# Reference gene selection and RNA preservation protocol in the cat flea, Ctenocephalides felis, for gene expression studies.

| | |
|---|---|
| Journal: | *Parasitology* |
| Manuscript ID | PAR-2016-0105.R1 |
| Manuscript Type: | Research Article - Standard |
| Date Submitted by the Author: | 02-May-2016 |
| Complete List of Authors: | McIntosh, Catriona; Univ of Aberdeen, Institute of Biological and Environmental Sciences<br>Baird, John; University of Aberdeen, School of Biological Sciences (Zoology)<br>Zinser, Erich; Zoetis, Global Parasitology Research<br>Woods, Debra; Zoetis, Global Parasitology Research<br>Campbell, Ewan; Univ of Aberdeen, School of Biological Sciences<br>Bowman, Alan; University of Aberdeen, Institute of Biological and Environmental Sciences; University of Aberdeen, School of Biological Sciences (Zoology) |
| Key Words: | Ctenocephalides felis, quantitative real-time PCR, normalization, RNA, gene expression, cat flea, RNA quality, RNA degradation |
| | |

# Reference gene selection and RNA preservation protocol in the cat flea, *Ctenocephalides felis*, for gene expression studies.

Authors: Catriona H McIntosh[a], John Baird[a], Erich Zinser[b], Debra J. Woods[b], Ewan M Campbell[a] , Alan S Bowman[a]

[a] Institute of Biological and Environmental Sciences, University of Aberdeen, Tillydrone Avenue, Aberdeen AB24 2TZ, UK

[b] Zoetis Inc., 333 Portage Street, Kalamazoo, MI 49007, USA

Running title: Reference genes for the cat flea.

**Corresponding Author:**

**Dr Alan S. Bowman**
**School of Biological Sciences (Zoology)**
**University of Aberdeen**
**Aberdeen**
**AB24 2TZ**

Tel: +44 1224 272877
Fax: +44 1224 272396
Email: a.bowman@abdn.ac.uk

## SUMMARY

The cat flea, *Ctenocephalides felis*, is a major pest species on companion animals thus of significant importance to the animal health industry. The aim of this study was to develop sampling and storage protocols and identify stable reference genes for gene expression studies to fully utilise the growing body of molecular knowledge of *C. felis*. RNA integrity was assessed in adult and larvae samples, which were either pierced or not pierced and stored in RNAlater at ambient temperature. RNA quality was maintained best in pierced samples, with negligible degradation evident after 10 days. RNA quality from non-pierced samples was poor within 3 days. Ten candidate reference genes were evaluated for their stability across four group comparisons (developmental stages, genders, feeding statuses and insecticide-treatment statuses). Glyceraldehyde 3 phosphate dehydrogenase (GAPDH), 60S ribosomal protein L19 (RPL19) and elongation factor-1α (Ef) were ranked highly in all stability comparisons, thus are recommended as reference genes under similar conditions. Employing just two of these three stable reference genes was sufficient for accurate normalization. Our results make a significant contribution to the future of gene expression studies in *C. felis*, describing validated sample preparation procedures and reference genes for use in this common pest.


Key words: *Ctenocephalides felis*, quantitative real-time PCR, normalization, RNA, gene expression, cat flea, RNA quality, RNA degradation

KEY FINDINGS

- Piercing *C. felis* samples in RNA-stabilizing solutions is critical to maintain RNA quality
- GAPDH, Elongation factor 1α and RPL19 are the most stable reference genes in *C. felis*
- The above three genes are stable across life stages, feeding status and insecticide treatment
- Employing just two of these three stable reference genes was sufficient for accurate normalization.
- Case study of vitellogenin expression demonstrates necessity to use multiple reference genes

## INTRODUCTION

The cat flea, *Ctenocephalides felis*, is one of the most common ectoparasites infesting companion animals worldwide and is of major importance to pet owners and the animal health industry (Rust and Dryden 1997; Beugnet et al. 2014). As well as irritation, cat flea infestations can trigger a severe allergic reaction in companion animals, known as flea allergy dermatitis (FAD), and act as a vector for several bacterial infections, most notably *Rickettsia felis*, and the parasitic worm *Dipylidium caninum* (Traversa 2013). For these reasons, and also the potential for current treatments to become ineffective, there is a constant need for more insight into this species. In recent years several cat flea expressed sequence tag (EST) and transcriptome studies have become available (Gaines et al. 2002; Ribeiro et al. 2012; Misof et al. 2014; Green et al. 2015), adding to a growing body of molecular knowledge that opens new opportunities for control. Techniques such as reverse-transcription quantitative real-time PCR (RT-qPCR) can be used to explore gene expression and this information can be used to find new ways to control *C. felis*.

RT-qPCR allows precise measurement of differential gene expression between samples. The sensitivity of the technique makes detection of small changes possible; however it also makes the results susceptible to the introduction of errors from experimental technique, such as differences in initial sample size, RNA extraction efficacy and reverse transcriptase enzyme efficiency during cDNA synthesis. To correct for these errors normalization is performed. Several normalization strategies can be used, such as accounting for the amount of total RNA, standardising sample size, or utilising internal reference genes, which are subject to conditions similar to the mRNA of interest (Huggett et al. 2005). Use of one or more endogenous reference genes has emerged as the preferred method for relative quantification and because they undergo the same processes as the mRNA of interest, reference genes can be used to correct for experimentally-introduced differences between samples (Derveaux et al. 2010). An ideal reference gene would be stably expressed across all experimental groups.

While normalization using endogenous reference genes is common, it is often the case that such reference genes are chosen without proper validation. Traditional “housekeeping” genes, such as β-Actin and glyceraldehyde 3-phosphate dehydrogenase (GAPDH), have frequently been used as reference genes for RT-qPCR without proper assessment of their suitability, largely due to their historic use as controls in less sensitive quantitative approaches such as Northern blotting (Boda et al. 2008). When tested, many commonly used control genes have

been shown to exhibit unstable expression across treatments under various conditions (Thellin et al. 1999). Several studies demonstrating the impact of unstable reference genes on the assessment of target gene expression levels have highlighted the need to validate reference genes for specific experimental design, cell and tissue type (Kidd et al. 2007; Boda et al. 2008; Kosir et al. 2010).

The aim of the current study was to develop procedures and tools for working with cat flea specimens at a molecular level. Understanding how storage can impact RNA integrity is vital for implementation of collaboration between research centres, allowing the transfer of reliable RNA between groups. Reliable reference genes are essential for robust gene expression studies (Bustin et al. 2009). Therefore the main tasks were to investigate how sample collection and storage procedures affect integrity of RNA that will be used in downstream gene expression studies and to screen and validate reference genes for use in RT-qPCR screens in the cat flea. Ten candidate reference genes in *C. felis* were assessed across the following 4 groups: developmental stage, sex, feeding status (fed versus unfed) and insecticide treatment-status (treated or untreated).

## MATERIALS AND METHODS

### *Insect rearing*

All *C. felis* samples were obtained from an artificially reared colony kept by Zoetis Inc (Kalamazoo, MI, USA), developed from fleas supplied by Elward II, California, USA, using methods similar to Kernif et al. (2015). Adults were fed *ad libitum* on bovine blood, after which eggs were collected three times per week and placed in containers with larval rearing media, consisting of 74 % finely ground laboratory canine diet, 25 % dried Brewer's yeast and 1 % part dried bovine blood, and fine sand. Larval containers were left undisturbed until emergence of adults approximately three weeks after egg collection. All life stages were reared in an insectary at ≈26 °C and 80 % relative humidity with a 12:12 L:D cycle.

### *Biological samples and cDNA synthesis*

Fed adult *C. felis* of mixed ages were collected from adult feeding chambers. Larvae and pupae were collected from culture pots approximately 7 and 12 days post-hatch, respectively. Unfed adults were collected approximately 30 days post-hatch (within 3 days of emergence from pupal case). For insecticide treatment, adults of mixed age were allowed to feed on 1

124 µM selamectin (Zoetis Inc, USA) in bovine blood for 24 hours prior to collection. Cat flea
125 samples were pierced once, centrally, with a 23 gauge needle, and groups of 10 placed
126 directly in 1 mL RNAlater (Life Technologies, ThermoFisher Scientific, Grand Island, NY,
127 USA) and kept at 4 °C overnight before storage at -80°C. Samples were sent to the University
128 of Aberdeen, UK, on dry ice. Prior to RNA extraction, pupae were removed from their cases
129 using 23 gauge needles. On the basis of size, females being larger than males, a subset of fed
130 adults were sorted into males and females.

131 For RNA extraction, pools of 3-10 fleas were removed from RNAlater and then homogenised
132 in 1 mL Tri-reagent (Sigma-Aldrich, UK) by crushing in 1.5 mL microfuge tubes with
133 micropestles. RNA was extracted according to manufacturer's instructions, with the phase
134 separation and ethanol washes repeated twice. RNA was resuspended in 8 µL (selamectin-
135 treated samples, as fewer fleas were available for RNA extraction) or 20 µL RNase-free $H_2O$
136 and quantified using a Nanodrop ND-1000 spectrophotometer (Thermo Scientific,
137 Loughborough, UK). RNA was treated with RQ1 DNase (Promega, UK) and 1 µg used as
138 template for cDNA synthesis with BioScript reverse transcriptase (Bioline Reagents Limited,
139 London, UK).

140 *Assessing influence of sampling procedure and storage conditions on RNA integrity*

141 Groups of 10 larvae and fed adults were either pierced once with a 23 gauge needle or not
142 pierced and placed in 1 mL RNAlater (Life Technologies). All samples were incubated at 4
143 °C overnight then stored at room temperature for 0, 3 or 10 days before being frozen at -80
144 °C until processing. RNA was extracted from groups of 10 fleas, as above. Total RNA
145 concentration was measured using a ND-1000 Nanodrop spectrophotometer (Thermo-
146 scientific) and RNA quality was assessed using an Agilent 2100 Bioanalyzer and Agilent
147 RNA 6000 Nano kit. Due to a hidden 18S/28S break in the rRNA of many arthropod species
148 (also apparent in *C. felis*) an accurate RNA Integrity Number (RIN) cannot be calculated
149 (Winnebeck et al. 2010). RNA integrity was therefore assessed by visual inspection of
150 electropherograms for each sample, assessing two replicates for each treatment. The time
151 points of 3 and 10 days were selected for study relates to the approximate time for
152 international courier by air (3 days) and international surface mail (10 days).

*Candidate reference gene selection & primer design*

Ten reference gene candidates were selected based on housekeeping genes previously used for RT-qPCR in the cat flea (Dreher-Lesnick et al. 2010) or transcripts commonly used as references in other insect species (Scharlaken et al. 2008; Li et al. 2013; Zhai et al. 2014; Tan et al. 2015). Ten candidate primer sets, representing transcripts from different functional classes, were initially assessed (Table 1). Sequences were obtained from annotated sequences in Ribeiro et al. (2012) (18S ribosomal RNA (18S), 28S ribosomal RNA (28S), 60S ribosomal protein (RPL19)), the BLAST Transcriptome Shotgun Assembly database (Elongation factor 1 α (Ef), Act (β-Actin)), or by using tBLASTn to search the cat flea EST database using *Drosophila melanogaster* sequences obtained from Flybase (Dos Santos et al. 2015) (GAPDH, Heat shock protein 22 (HSP22), NADH dehydrogenase/ ubiquinone reductase (NADH), α-Tubulin (αTub)). Primer3Plus (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi/) was used for primer design, implementing the qPCR setting and then checked manually. For comparison, a primer set targeting muscle-specific actin (DLAct) used in Dreher-Lesnick et al. (2010) was also included in the analysis. PCR was performed for each primer set using 25 µL BioMix Red (Bioline), 22 µL $H_2O$, 2 µL mixed *C. felis* cDNA and 1 µL 10 mM primer sets. Reactions were performed with the following conditions: 95 °C 5 min, 35 cycles of 95 °C 30 s, 58 °C 45 s, 72 °C 45 s, followed by incubation at 72 °C for 10 min. PCR products were electrophoresed in a 2 % agarose Tris-borate-EDTA (TBE) gel to confirm there was a single product of the expected size.

*Quantitative real-time PCR*

RT-qPCR was carried out in 96-well plates CFX96 Touch Real-Time PCR detection system (Bio-Rad Laboratories, USA). Reactions were run in 20 µL volumes (10 µL iTaq Sybr Green supermix (Bio-Rad), 1 µL 10 mM primer mix, 5 µL $H_2O$ and 4 µL template cDNA (1/20 dilution of cDNA produced from 1 µg DNase-treated RNA). PCR cycling conditions were: 95 °C 3min, 40 cycles of 95 °C 10 s, 58 °C 30 s followed by a melt-curve analysis step consisting of 0.5 °C incremental rises every 5 s, rising from 65°C to 95 °C. No template controls in duplicate were run for each primer set. Three replicates were run in triplicate for each treatment, except unfed and fed adults, where two and four replicates were used, respectively. Four-step 10-fold serial dilutions of mixed standard cDNA were performed in duplicate to create standard curves to calculate primer efficiencies. CFX manager software (version 3.1) (Biorad) was used to calculate efficiencies from a standard serial dilution curve.

Melt-curve analysis utilised CFX manager software to confirm correct product profiles for each primer set and Cq values extracted for further analysis.

*Data analysis*

Reference gene stability was assessed using three software programmes: geNorm (version 3.4) (Vandesompele et al. 2002), Normfinder (version 0.953) (Andersen et al. 2004) and Bestkeeper (version 1.0) (Pfaffl et al. 2004). Cq values were transformed using the delta-Ct method for analysis in GeNorm. For Normfinder, Cq values were transformed to a linear scale using the calculation $(2E)^{-Cq}$. Cq and efficiency values were input directly into Bestkeeper.

GeNorm ranks reference genes from most to least stable by calculating the gene expression stability M, the average pairwise variation of the expression ratio of a particular gene compared to all other tested genes (Vandesompele et al. 2002). Low M value is indicative of gene stability, with $M < 1.5$ necessary for utility as a reference gene. GeNorm gives two informative outputs. Firstly, a ranking of genes in order of stability based on calculation of average M for all genes and step-wise exclusion of the least stable gene and recalculation of the average M. Secondly, stability rankings generated from geNorm software can be used to assess the number of reference genes needed for accurate normalization, based on the pairwise variation ($V_n/V_{n+1}$) between sequential normalization factors, based on geometric means of the most stable genes which is recalculated following addition of each subsequent gene. The lowest number of genes giving $V_n/V_{n+1} < 0.15$ is the minimal number that should be used for normalization.

Normfinder utilises a model-based approach to assess reference gene stability, based on measures on intra- and inter-group variations, which are based on user-specified groupings (Andersen et al. 2004). This generates a stability value (SV) for each gene, as well as for the best combination of two reference genes. Low SV is indicative of gene stability, with $SV > 1$ suggesting a candidate is unstable and not suitable for use as a reference gene.

Bestkeeper uses input Cq and efficiency data to generate descriptive statistics for each gene, before generating a Bestkeeper index value (r) for each sample based on the geometric mean of its Cq values for each reference gene tested (Pfaffl et al. 2004). Stability can be assessed, based on standard deviation (SD) ± Cq and coefficient of variation. Only candidates where SD ± Cq is < 1 are suitable for use as reference genes.

Analysis was conducted in each of the programmes to assess reference genes most suitable for use in four groups: Developmental stages (larvae vs pupae vs unfed adults vs fed adults), Sexes (male vs female fed adults), Feeding statuses (fed vs unfed adults) and Treatment statuses (selamectin treated vs untreated fed adults). An overall ranking was produced using a points-based system to combine the rankings from all of the programmes used.

*Validation of reference genes – a case study in vitellogenin C expression*

Vitellogenins are key components of yolk in insect, produced in the fat body of adult females (Pan et al. 1969). Due to this function it is expected that levels of vitellogenin transcripts will be significantly higher in females than in males. The expected large difference made this a promising target to validate candidate reference genes for their utility in normalization. Primers were designed from an EST sequence representing vitellogenin C (Ribeiro et al. 2012), tested for specificity by melt-curve analysis and PCR followed by gel electrophoresis to confirm a single product of the expected size was produced (Table 1). The efficiency of this primer set was assessed by creating a standard curve using CFX Manager software (version 3.1) (Biorad) from duplicate 4-step 10-fold serial dilutions of mixed standard *C. felis* cDNA. RT-qPCR was performed to measure the expression of vitellogenin in samples from male (n = 3) and female (n = 3) fed adult *C. felis*, and normalized using the best single reference genes (GAPDH, Ef), best two reference genes (GAPDH + Ef), best three reference genes (GAPDH + Ef + RPL19) or least stable reference gene (18S) as listed in the overall ranking of reference genes for this comparison (Table 2). First the R0 for each sample was calculated for each gene for each sample using the equation $R0 = 1/(1+E)^{Cq}$, then the normalized values were calculated by dividing Vit R0 by the reference gene R0 or geometric mean of R0 for normalization with multiple reference genes.

RESULTS

*Impact of sample storage method on RNA quality*

The electropherograms for pierced larvae and adult samples are similar after 0, 3 and 10 days storage in RNAlater at room temperature (Fig. 1), with no appreciable accumulation of small RNA fragments visible. In contrast, degradation was clear in unpierced samples within 3 days, particularly in larvae samples (Fig. 1B). By day 10 at room temperature the majority of large RNA transcripts appeared to be fragmented, demonstrating RNA quality had dropped significantly.

*PCR efficiencies and expression levels of candidate reference genes*

Primer efficiencies ranged from 83.5-97.5 %, with most primer sets having efficiency greater than 90 %. The DLAct primers had a lower efficiency than preferable (83.5 %) and would have been discarded based on normal acceptable efficiency criteria. However, the DLAct primers were still used in reference gene testing for comparison due to their prior use in a publication (Dreher-Lesnick et al. 2010). NADH primers were not used for further analysis due to their highly variable efficiency (E = 90.7 %, SD = 16.7 %).

Cq values across all treatment samples (Mean ± SD, n = 30) for the 9 analysed reference genes ranged from 15.34 ± 1.65 (28S) to 22.44 ± 1.34 (α-Tubulin) (Fig. 2). GAPDH was the least variable reference gene tested across all samples (coefficient of variation (CV) = 3.45 %), while 28S was the most variable (CV = 10.93 %). Several genes (18S, 28S, DLAct) had clear outlying values, which suggested instability (Fig. 2).

*Expression stability of reference genes across developmental stages*

Three software programs were used to rank the nine candidate reference genes in *C. felis* for their stability across different developmental stages (larvae n = 3, pupae n = 3, unfed adults n = 2, fed adults n = 3; throughout the study n = number of pooled samples tested, each pool contained between 3 and 10 fleas) (Table 2). GeNorm ranked the genes based on their average expression stability (M), calculating this value with all genes included then removing the least stable gene and recalculating M until only two genes remain which cannot be further differentiated (Fig. 3). Ef and RPL19 were identified as the most stable genes by geNorm (M = 0.132) and 28S the least stable (M = 1.203) (Table 2, Fig. 3). However, all genes tested had M < 1.5 therefore can be considered stable enough to use as reference genes according to this analysis. A pairwise variation analysis between normalization factors $V_n/V_{n+1}$ was also performed by geNorm to assess the minimal number of reference genes needed for accurate normalization. Pairwise variation (V) < 0.15 indicates additional reference genes are unnecessary. For comparisons across all developmental stages V2/3 V = 0.048, indicating two reference genes are sufficient for normalization in this case (Fig. 4) and no significant benefit is gained by using > 2 reference genes.

The best gene determined by Normfinder analysis for comparisons between developmental *C. felis* groups was RPL19 (SV = 0.270) and the best combination of two genes was actin and GAPDH (SV = 0.210) (Table 2). HSP and 28S were found to be the least stable genes, with

SV > 1 suggesting they were unsuitable for use as reference genes in *C. felis* studies (Table 2).

Cq and efficiency values were input into Bestkeeper to produce descriptive statistics. The standard deviation ± Crossing Point (SD ± CP) can be used to rank stability. Under this criteria 18S was ranked as the most stable *C. felis* gene (SD ± CP = 0.54), followed by GAPDH (SD = 0.63) and Ef (SD = 0.76). HSP was the least stable gene (SD = 1.89) and considered too unstable for use as a reference gene as it had SD > 1.

The rankings for each program were combined using a points-based system to estimate an overall ranking of reference gene stability. This ranking found Ef, RPL19 and Act to be the most stable genes across *C. felis* developmental stages and 28S and HSP to be the least stable candidates (Table 2).

*Expression stability of reference genes across sexes*

Comparing the stability of candidate reference genes between male (n = 3) and female (n = 3) fed *C. felis* adults, geNorm ranked GAPDH and RPL19 as the most stable (M = 0.112) (Table 2, Fig. 3). 18S was the least stable gene based on this comparison; although all genes had M < 1.5 therefore can be considered as potentially suitable reference genes in *C. felis*. Pairwise comparison of normalization factors suggested two genes are sufficient for accurate normalization (V = 0.049) (Fig. 4). Normfinder ranked GAPDH as the most stable gene (SV = 0.144), Act and Ef as the best combination of two genes (SV = 0.111) and DLAct the least stable (SV = 0.510) (Table 2). DLAct was ranked as the most stable gene by Bestkeeper (SD = 0.51), while suggesting 18S, HSP and αTub are unsuitable as reference genes (SD > 1). The combined overall ranking placed GAPDH, Ef and RPL19 as the most stable candidate reference genes across *C. felis* and 18S as the least stable (Table 2).

*Expression stability of reference genes across feeding statuses*

GeNorm ranked Act and Ef as the most stable genes across feeding statuses (unfed adults n = 2, fed adults n = 4) (M = 0.112) (Table 2). Two genes were found to be sufficient for normalization (Fig. 4). 18S was the least stable gene according to both geNorm and Normfinder. Normfinder placed GAPDH as the most stable gene (SV = 0.092) and GAPDH and RPL19 to be the best combination of two genes (SV = 0.065). Bestkeeper estimated 28S and DLAct as the most and least stable genes respectively. Each candidate met the requirements to be classed as a suitable reference gene by all programs in this comparison.

The overall points-system ranking placed RPL19, GAPDH and Ef as the most stable reference genes across fed and unfed *C. felis* adults and 18S as the least stable candidate.

*Expression stability of reference genes across insecticide treatment statuses*

Stability of reference genes across treated (1 µM selamectin, n = 3) and untreated (n = 3) fed adult *C. felis* was investigated. Act and RPL19 were the most stable candidates according to geNorm (M = 0.104) (Table 2, Fig. 3). Bestkeeper also ranked these as the top two reference genes (Table 2). Two candidates were estimated to be sufficient for accurate normalization (V = 0.150) (Fig. 4). Ef (SV = 0.035) or a combination of Ef and αTub (SV = 0.042) were the best candidates according to Normfinder (Table 2). All programmes ranked 18S as the least stable gene across treatment statuses, with geNorm and Bestkeeper both classing it as unsuitable for use as a reference gene. Bestkeeper also found αTub, DLAct and 28S to be unsuitable candidates, perhaps because samples within this group account for several of the outliers seen in Fig. 2, which are likely to lead to a high standard deviation. The most stable genes in the overall ranking were Act, RPL19 and GAPDH, with 18S as the least stable candidate by this estimate.

*Validation of reference genes – a case study in vitellogenin C levels across sexes*

In all cases vitellogenin C was found to be upregulated in females relative to males, with reported fold-changes ranging from 8.46x to 12.32x (Fig. 5). Normalization with the two best reference genes individually led to disagreement in fold-change (GAPDH = 8.46x, Ef = 11.08x), whereas results were more consistent when using 2 or 3 reference genes in combination (9.69x ± 1.07 & 9.32x ±0.80) respectively). The coefficient of variation of the normalised fold change was much higher when using the least stable gene (18S) to normalise (37.98 %) compared to any of the combinations of single of multiple more stable genes, where the coefficient of variation ranged from 8.60-12.70 %.

## DISCUSSION

RNA samples are highly susceptible to breakdown from endogenous RNases following collection. RNAlater, a high density salt solution, acts to stabilise RNA by preventing action of such RNases. In order to work effectively RNAlater must enter tissues (Chen et al. 2007) but external structures, such as fine hairs on the surface of many arthropods, can prevent the solution from contacting internal tissues. Thus, it is often necessary to penetrate the sample tissues for proper exposure to RNAlater. Piercing individual cat fleas is a relatively laborious

process due to their small size and could dissuade some potential collaborators (e.g. veterinarians, kennel staff, the general public) from collecting fleas for downstream gene expression work. Thus, it was investigated if piercing is actually necessary for maintenance of RNA integrity by RNAlater. This study found that penetrating *C. felis* specimens is essential for preservation of RNA, with degradation clearly apparent in unpierced larvae and adult samples after even 3 days at room temperature (Fig. 1). However when specimens were pierced prior to submergence in RNAlater they could be stored at room temperature for up to 10 days with little degradation visible on electropherogram traces. A small peak at around 25 seconds was visible in pierced adult electropherograms after 3 and 10 days, representing small RNAs which could be indicative of a small amount of degradation. Thus, samples could be shipped at ambient temperature nationally and internationally for collaboration between research groups, if pierced upon collection and placed in RNAlater. Such an approach may be particularly useful when fleas are to be collected by veterinary practices or pet owners before being passed onto the research organisation. However if a particularly sensitive technique is to be utilized such as RNASeq it may still be beneficial to freeze samples before transportation on dry ice.

Reference genes which are stable across experimental conditions are essential to reliable interpretation of RT-qPCR data. Although several studies have used RT-qPCR to look at *R. felis* bacterial replication within the cat flea (Henry et al. 2007; Obhiambo et al. 2014), few have utilised the technique to study endogenous cat flea gene expression (Dreher-Lesnick et al. 2010). Past historical "housekeeping genes" have often been used in arthropod studies without proper validation. Recently, systematic screening of candidate reference genes has been performed for many insect species (Scharlaken et al. 2008; Teng et al. 2012; Li et al. 2013; Omondi et al. 2015; Shakeel et al. 2015), with many of these studies highlighting the importance of validating references in all experimental conditions and tissues of interest. In this study we systematically assessed ten candidate reference genes for stability within 4 groups of *C. felis*: developmental stages, sexes, feeding statuses and insecticide-treatment statuses. Transcripts commonly used in other insect species were selected for comparison (Scharlaken et al. 2008; Li et al. 2013; Zhai et al. 2014; Tan et al. 2015).

Three programs were used to estimate the stability of the candidate reference genes, geNorm, Normfinder and Bestkeeper. Each program uses a different algorithm to assess stability,

leading to differences in the rankings between programmes. This was particularly apparent for Bestkeeper in this study, which often highlighted as the best gene a candidate which was ranked low by other programmes (Table 2). To give an easy guide to stable reference genes an overall ranking was produced for each comparison. This overall ranking showed GAPDH, RPL19 and Ef to rank highly in all comparisons (Table 2). Ribosomal proteins, GAPDH and Ef have all been characterised recently as stable reference genes in other arthropod species (Scharlaken et al. 2008; Teng et al. 2012; Li et al. 2013; Omondi et al. 2015; Shakeel et al. 2015). While it is important to assess stability of references in specific experimental conditions, GAPDH, RPL19 and Ef would be recommended as reference genes for any of the comparisons tested here in *C. felis*.

The use of unstable reference genes can have a large impact on the interpretation of RT-qPCR results (Kidd et al. 2007; Boda et al. 2008; Kosir et al. 2010). To validate the ranking of our candidate genes levels of vitellogenin C in male and female fed adult *C. felis* were investigated, using the best three (GAPDH + Ef + RPL19), two (GAPDH + Ef) or single (GAPDH, Ef) genes and the least stable (18S). Vitellogenin C levels were found to be approximately 9-fold higher in females compared to males. While all normalization strategies demonstrated an increase in vitellogenin C in females, the estimated fold change varied from 8.5-fold to 12.3-fold (Fig. 5). Using the least stable gene for normalization gave a high coefficient of variation (37.98 %) compared to the other normalization strategies (CV 8.60-12.70 %), demonstrating the uncertainty introduced by use of an inappropriate reference gene. This is particularly important when trying to detect small changes in gene expression between samples, where instability of a reference gene could lead to misinterpretation of results (Omondi et al. 2015). Use of two or three reference genes generated a more consistent fold change estimate (9.69-fold & 9.32-fold respectively), with single reference genes generating different estimates (GAPDH = 8.46x, Ef = 11.08x). This highlights the importance of using multiple reference genes for accurate normalization.

The present study provides insight into sample preparation and reference genes suitable for use across a variety of conditions for *C. felis* specimens. In summary, our findings recommend piercing of *C. felis* before placing in an RNA-stabilizing solution and storing at room temperature and that two reference genes selected from GAPDH, Ef and RPL19 are suitable and suffice for accurate gene expression studies in *C. felis* in the given experimental

conditions. This paves the way for new investigations into *C. felis* gene expression, opening new avenues for the research community to utilise to find ways to tackle this common pest.

ACKNOWLEDGEMENTS

Many thanks to Zoetis Inc for providing the cat flea specimens for this analysis.

FINANCIAL SUPPORT

This work was supported by a Knowledge Transfer Network BBSRC Industrial Case (# BB/L502467/1) studentship in association Zoetis Inc.

CONFLICT OF INTEREST

None

## REFERENCES

**Andersen C. L., Jensen J. L. and Orntoft T. F.** (2004). Normalization of real-time quantitative reverse transcription-PCR data: a mode-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Research* **64**, 5245-5250. doi:10.1158/0088-5472.CAN-04-0496

**Beugnet F., Bourdeau P., Chalvet-Monfray K., Cozma V., Farkas R., Guillot J., Halos L., Joachim A., Losson B., Miro G., Otranto D., Renaud M. and Rinaldi L.** (2014). Parasites of domestic cats in Europe: co-infestations and risk factors. *Parasites & Vectors* **7**: 291. doi: 10-1186/1756-3305-7-291.

**Boda E., Pini A., Hoxha E., Parolisi R. and Tempia F.** (2008). Selection of reference genes for quantitative real-time RT-PCR studies in mouse brain. *Journal of Molecular Neuroscience* **37**, 238-253. doi: 10.1007/s12031-008-9128-9

**Bustin S. A., Benes V., Garson J. A., Hellemans J., Huggett J., Kubista M., Mueller R., Nolan T., Pfaffl M. W., Shipley G. L., Vandescompele J. and Wittwer C. T.** (2009). The MIQE Guidelines: *M*inimum *I*nformation for Publication of *Q*uantitative Real-Time PCR *E*xpermiments. *Clinical Chemistry* **55**, 611-622. doi: 10.1373/clinchem.2008.112797

**Chen Y., Evans J., Hamilton M. and Feldlaufer M.** (2007). The influence of RNA integrity on the detection of honey bee viruses: Molecular assessment of different sample storage methods. *Journal of Apicultural Research* **46**, 81-87. doi: 10.1080/00218839.2007.11101372

**Derveaux S., Vandesompele J. and Hellemans J.** (2010). How to do successful gene expression analysis using real-time PCR. *Methods* **50**, 227-230. doi: 10.1016/j.ymeth.2009.11.001

**Dos Santos G., Schroeder A. J., Goodman J. L., Strelets V. B., Crosby M. A., Thurmond J., Emmert D. B. and Gelbart W. M.** (2015). FlyBase: Introduction of the *Drosophila melanogaster* release 6 reference genome assembly and large-scale migration of genome annotations. *Nucleic Acids Research* **43**, D690-D697. doi: 10.1093/nar/gku1099

**Dreher-Lesnick S. M., Ceraul S. M., Lesnick S. C., Gillespie J. J., Anderson J. M., Jochim R. C., Valenzuela J. G. and Azad A. F.** (2010). Analysis of *Rickettsia typhi*-infected and uninfected cat flea (*Ctenocephalides felis*) midgut cDNA libraries: deciphering molecular pathways involved in host response to *R. typhi* infection. *Insect Molecular Biology* **19**, 229-241. doi: 10.1111/j.1365-2583.2009.00978.x

**Gaines P. J., Brandt K. S., Eisele A. M., Wagner W. P., Bozic C. M. and Wisnewski N.** (2002). Analysis of expressed sequence tags from subtracted and unsubtracted

*Ctenocephalides felis* hindgut and Malpighian tubule cDNA libraries. *Insect Molecular Biology* **11**, 299-306. doi: 10.1046/j.1365-2583.2002.00337.x

**Greene W.K., Macnish M.G., Rice K.L. and Thompson, R.C.A.** (2015). Identification of genes associated with blood feeding in the cat flea,*Ctenocephalides felis. Parasites and Vectors* **8**, 368. doi: 10.1186/s13071-015-0972-5

**Henry K. M., Jiang J., Rozmajzl P. J., Azad A. F., Macaluso K. R. and Richards A. L.** (2007). Development of quantitative real-time PCR assays to detect *Rickettsia typhi* and *Rickettsia felis*, the causative agents of murine typhus and flea-borne spotted fever. *Molecular and Cellular Probes* **21**, 17-23.

**Huggett J., Dheda K., Bustin S. and Zumla A.** (2005). Real-time RT-PCR normalisation; strategies and considerations. *Genes and Immunity* **6**, 279-284. doi: 10.1038/sj.gene.6364190

**Kernif T., Stafford K., Coles G. C., Bitam I., Papa K., Chiaroni J., Raoult D. and Parola P.** (2015). Responses of artificially reared cat fleas *Ctenocephalides felis felis* (Bouche, 1835) to different mammalian bloods. *Medical and Veterinary Entomology* **29**, 171-177. doi: 10.1111/mve.12100

**Kidd M., Nadler B., Mane S., Eick G., Malfertheiner M., Champaneria M., Pfragner R. and Modlin I.** (2007). GeneChip, geNorm, and gastrointestinal tumors: Novel reference genes for real-time PCR. *Physiological Genomics* **30**, 363-370. doi: 10.1152/physiolgenomics.00251.2006

**Kosir R., Acimovic J., Golicnik M., Perse M., Majdic G., Fink M. and Rozman D.** (2010). Determination of reference genes for circadian studies in different tissues and mouse strains. *BMC Molecular Biology* **11**:60. doi: 10.1186/1471-2199-11-60

**Li R., Wang S., Wu Q., Yang N., Yang X., Pan H., Zhou X., Bai L., Xu B., Zhou X. and Zhang Y.** (2013). Reference gene selection for qRT-PCR analysis in the sweetpotato whitefly, *Bemisia tabaci* (Hemiptera: Aleyrodidae). *PLoS ONE* **8**, e53006. doi: 10.1371/journal.pone.0053006

**Misof B., Liu S., Meusemann K., Peters R. S., Donath A., Mayer C., Frandsen P. B., Ware J., Flouri T., Beutel R. G., et al.** (2014). Phylogenomics resolves the timing and pattern of insect evolution. *Science* **346**, 763-767. doi: 10.1126/science.1257570

**Odhiambo A. M., Maina A. N., Taylor M. L., Jiang J. and Richards A. L.** (2014). Development and validation of a quantitative real-time polymerase chain reaction assay specific for the detection of *Rickettsia felis* and not *Rickettsia felis*-like organisms. *Vector-Borne and Zoonotic Diseases* **14**, 476-481. doi: 10.1089/vbz.2013.1518

**Omondi B. A., Latorre-Estivalis J. M., Oliveira I. H. R., Ignell R. and Lorenzo M. G.** (2015). Evaluation of reference genes for insect olfaction studies. *Parasites & Vectors* **8**: 243. doi: 10.1186/s13071-015-0862-x

**Pan M. L., Bell W. J. and Telfer W. H.** (1969). Vitellogenic blood protein synthesis by insect fat body. *Science* **165**, 393-394.

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

**Ribeiro J. M. C., Assumpção T. C. F., Ma D., Alvarenga P. H., Pham V. M., Andersen J. F., Francischetti I. M. B. and Macaluso K. R.** (2012). An insight into the

497 sialotranscriptome of the cat flea, *Ctenocephalides felis*. *PLoS ONE* **7**, e44612. doi:
498 doi:10.1371/journal.pone.0044612

499 **Rust M. K. and Dryden M. W.** (1997). The biology, ecology, and management of the cat
500 flea. *Annual Review of Entomology* **42**, 451-473.

501 **Scharlaken B., de Graaf D., Goossens K., Brunain M., Peelman L. J. and Jacobs F. J.**
502 (2008). Reference gene selection for insect expression studies using quantitative real-time
503 PCR: The head of the honeybee, *Apis mellifera*, after a bacterial challenge. *Journal of Insect*
504 *Science* **8**, 33. doi: 10.1673/031.008.3301

505 **Shakeel M., Zhu X., Kang T., Wan H. and Li J.** (2015). Selection and evaluation of
506 reference genes for quantitative gene expression studies in cotton bollworm, *Helicoverpa*
507 *armigera* (Lepidoptera: Noctuidae). *Journal of Asia-Pacific Entomology* **18**, 123-130. doi:
508 10.1016/j.aspen.2015.01.001

509 **Tan Q. Q., Zhu L., Li Y., Liu W., Ma W. H., Lei C. L. and Wang X. P.** (2015). A *de Novo*
510 transcriptome and valid reference genes for quantitative real-time PCR in *Colaphellus*
511 *bowringi*. *PLoS ONE* **10**, e0118693. doi: 10.1371/journal.pone.0118693

512 **Teng X., Zhang Z., He G., Yang L. and Li F.** (2012). Validation of reference genes for
513 quantitative expression analysis by real-time RT-PCR in four lepidopteran insects. *Journal of*
514 *Insect Science* **12**:60. doi:10.1673/031.012.6001

515 **Thellin O., Zorzi W., Lakaye B., De Borman B., Coumans B., Hennen G., Grisar T.,**
516 **Igout A. and Heinen E.** (1999). Housekeeping genes as internal standards: Use and limits.
517 *Journal of Biotechnology* **75**, 291-295.

**Traversa D.** (2013). Fleas infesting pets in the era of emerging extra-intestinal nematodes. *Parasites and Vectors* **6**:59. doi: 10.1186/1756-3305-6-59

**Vandesompele J., De Preter K., Pattyn F., Poppe B., Van Roy N., De Paepe A. and Speleman F.** (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology* **3**, RESEARCH0034.

**Winnebeck E. C., Millar C. D. and Warman G. R.** (2010). Why does insect RNA look degraded? *Journal of Insect Science* **10**:159. doi: insectscience.org/10.159.

**Zhai Y., Lin Q., Zhou X., Zhang X., Liu T. and Yu Y.** (2014). Identification and validation of reference genes for quantitative real-time PCR in *Drosophila suzukii* (Diptera: Drosophilidae). *PloS ONE* **9**, e106800. doi: 10.1371/journal.pone.0106800.

532 **Table 1. Candidate reference genes assessed for stability**.

533 [a] Tm, melting temperature for oligos; [b] E, efficiency of primers, assessed by standard curve

534 slope. SD calculated for efficiencies between runs.

| Transcript Name | Function | Oligo Sequence | Product size (bp) | Tm (°C)[a] | E ±SD (%)[b] |
|---|---|---|---|---|---|
| 18S ribosomal RNA gene (18S) | Structural protein in ribosome | F: CCTGCGGCTTAATTTGACTC<br>R: AGACAAATCGCTCCACCAAC | 135 | 59.8<br>60.1 | 94.1±5.1 |
| 28S ribosomal RNA gene (28S) | Structural protein in ribosome | F: AAACGGTCCTTGTGACTTGG<br>R: TCTGAGCTGACCGTTGAATG | 136 | 60<br>60 | 92.7±2.6 |
| β-Actin (Act) | Cytoskeletal structure | F: AGGAATTGCTGACCGTATGC<br>R: TTGGAAGGTGGATAGGGATG | 139 | 60.1<br>59.7 | 97.5±1.8 |
| Muscle specific actin (DLAct) | Cytoskeletal structure | F: GGTCGGTATGGGACAAAAGGAC<br>R: GTAGATTGGGACGGTGTGAGAGAC | 367 | 59.9<br>62.3 | 83.5±1.3 |
| Elongation factor 1 α (Ef) | Translation | F: TCGTACTGGCAAATCCACAG<br>R: CATGTCACGGACAACGAAAC | 145 | 59.7<br>60 | 95.3±4.8 |
| Glyceraldehyde 3 phosphate dehydrogenase (GAPDH) | Glycolysis | F: ACCCAAAAGACTGTGGATGG<br>R: CGGAATGACTTTGCCTACAG | 117 | 59.8<br>58.4 | 91.4±1.5 |
| Heat shock protein 22 (HSP) | Stress response | F: ACCCAATGCGTCTTATGGAC<br>R: TAATAACCGCCACGGAAGAG | 103 | 59.8<br>60.1 | 93.7±2.3 |
| NADH dehydrogenase/ ubiquinone reductase (NADH) | Respiratory chain enzyme | F: GTCGCTGGTGTAGATGATCTTG<br>R: TTCGACGTTAAGCACCACAG | 133 | 59.8<br>59.9 | 90.7±16.7 |
| 60S ribosomal protein L19 (RPL19) | Structural protein in ribosome | F: TACAGCTAATGCCCGTACACC<br>R: TTCAACAAACGCCTCAGGAC | 72 | 60<br>61.2 | 91.7±1.1 |
| α-Tubulin (αTub) | Cytoskeletal structure | F: AACTATTGGAGGCGGTGATG<br>R: TTGACGGTATGTTCCAGTGC | 125 | 60<br>59.6 | 91.6±3.4 |
| Vitellogenin | Reproduction | F: CAAGAATCCAGCTCCTCCAG<br>R: ACGGATGCTGAAGCAGAGTT | 204 | 59.9<br>60 | 91.2±2.0 |

535

**Table 2. Rankings of candidate reference genes**

M = average expression stability (geNorm), SV = stability value (Normfinder), SD ± CP = standard deviation ± crossing point (Bestkeeper), * = not considered a suitable reference gene by this programme. Overall ranking is a based on a points-based system to combine the rankings from all programmes used. All rankings are stated from most stable (1) to least stable (10).

**Developmental Stages**

| Ranking | GeNorm | M | Normfinder | SV | Bestkeeper | SD ± CP | Overall Ranking |
|---|---|---|---|---|---|---|---|
| **1** | Ef / RPL19 | 0.132 | RPL19 | 0.27 | 18S | 0.54 | Ef |
| **2** | - | - | Ef | 0.276 | GAPDH | 0.63 | RPL19 |
| **3** | Act | 0.149 | Act | 0.29 | Ef | 0.76 | Act |
| **4** | aTub | 0.257 | GAPDH | 0.318 | Act | 0.8 | GAPDH |
| **5** | GAPDH | 0.441 | aTub | 0.365 | RPL19 | 0.88 | 18S |
| **6** | DLAct | 0.626 | DLAct | 0.558 | DlAct | 0.91 | aTub |
| **7** | 18S | 0.845 | 18S | 0.902 | 28S | 0.97 | DlAct |
| **8** | HSP | 1.034 | HSP | 1.054* | aTub | 0.98 | 28S / HSP |
| **9** | 28S | 1.203 | 28S | 1.145* | HSP | 1.89* | - |

**Males vs Females**

| Ranking | GeNorm | M | Normfinder | SV | Bestkeeper | SD ± CP | Overall Ranking |
|---|---|---|---|---|---|---|---|
| **1** | GAPDH / RPL19 | 0.114 | GAPDH | 0.144 | DlAct | 0.51 | GAPDH |
| **2** | - | - | Ef | 0.188 | Ef | 0.54 | Ef |
| **3** | Ef | 0.142 | Act | 0.197 | RPL19 | 0.61 | RPL19 |
| **4** | Act | 0.196 | RPL19 | 0.216 | GAPDH | 0.7 | Act |
| **5** | HSP | 0.408 | HSP | 0.412 | 28S | 0.76 | 28S / DLAct / HSP |
| **6** | aTub | 0.497 | 28S | 0.414 | Act | 0.78 | - |
| **7** | 28S | 0.629 | aTub | 0.469 | 18S | 1.21* | - |
| **8** | DLAct | 0.73 | 18S | 0.491 | HSP | 1.23* | aTub |
| **9** | 18S | 0.83 | DLAct | 0.51 | aTub | 1.34* | 18S |

**Fed Adults vs Unfed Adults**

| Ranking | GeNorm | M | Normfinder | SV | Bestkeeper | SD ± CP | Overall Ranking |
|---|---|---|---|---|---|---|---|
| **1** | Act / Ef | 0.112 | GAPDH | 0.092 | 28S | 0.27 | RPL19 |
| **2** | - | - | RPL19 | 0.147 | GAPDH | 0.29 | Ef / GAPDH |
| **3** | RPL19 | 0.141 | aTub | 0.16 | RPL19 | 0.42 | - |
| **4** | HSP | 0.163 | Ef | 0.206 | Ef | 0.47 | aTub |
| **5** | aTub | 0.195 | DLAct | 0.23 | HSP | 0.51 | Act |
| **6** | GAPDH | 0.221 | Act | 0.236 | aTub | 0.52 | HSP |
| **7** | DLAct | 0.266 | HSP | 0.241 | Act | 0.54 | 28S |
| **8** | 28S | 0.383 | 28S | 0.431 | 18S | 0.57 | DLAct |
| **9** | 18S | 0.489 | 18S | 0.491 | DLAct | 0.71 | 18S |

**Insecticide Treated vs Untreated**

| Ranking | GeNorm | M | Normfinder | SV | Bestkeeper | SD ± CP | Overall Ranking |
|---|---|---|---|---|---|---|---|
| **1** | Act / RPL19 | 0.104 | Ef | 0.035 | Act | 0.22 | Act / RPL19 |
| **2** | - | - | aTub | 0.075 | RPL19 | 0.25 | - |
| **3** | GAPDH | 0.337 | HSP | 0.135 | GAPDH | 0.45 | Ef |
| **4** | HSP | 0.459 | DLAct | 0.138 | Ef | 0.63 | GAPDH |
| **5** | Ef | 0.502 | GAPDH | 0.252 | HSP | 0.66 | HSP |
| **6** | aTub | 0.597 | RPL19 | 0.408 | aTub | 1.07* | aTub |
| **7** | DLAct | 0.734 | Act | 0.416 | DLAct | 1.45* | DLAct |
| **8** | 28S | 1.115 | 28S | 0.581 | 28S | 2.21* | 28S |
| **9** | 18S | 1.533* | 18S | 0.952 | 18S | 2.82* | 18S |

**Fig. 1 Electopherogram assessment of RNA quality for pierced and unpierced C. felis specimens stored at room temperature for 0, 3 or 10 days.**

Total RNA was extracted from pierced or unpierced *C. felis* larvae (Fig. 1A) or adults (Fig. 1B) which had been stored in RNAlater at room temperature for 0, 3 0r 10 days. 40-120 ng of RNA were run on the Agilent Bioanalyzer 2100 microfluidics gel analysis platform to determine RNA quality.

**Figure 2. Average cycle thresholds of candidate reference genes.**

The boxplot represents median, and indicates 25th and 75th percentile. Whiskers represent the 10th and 90th percentiles. Black dots indicate outliers. Cq values for all tested samples (n = 30) across all groups (*C. felis* developmental groups, sexes, feeding statuses and insecticide treatment statuses).

**Figure 3. Average expression stability of candidate reference genes**

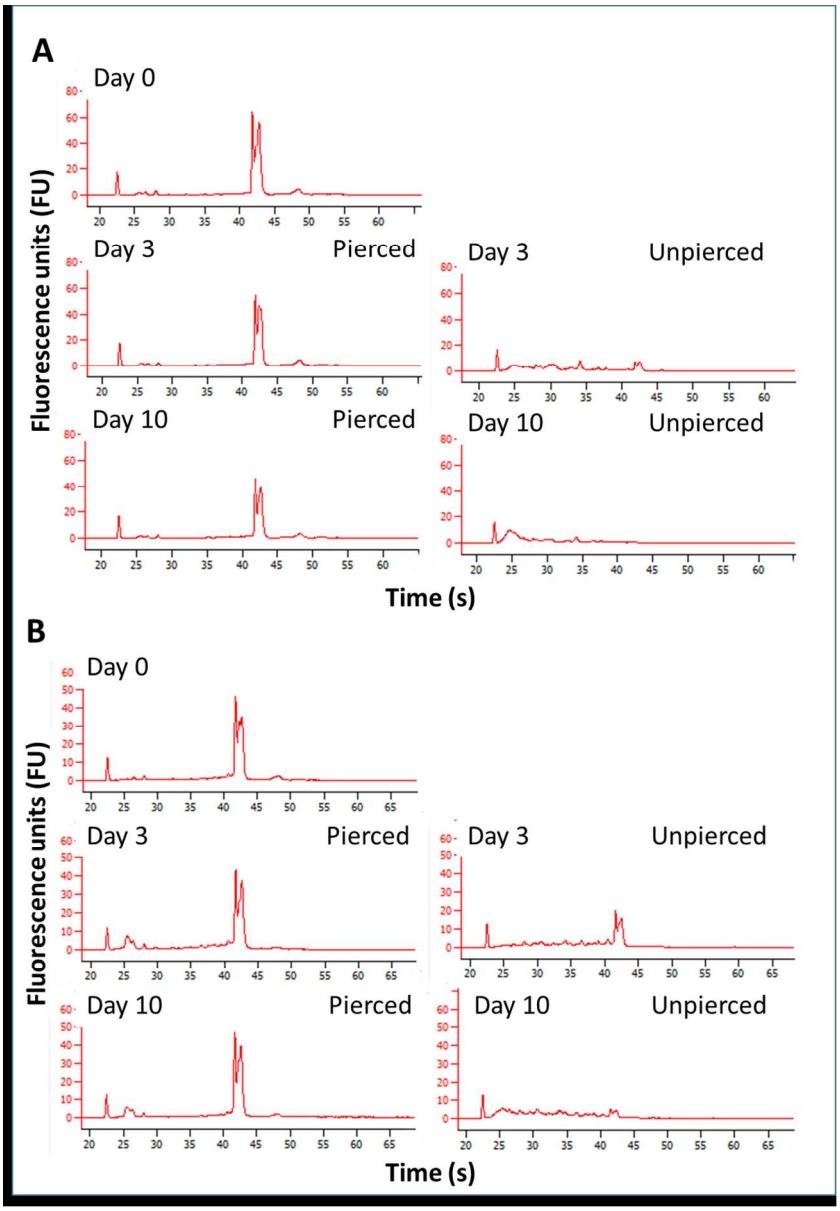
Values for average expression stability (M) as calculated by geNorm (v. 3.4) by pairwise comparison and stepwise exclusion of the least stable reference gene, for four group comparisons: (A) *C. felis* developmental stages (larvae, pupae, unfed adults, fed adults); (B) feeding statuses (unfed and fed adults); (C) sexes (males and females); (D) Insecticide-treatment statuses (treated with 1 µM selamectin or untreated fed adults).

**Figure 4. Pairwise variation values for assessment of necessary number of reference genes**

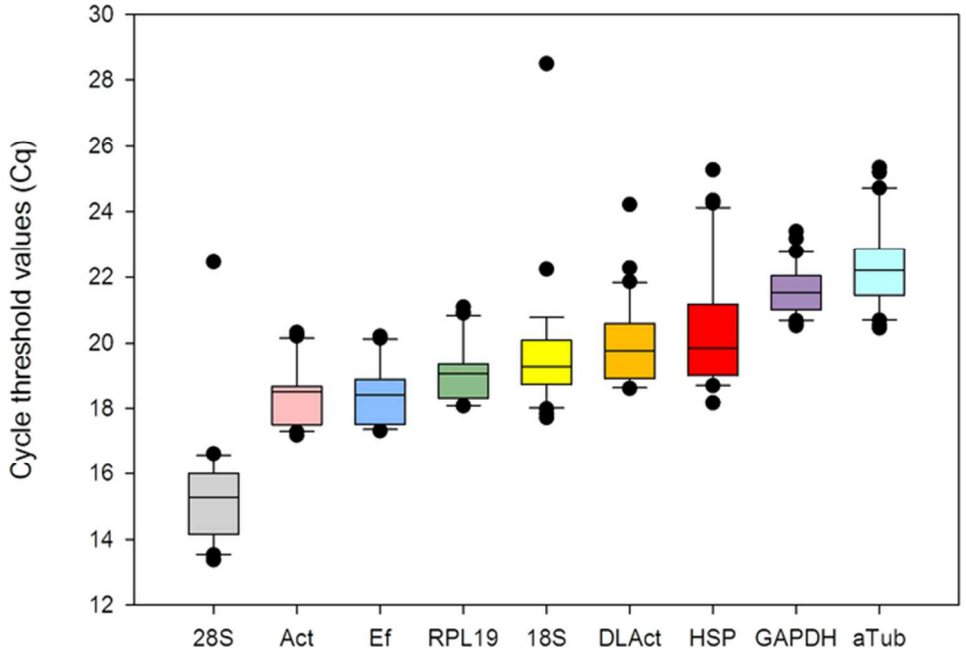
geNorm (v. 3.4) calculated pairwise variation (V) for assessment of sufficient number of reference genes for accurate normalization in each of four group comparisons of *C. felis*: developmental stages (larvae, pupae, unfed adults, fed adults); feeding statuses (unfed and fed adults); sexes (males and females); insecticide-treatment statuses (treated with 1 µM selamectin or untreated fed adults). $V < 0.15$ indicates inclusion of a further reference gene is of negligible benefit.

**Figure 5. Validation of reference genes by testing vitellogenin C expression levels.**

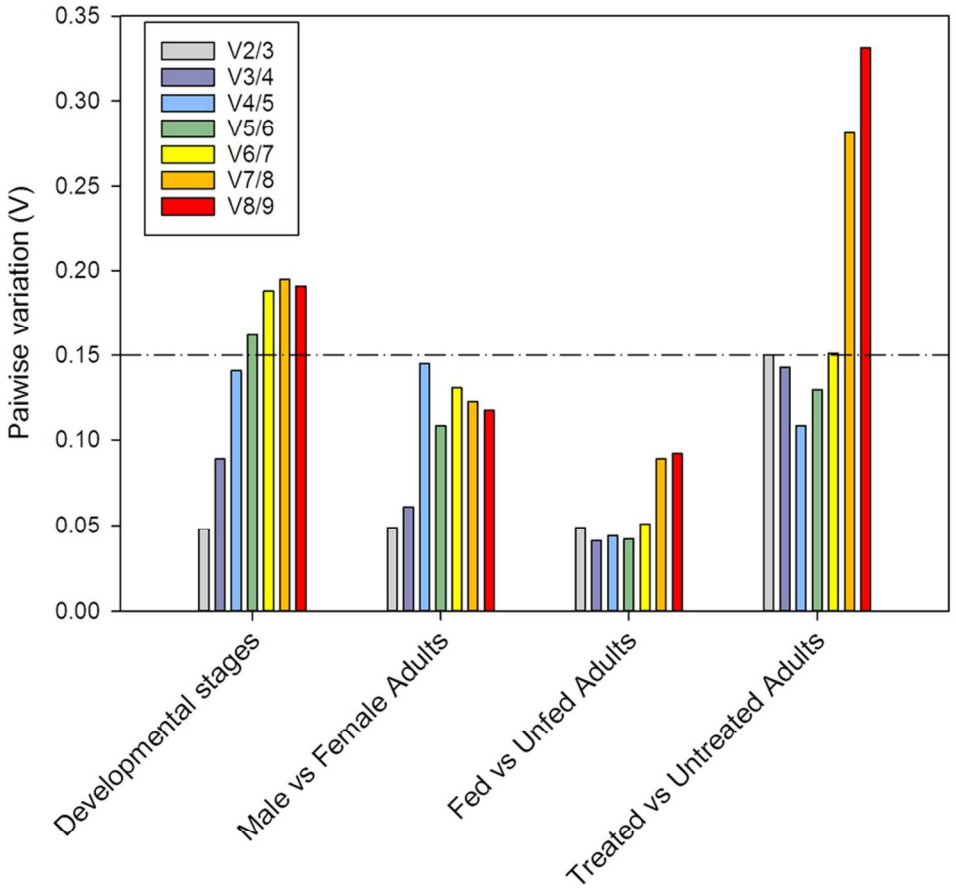
Vitellogenin C levels were measured in female relative to male fed *C. felis* adults. Vitellogenin C expression levels were assessed relative to single most (GAPDH, Ef) or least (18S) stable reference genes or combinations of the two (Ef + GAPDH) or three (Ef + GAPDH + RPL19) most stable reference genes. Data are means +/- SEM, n = 3.



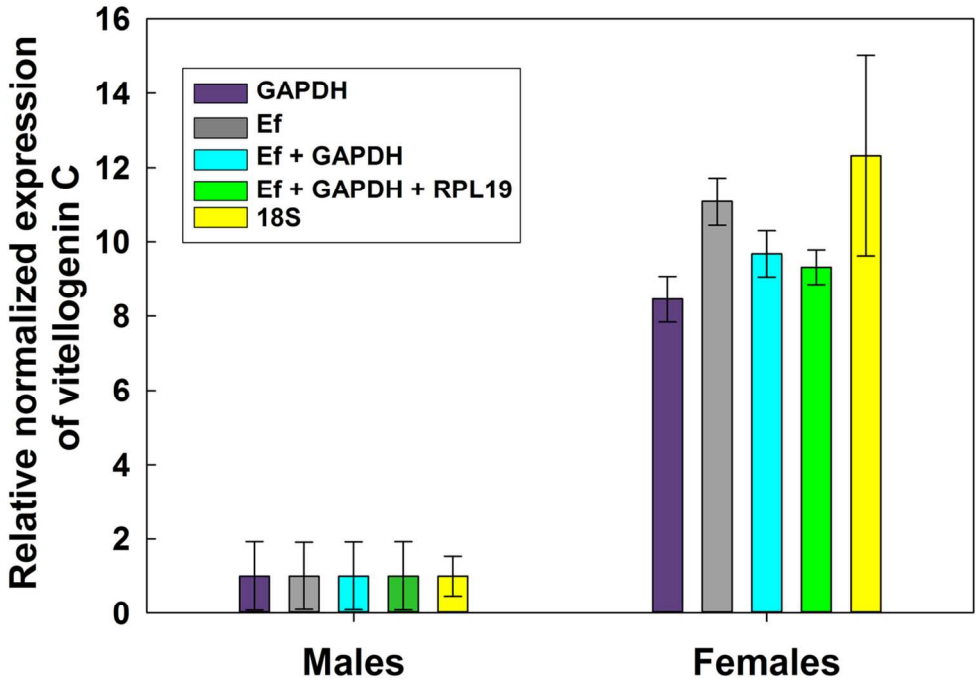

401x578mm (72 x 72 DPI)




59x44mm (300 x 300 DPI)




80x80mm (300 x 300 DPI)




57x41mm (600 x 600 DPI)