. We specifically focused on the blood feeding phases of female ticks, and samples were collected daily after the onset of blood feeding. To obtain tissue-specific expression data, we tested gene expressions in the salivary glands (SG) and synganglia (SYN). This study provides the most appropriate reference genes for future gene expression studies in *I. scapularis*.

Materials and Methods

Ticks and Tick Feeding

Adult *I. scapularis* ticks were obtained from a tick rearing facility at Oklahoma State University. Ticks were fed on rabbits in rubber capsules ($\approx 5 \times 5$ cm) glued onto the shaved backs of the animals using nonirritant Latex glue covered with a fine mesh net. Upon initiation of feeding, ticks were collected every day until full engorgement. Thus, the samples consisted of unfed ticks (D0), ticks feeding from 1 to 6 d (D1–D6) and fully engorged (FE) ticks at the D7. Three replications of this experiment were performed using different rabbits. This protocol was approved by the Kansas State University Institutional Animal Care and Use Committee (IACUC). All procedures were as described in the IACUC-approved protocol to minimize suffering.

RNA Isolation and cDNA Synthesis

SG and SYN were dissected and pooled from four to nine females per feeding stage in each biological replication in ice-cold phosphate-buffered saline. Samples were kept on dry ice immediately after dissection and then stored at -80°C until the RNA was extracted. Total RNA from the tissues was isolated using the RNeasy Plus Micro kit (Qiagen, Valencia, CA), which included on-column DNase treatment. The RNA concentrations and purities were determined with a NanoDrop 2000 spectrophotometer (Thermoscientific, Wilmington, DE). All the samples routinely reached the A260/280 ratio in between 1.8 and 2.2. Total RNA, 340 ng from SG and 40–60 ng from SYN, was used to synthesize first-strand cDNA using the SuperScript III system (Invitrogen, Life sciences, Grand Island, NY) and an oligo-dT-based reverse transcriptase protocol.

Real-Time qPCR and Data Analysis

A SYBR Green-based qPCR assay was performed for six commonly used reference genes: *gapdh*, *l13a*, *tbp*, *rps4*, *gpdh*, and *gusb* (Table 1). Primers were designed using Oligo Explorer (Gene Link, Hawthorne, NY) and Beacon Designer (Premier Biosoft, Palo Alto, CA) (Table 1). To determine amplification efficiencies of the tested reference genes,

standard calibration curves were constructed using 10-fold dilution series of the cDNA obtained from carcasses of partially fed tick. The 20- μ l qPCR reactions in opaque plates (Bio-Rad, Hercules, CA) contained 1 $\times$ SYBR Premix Ex *Taq* (Takara Bio Inc., Otsu, Japan), 1 $\times$ ROX reference dye, 250 nM of each primer, and 1 μ l of diluted cDNA (3.2 and 0.6 ng of SG and SYN total RNA equivalent, respectively). The qPCR reactions were carried out on an iCycler IQ thermal cycler (Bio-Rad) in duplicate wells. The cycling protocol comprised a 2-min polymerase activation step at 95°C followed by 40 cycles at 95°C for 15 s, 55°C for 30 s, and 72°C for 30 s. The annealing temperature 55°C was found to be the optimal for multiple genes in the optimization processes using a gradient thermocycler. Each assay included melting curve analysis from 55 to 95°C to ensure the absence of primer-dimer formation and amplicon specificity. In addition, the specificities of the PCR products were confirmed by determining the amplicon sizes using agarose gel electrophoresis.

The amplification efficiency for each tested gene was calculated based on calibration curves using iQ5 software (Bio-Rad). Raw C_q (quantification cycle, also referred to as C_t , threshold cycle) were obtained by assuming the same threshold fluorescent values for all genes, converted by the specific amplification efficiencies for each gene, and analysis using the geNorm (Vandesompele et al. 2002) and NormFinder (Andersen et al. 2004) algorithms included in the GenEx qPCR analysis software (Exiqon Inc., Woburn, MA). In geNorm, an expression stability value M for every tested gene in each tick tissue was assessed for three biological replications and averaged. Analysis using NormFinder produced both intergroup and intragroup expression variation values. Initial analyses were performed to obtain the temporal variations of each gene within each tissue. Then, we tested the variations between the tissues SG and SYN.

Results and Discussion

We measured the total RNA content for each tissue over the course of the feeding period. In unfed female ticks, the amounts of total RNA isolated from SG and SYN were 62 and 7.1 ng per tick, respectively (Fig. 1). After the onset of feeding, total RNA in the D1 sample increased by 4.3 $\times$ in the SG and 3.2 $\times$ in the SYN. In the SG, an apparent pattern of increasing RNA levels from D1 through repletion, when the total RNA reached 2.1 μ g or 34 $\times$ of the unfed value, was observed, except for a small decrease in the D6 samples. The SG gradually increases its size in the diameter of each acinus from $\approx$ 40 μ m in unfed acini to $\approx$ 150 μ m in fully engorged in acini types II and III (L. Simo et al. unpublished data; Kemp et al. 1982) during feeding and slightly decreases at repletion. The changes in the levels of RNA are correlated with the SG size and the gland activities for synthesis and secretion of bioactive salivary components, such as collagen-like peptides, Kunitz domain-containing peptides, and metalloproteases, which are produced mainly in the early feeding phase (Ribeiro et al. 2006, Chmelar et al. 2008). In the SYN, a gradual increase in the levels of RNA was found until D4 (41 ng or 5.8 $\times$ of the unfed value), followed by a gradual decrease until full engorgement, while the SYN size does not change over the course of feeding (L. Simo et al., unpublished data). We consider D5 and D6 to comprise the rapid engorgement phase, during which time a major uptake of blood and an increase in body size occur before the repletion of female *I. scapularis*. The rapid engorgement associated with large changes in the physiology of the tick may require the increased transcriptional activities of SYN. Overall, the patterns of the changes in RNA quantity in both the SG and the SYN tissues are likely explained in part by the activities of the SG and the SYN during the rapid engorgement phase. The high magnitude of the increases in RNA levels (up to 34 $\times$ and 5.8 $\times$ of total RNA in the SG and the SYN tissues, respectively) as measured using a NanoDrop is surprising, although validation of this result by using other independent RNA quantification

methods and further investigation are required for a better understanding of the observed changes.

The amplification efficiencies of the six genes, assuming the same threshold, were in the range of 95–118%, except for *l13a*, which showed an amplification efficiency of 62% (Table 1). The amplification efficiencies obtained for each gene were used in subsequent calculations to identify the expression stabilities using geNorm and NormFinder. In general, C_q values were high because low amounts of cDNA (equivalent to 3.2 and 0.6 ng total RNA of the SG and SYN tissues per PCR reaction, respectively) were used in these experiments. The C_q values for each gene suggested that *rps4* was the most highly expressed gene, followed by *gpdh*, while the expression of *gusb* was very low or undetectable in our experimental conditions (Fig. 2). The data for *gusb* were excluded from expression stability tests because the high C_q values (almost reaching the maximum 40 cycles) with large variations could otherwise distort the results.

The gene expression stability values in each SG and SYN sample across the temporal scale, eight daily samples from unfed to fully engorged, were calculated for each biological replication, and the stability values were averaged (Tables 2 and 3). The expression stability value (M) obtained from geNorm analysis ranked *rps4* and *l13a* as the most stable gene pair, in both the SG and SYN samples (Table 2). Expression levels of the gene pair *rps4/l13a* were consistently stable across the three biological replications. Norm-Finder analysis also demonstrated the stable expression of both *rps4* and *l13a* (Table 3). The average intragroup variation, variation of a gene over time, across three biological replications in SG was the lowest for *l13a* (0.2), followed by *rps4* (0.23), with a relatively small difference between the two. The lowest intragroup variation in the SYN sample was observed for *rps4* (0.05). Summing up the results of two different analyses, we find that the *rps4/l13a* pair, or *rps4* alone, has the most stable expression, suggesting that these genes are appropriate internal normalization factors for feeding-related temporal comparisons in the SYN and the SG.

Because we found that the changes in expression levels of the candidate genes over the feeding period did not show a specific pattern, such as decreasing or increasing over time (Fig. 2), we tested whether the candidate genes could also be applicable for tissue-specific expression studies, specifically in comparisons between the SG and SYN (Fig. 3). In NormFinder, we find the smallest intertissue variation in *l13a* (Fig. 3). Intratissue variation was smallest in *rps4*, followed by *l13a*, supporting the assertion that the *rps4/l13a* pair (or either of the genes individually) could be used as a normalization factor in gene expression comparisons between the SG and the SYN, as well as in feeding-related temporal changes in the SG and the SYN. We find that the high expression of *rps4*, in particular, makes it useful as a normalization factor when we perform studies using minute amounts of dissected tissue for qPCR.

Ribosomal proteins, such as *l13a* and *rps4*, are considered as housekeeping genes which are constitutively expressed at a high level to maintain cellular function. Many of the genes encoding ribosomal proteins may be co-expressed in the same tissues, where a high demand for ribosomes for protein syntheses and cell divisions are expected (Thorrez et al. 2008). Indeed, genes encoding ribosomal proteins have been reported as useful and reliable internal controls in various gene expression studies ranging from insects (Mamidala et al. 2011, Scharlaken et al. 2008) to ticks (Nijhof et al. 2009), and to mammals (Mehta et al. 2010), although care must be taken when selecting the candidate ribosomal protein genes because of the likely tissue-specific differences in expression (Thorrez et al. 2008). In conclusion, we recommend *rps4* and *l13a*, either individually or as a pair, as suitable internal reference genes for qPCR studies in the SG and the SYN during blood feeding in female *I. scapularis*.

Acknowledgments

This study was funded by National Institutes of Health NIH-1R21AI081136, NIH-1R01AI090062-01, and the Arthropod Genomics Center at Kansas State University. This article is contribution no. 12-337-J from the Kansas Agricultural Experiment Station.

References Cited

- Andersen CL, Jensen JL, Orntoft TF. 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.* 2004; 64:5245–5250. [PubMed: 15289330]
- Bustin SA. Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. *J. Mol. Endocrinol.* 2002; 29:23–39. [PubMed: 12200227]
- (CDC) Centers for Disease Control and Prevention. Ticks. 2011. (<http://www.cdc.gov/ticks/index.html>).
- Chmelar J, Anderson JM, Mu J, Jochim RC, Valenzuela JG, Kopecky J. Insight into the sialome of the castor bean tick, *Ixodes ricinus*. *BMC Genomics.* 2008; 9:233. [PubMed: 18489795]
- de Jonge HJ, Fehrman RS, de Bont ES, Hofstra RM, Gerbens F, Kamps WA, de Vries EG, van der Zee AG, te Meerman GJ, ter Elst A. Evidence based selection of housekeeping genes. *PLoS ONE.* 2007; 2:e898. [PubMed: 17878933]
- Kemp, DH.; Stone, BF.; Binnington, KC. Tick attachment and feeding: role of the mouthparts, feeding apparatus, salivary gland secretions and the host response. In: Obenchain, F.; Galun, R., editors. *Physiology of Ticks*. Oxford, United Kingdom: Pergamon, Ltd; 1982. p. 119-168.
- Lee PD, Sladek R, Greenwood CM, Hudson TJ. Control genes and variability: absence of ubiquitous reference transcripts in diverse mammalian expression studies. *Genome Res.* 2002; 12:292–297. [PubMed: 11827948]
- Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods.* 2001; 25:402–408. [PubMed: 11846609]
- Mamidalá P, Rajarapu SP, Jones SC, Mittapalli O. Identification and validation of reference genes for quantitative real-time polymerase chain reaction in *Cimex lectularius*. *J. Med. Entomol.* 2011; 48:947–951. [PubMed: 21845960]
- Megy K, Emrich SJ, Lawson D, Campbell D, Dialynas E, Hughes DS, Koscielny G, Louis C, Maccallum RM, Redmond SN, Sheehan A, Topalis P, Wilson D. VectorBase: improvements to a bioinformatics resource for invertebrate vector genomics. *Nucleic Acids Res.* 2012; 40:D729–D734. [PubMed: 22135296]
- Mehta R, Birerdinc A, Hossain N, Afendy A, Chandhoke V, Younossi Z, Baranova A. Validation of endogenous reference genes for qRT-PCR analysis of human visceral adipose samples. *BMC Mol. Biol.* 2010; 11:39. [PubMed: 20492695]
- Nijhof AM, Balk JA, Postigo M, Jongejan F. Selection of reference genes for quantitative RT-PCR studies in *Rhipicephalus (Boophilus) microplus* and *Rhipicephalus appendiculatus* ticks and determination of the expression profile of Bm86. *BMC Mol. Biol.* 2009; 10:112. [PubMed: 20040102]
- Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* 2001; 29:D2002–D2007.
- Ribeiro JMC, Alarcon-Chaidez F, Francischetti IMB, Mans BJ, Mather TN, Valenzuela JG, Wikel SK. An annotated catalog of salivary gland transcripts from *Ixodes scapularis* ticks. *Insect Biochem. Mole. Biol.* 2006; 36:111–129.
- Scharlaken B, de Graaf DC, Goossens K, Brunain M, Peelman LJ, Jacobs FJ. Reference gene selection for insect expression studies using quantitative realtime PCR: the head of the honeybee, *Apis mellifera*, after a bacterial challenge. *J. Insect. Sci.* 2008; 8:33.
- Thellin O, Zorzi W, Lakaye B, DeBorman B, Coumans B, Hennen G, Grisar T, Igout A, Heinen E. Housekeeping genes as internal standards: use and limits. *J. Biotechnol.* 1999; 75:291–295. [PubMed: 10617337]

- Thorrez L, Van Deun K, Tranchevent LC, Van Lommel L, Engelen K, Marchal K, Moreau Y, Van Mechelen I, Schuit F. Using ribosomal protein genes as reference: a tale of caution. PLoS ONE. 2008; 3:e1854. [PubMed: 18365009]
- Valente V, Teixeira SA, Neder L, Okamoto OK, Oba-Shinjo SM, Marie SK, Scrideli CA, Paco-Larson ML, Carlotti CG Jr. Selection of suitable housekeeping genes for expression analysis in glioblastoma using quantitative RT-PCR. BMC Mol. Biol. 2009; 10:17. [PubMed: 19257903]
- Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol. 2002; 3:1-0034.11.

Fig. 1. Amount of total RNA isolated from the salivary glands (SG) and synganglia (SYN) over the feeding duration. The different letters on top of the bars indicate statistically significant differences ($P < 0.05$; pair-wise student t -test).

Salivary glands

Synganglia

Fig. 2.

Expression levels of the tested reference genes in the tick SG and SYN over the course of feeding. The figure for each tested tissue and biological replication (Rep. 1–3) represents mean C_q values with standard deviations for two technical replicates per tested tissue. Box-and-whisker plots represent lower and upper quartiles of the quantification cycle range with medians (–) and means ($\square$) indicated. Whiskers represent 10th and 90th percentiles. The three boxes shown per gene indicate three biological replications.

Fig. 3. Expression variations between SG and SYN obtained using NormFinder software over the blood-feeding duration. The data are the average of three biological replications. The circles are for intergroup variation and error bars are for intragroup variation (intergroup variation, variation of the genes between the tissues; intragroup variation, variation of the genes over the feeding duration of each tissue).

Table 1

List of candidate reference genes used in this study for finding normalization factors in spatiotemporal comparisons of the salivary glands (SG) and the synganglia (SYN) over a blood-feeding duration of female *I. scapularis*

Gene symbol	Gene name	Function	GenBank Accession no.	5' primer sequence 3'	Amplicon length (bp)	PCR efficiency (%)
<i>gapdh</i>	Glyceraldehyde-3-phosphate dehydrogenase	Oxidoreductase in glycolysis and gluconeogenesis	XM_002434302	F: AGTGAAAGCGTGTCATCG R: GGGTAAACCTTGTTGCTGTC	76	118
<i>l13a</i>	Ribosomal protein L13A	Structural component of 60S ribosomal subunit	XM_002436237	F: GCATTTTCGGGCAGTTTCTTG R: TACTTTTCGCCCTGGGGTTGA	280	62
<i>tbp</i>	TATA box-binding protein	Transcription factor	XM_002402081	F: AAGTTGGACCTGAAGAAG R: TGGATGATACGAGGCATAC	197	95
<i>rps4</i>	Ribosomal protein S4	Structural component of 40S ribosomal subunit	DQ066214	F: GGTGAAGAAAGATTGTCAAGCAGAG R: TGAAAGCCAGCAGGGTAGTG	80	99
<i>gpdh</i>	Glucose-6-phosphate dehydrogenase	Involved in pentose phosphate pathway	XM_002411084	F: AGCATAGCGAGCGTTGTG R: TCGGCAITTTGTAGAGACTGG	161	105
<i>gusb</i>	Beta-glucuronidase	Exoglycosidase in lysosomes	XM_002407775	F: GACCGAGAGCACCATTCC R: AAGTCAAAGGGCACATTCTG	96	115

Primer sequence, amplicon length, and PCR efficiencies are listed.

Temporal expression stabilities over the blood-feeding duration obtained using geNorm software in each tissue (SG and SYN) for three biological replications (Rep. 1–3)

Table 2

Rank	SG				SYN			
	Rep. 1	Rep. 2	Rep. 3	Avg ± SD	Rep. 1	Rep. 2	Rep. 3	Avg ± SD
1/2	<i>tps4+113a</i> (0.24)	<i>tps4+113a</i> (0.31)	<i>tps4+113a</i> (0.44)	<i>tps4+113a</i> (0.33 ± 0.09)	<i>tps4+113a</i> (0.41)	<i>tps4+113a</i> (0.43)	<i>tps4+113a</i> (0.47)	<i>tps4+113a</i> (0.44 ± 0.03)
3	<i>gpdh</i> (0.47)	<i>gpdh</i> (0.5)	<i>tbp</i> (0.58)	<i>gpdh</i> (0.54 ± 0.09)	<i>tbp</i> (0.68)	<i>gpdh</i> (0.77)	<i>tbp</i> (0.7)	<i>gpdh</i> (0.88 ± 0.17)
4	<i>tbp</i> (0.83)	<i>tbp</i> (0.64)	<i>gpdh</i> (0.65)	<i>tbp</i> (0.68 ± 0.12)	<i>gapdh</i> (0.89)	<i>gapdh</i> (1.11)	<i>gpdh</i> (0.8)	<i>tbp</i> (0.93 ± 0.42)
5	<i>gapdh</i> (1.31)	<i>gapdh</i> (1.08)	<i>gapdh</i> (1.09)	<i>gapdh</i> (1.16 ± 0.12)	<i>gpdh</i> (1.08)	<i>tbp</i> (1.42)	<i>gapdh</i> (1.08)	<i>gapdh</i> (1.03 ± 0.12)

Increasing stability is expressed by smaller M values. The avg and SD for three biological replications are shown for each tissue.

Table 3

Temporal expression variations over the blood-feeding duration obtained using NormFinder software in each tissue (SG and SYN) for three biological replications (Rep. 1–3)

	SG				SYN			
	Rep. 1	Rep. 2	Rep. 3	Avg	Rep. 1	Rep. 2	Rep. 3	Avg
<i>Il3a</i>	0.02/0.27	-0.14/0.28	0.12/0.04	$3.3 \times 10^{-5}/0.2$	0.17/0.06	-0.06/0.51	-0.1/0.67	$3.3 \times 10^{-5}/0.42$
<i>rps4</i>	-0.07/0.33	0.08/0.09	-0.01/0.27	0/0.23	0.13/0.04	-0.13/0.009	0.003/0.1	$-2.8 \times 10^{-18}/0.05$
<i>gpdh</i>	0.51/0.68	0.11/0.19	-0.63/0.15	0/0.34	0.42/1.61	-0.3/0.92	-0.11/0.19	0/0.91
<i>thp</i>	0.01/0.49	-0.16/0.34	0.14/0.43	0/0.42	-0.4/0.39	0.39/2.94	0.005/0.38	$9.5 \times 10^{-18}/1.24$
<i>gapdh</i>	-0.48/4.05	0.1/3.16	0.37/3.25	0/3.49	-0.33/1.42	0.11/1.5	0.21/2.17	$-3.3 \times 10^{-5}/1.7$

The values are for intergroup variation/intragroup variation (intergroup variation, variation of the gene among the replications; intragroup variation, variation of the genes over the feeding duration within a replication).