Selection of Reference Genes for Expression Analysis Using Quantitative Real-Time PCR in the Pea Aphid, *Acyrthosiphon pisum* (Harris) (Hemipter, Aphidiae)

Chunxiao Yang  
*University of Kentucky*, chunxiaoyang@uky.edu

Huipeng Pan  
*University of Kentucky*, huipengpan@uky.edu

Yong Liu  
*Institute of Plant Protection, China*

Xuguo Zhou  
*University of Kentucky*, xuguozhou@uky.edu

Follow this and additional works at: [https://uknowledge.uky.edu/entomology_facpub](https://uknowledge.uky.edu/entomology_facpub)

Part of the Entomology Commons

Right click to open a feedback form in a new tab to let us know how this document benefits you.

Repository Citation  
Yang, Chunxiao; Pan, Huipeng; Liu, Yong; and Zhou, Xuguo, "Selection of Reference Genes for Expression Analysis Using Quantitative Real-Time PCR in the Pea Aphid, *Acyrthosiphon pisum* (Harris) (Hemiptera, Aphidiae)" (2014). *Entomology Faculty Publications*. 89.  
[https://uknowledge.uky.edu/entomology_facpub/89](https://uknowledge.uky.edu/entomology_facpub/89)

This Article is brought to you for free and open access by the Entomology at UKnowledge. It has been accepted for inclusion in Entomology Faculty Publications by an authorized administrator of UKnowledge. For more information, please contact UKnowledge@lsv.uky.edu.
Selection of Reference Genes for Expression Analysis Using Quantitative Real-Time PCR in the Pea Aphid, *Acyrthosiphon pisum* (Harris) (Hemiptera, Aphidiae)

Digital Object Identifier (DOI)
http://dx.doi.org/10.1371/journal.pone.0110454

Notes/Citation Information
Published in *PLOS One*, v. 9, no. 11, article e110454, p. 1-10.

© 2014 Yang et al.

This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

This article is available at UKnowledge: https://uknowledge.uky.edu/entomology_facpub/89
Selection of Reference Genes for Expression Analysis Using Quantitative Real-Time PCR in the Pea Aphid, *Acyrthosiphon pisum* (Harris) (Hemiptera, Aphidiae)

Chunxiao Yang<sup>1,2,*</sup>, Huipeng Pan<sup>2,*</sup>, Yong Liu<sup>1,*</sup>, Xuguo Zhou<sup>2,*</sup>

1. Hunan Academy of Agricultural Sciences, Institute of Plant Protection, Changsha, Hunan, China, 2. Department of Entomology, University of Kentucky, Lexington, Kentucky, United States of America

*haosliu@163.com (YL); xuguozhou@uky.edu (XGZ)

These authors contributed equally to this work.

Abstract

To facilitate gene expression study and obtain accurate qRT-PCR analysis, normalization relative to stable expressed housekeeping genes is required. In this study, expression profiles of 11 candidate reference genes, including *actin* (*Actin*), *elongation factor 1a* (*EF1A*), *TATA-box-binding protein* (*TATA*), *ribosomal protein L12* (*RPL12*), *β-tubulin* (*Tubulin*), *NADH dehydrogenase* (*NADH*), *vacuolar-type H<sup>+</sup>-ATPase* (*v-ATPase*), *succinate dehydrogenase B* (*SDHB*), *28S ribosomal RNA* (*28S*), *16S ribosomal RNA* (*16S*), and *18S ribosomal RNA* (*18S*) from the pea aphid *Acyrthosiphon pisum*, under different developmental stages and temperature conditions, were investigated. A total of four analytical tools, *geNorm*, *Normfinder*, *BestKeeper*, and the *ΔCt* method, were used to evaluate the suitability of these genes as endogenous controls. According to *RefFinder*, a web-based software tool which integrates all four above-mentioned algorithms to compare and rank the reference genes, *SDHB*, *16S*, and *NADH* were the three most stable house-keeping genes under different developmental stages and temperatures. This work is intended to establish a standardized qRT-PCR protocol in pea aphid and serves as a starting point for the genomics and functional genomics research in this emerging insect model.
Introduction

Quantitative real-time PCR (qRT-PCR) is a rapid and reliable method for the detection and quantification of gene expression levels during different biological processes [1]. Although qRT-PCR is often described as the gold standard, there are still some limitations of this assay such as RNA quality and quantity, reverse transcription and normalization, and efficiency of PCR reaction can influence threshold cycle (Ct) values [2, 3]. A common technique in qRT-PCR is to normalize data by measuring in parallel the expression of a reference gene from the same samples. Using housekeeping genes as a reference is the most widely adopted approach [1]. Housekeeping genes are believed to possess inherent stable and constitutive expression irrespective of physiological conditions in different samples or treatments under investigation [4, 5]. Several reports have demonstrated that some commonly used reference genes differentially expressed under different treatments or conditions [4–8]. In fact, no reference genes are stably expressed and suitable for the entire cell and tissue, and various experimental conditions [4–8].

The pea aphid, *Acyrthosiphon pisum* (Harris) (Hemiptera, Aphidiae), is an important cosmopolitan pest. It feeds on a wide range of legume plants (family Fabaceae) worldwide, including pea, clover, alfalfa, and broad bean, and is considered as the aphid species of major agronomical importance [9]. More importantly, it can transmit over 30 plant viruses [10]. In addition, *A. pisum* is an emerging model organism for the studies of insect-plant interactions, especially after the release of its genome in 2010 [11]. With the advent of omics tools, there is an unprecedented opportunity to investigate the genetic basis of its physiological and biological functions [12, 13]. There have been demonstrated needs for the systematic validation of references genes in qRT-PCR analysis, normalization procedures have yet received any attention for this emerging insect model.

The objective of this study was to address an important aspect of gene expression studies in the pea aphid, *Acyrthosiphon pisum*, as well as in other insects which is the selection of appropriate references genes with stable expression under different experimental conditions. Here, the expression profiles of 11 candidate reference genes, including *actin* (Actin), elongation factor 1α (*EF1A*), TATA-box-binding protein (*TATA*), ribosomal protein L12 (*RPL12*), β-tubulin (*Tubulin*), NADH dehydrogenase (*NADH*), vacuolar-type H+–ATPase (*v-ATPase*), succinate dehydrogenase B (*SDHB*), 28S ribosomal RNA (*28S*), 16S ribosomal RNA (*16S*), and 18S ribosomal RNA (*18S*) from the pea aphid genome [11], were examined under different developmental stages and temperatures. As a result, different sets of reference genes were recommended accordingly.
Materials and Methods

Insects
Pea aphid, *Acyrthosiphon pisum* (Harris) (Hemiptera, Aphidiae) colony was kindly provided by Dr. John Obrycki (University of Kentucky). Aphids were maintained at 20–28˚C on seedlings of fava bean, *Vicia faba* (Fabales, Fabaceae) in a greenhouse.

Samples preparation
Fifteen adult females were allowed to lay the offspring for 24 h on fava bean leaves resting on wet filter paper in a petri dish (9 cm diameter). Then 10 adults as one replicate and 20 nymphs (less than 24 h old) as one replicate, respectively were exposed to 10˚C, 22˚C, and 30˚C, respectively for 2 d in a climate chamber with a photoperiod of 14: 10 (L: D) and 50% relative humidity. All collected samples were preserved in 1.5 ml centrifuge tubes and stored at −80˚C after being frozen in liquid nitrogen. Each treatment was repeated three times independently, therefore, there are 18 biological samples in total.

Total RNA extraction and cDNA synthesis
Total RNA was extracted using TRIzol (Invitrogen, Carlsbad, CA) following the manufacturer’s instruction. First-strand cDNA was synthesized from 1 μg of total RNA using the M-MLV reverse transcription kit (Invitrogen, Carlsbad, CA) according the manufacturer’s recommendations.

Reference gene selection and primer design
Eleven commonly used reference genes were selected (Table 1). PCR amplifications were performed in 50 μl reactions containing 10 μl 5 × PCR Buffer (Mg^{2+} Plus), 1 μl dNTP mix (10 mM of each nucleotide), 5 μl of each primer (10 μM each), and 0.25 μl of Go Taq (5 u/μl) (Promega). The PCR parameters were as follows: one cycle of 94˚C for 3 min; 35 cycles of 94˚C for 30 s, 59˚C for 45 s, and 72˚C for 1 min; a final cycle of 72˚C for 10 min. Amplicons of the expected size were purified and cloned into the pCR4-TOPO vector (Invitrogen, Carlsbad, CA) for sequencing confirmation.

Quantitative real-time PCR
Gene-specific primers (Table 1) were used in PCR reactions (20 μl) containing 7 μl of ddH2O, 10 μl of 2 × SYBR Green MasterMix (Bio-Rad), 1 μl of each specific primer (10 μM), and 1 μl of first-strand cDNA template. The qPCR program included an initial denaturation for 3 min at 95˚C followed by 40 cycles of denaturation at 95˚C for 10 s, annealing for 30 s at 55˚C, and extension for 30 s at 72˚C. For melting curve analysis, a dissociation step cycle (55˚C for 10 s, and then 0.5˚C for 10 s until 95˚C) was added. The reactions were set up in 96-well format
Microseal PCR plates (Bio-Rad) in triplicates. All experiments were replicated in triplicate.

Reactions were performed in a MyiQ single Color Real-Time PCR Detection System (Bio-Rad). Existence of single peaks in melting curve analysis was used to confirm gene-specific amplification and rule out non-specific amplification and primer-dimer generation. The qRT-PCR was determined for each gene using slope analysis with a linear regression model. Relative standard curves for the transcripts were generated with serial dilutions of cDNA (1/5, 1/25, 1/125, 1/625, and 1/3125). The corresponding qRT-PCR efficiencies (E) were calculated according to the equation: 

$$E = \left(10^{-1/slope}\right) - 1 \times 100.$$  

### Stability of gene expression

All biological replicates were used to calculate the average Ct value. The stability of the ten housekeeping genes were evaluated by algorithms **geNorm** [1], **NormFinder** [14], **BestKeeper** [15], and the comparative $$\Delta Ct$$ method [16]. Finally, we

---

<table>
<thead>
<tr>
<th>Gene</th>
<th>Accession No.</th>
<th>Primer sequences (5’-3’)</th>
<th>Length (bp)</th>
<th>E (%)*</th>
<th>$$R^2$$**</th>
</tr>
</thead>
<tbody>
<tr>
<td>EF1A</td>
<td>AY219737</td>
<td>F: AGAATGGACAAACCCGCTGAA</td>
<td>104</td>
<td>104.8</td>
<td>0.9989</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R: GCTGTAGTTGTTGTCAGTAGAG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tublin</td>
<td>NM_001190398</td>
<td>F: TTGGTACACTTGGAAGGTTAG</td>
<td>103</td>
<td>103.5</td>
<td>0.9980</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R: AGCCGTAGCTTTGCTGATTTG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NADH</td>
<td>NM_001162323</td>
<td>F: CGAGGAGAACATGCTCTTAGAC</td>
<td>93</td>
<td>104.6</td>
<td>0.9989</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R: GATAGCTTGGCTGGAGCATATAG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RPL12</td>
<td>NM_001126171</td>
<td>F: AAGGCTATCAGTCTGACGAGAAAG</td>
<td>101</td>
<td>108.0</td>
<td>0.9992</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R: ACCAATGATGTCAGAGAAGGA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SDHB</td>
<td>NM_001162436</td>
<td>F: CTGAATTCTCTTGAGACCTATGG</td>
<td>90</td>
<td>102.9</td>
<td>0.9963</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R: ACGGCCAGAAGCCCTAAA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18S</td>
<td>X62623</td>
<td>F: CGGCGAACCACGTCAATTCA</td>
<td>101</td>
<td>103.9</td>
<td>0.9921</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R: GGAACCTCTCGGCATGTATTA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28S</td>
<td>S50426</td>
<td>F: CCGGTGCTAACCCTCTTACAA</td>
<td>118</td>
<td>97.2</td>
<td>0.9943</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R: CGACGCTTCTCTGATTTGTTCA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16S</td>
<td>FJ411411</td>
<td>F: AGAAACCAACCTGGCTTACAC</td>
<td>121</td>
<td>109.9</td>
<td>0.9989</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R: TTGCGACCTCTGATGTGTAATTA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>v-ATPase</td>
<td>NM_001126156</td>
<td>F: CTTCTTGCTGAGTGTCTGGT</td>
<td>90</td>
<td>99.4</td>
<td>0.9931</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R: GCCATCAGCAGCAGCTGATTA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Actin</td>
<td>NM_001126200</td>
<td>F: CGTTCAACTGGGAAGCTATGG</td>
<td>111</td>
<td>105.8</td>
<td>0.9985</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R: GGTTCAATGGGAGCTGTTTTA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TATA</td>
<td>NM_001162717</td>
<td>F: CACCTAATGTCACCACGCTATT</td>
<td>123</td>
<td>106.9</td>
<td>0.9992</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R: TGTGTCAAGCGCTTCTAAG</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**E**: PCR efficiency (calculated from the standard curve).

**R^2**: Regression coefficient.

doi:10.1371/journal.pone.0110454.t001
compared and ranked the tested candidates based on a web-based analysis tool RefFinder (http://www.leonxie.com/referencegene.php).

Results

Transcriptional profiling of candidate reference genes

First, 11 candidate reference genes were investigated by reverse transcription polymerase chain reaction (RT-PCR). All genes tested were expressed in pea aphid, and visualized as a single amplicon with expected size on a 1.5% agarose gel. All amplicons were sequenced and displayed 100% identity with their corresponding sequences. Furthermore, gene-specific amplification of these genes was confirmed by a single peak in real-time melting-curve analysis. A standard curve was generated for each gene, using five-fold serial dilution of the pooled cDNAs. The correlation coefficient and PCR efficiency for each standard curve were shown in Table 1.

We calculated the mean and the standard derivation (SD) of the C\textsubscript{t} values for all the samples together. v-ATPase had the most variable expression levels reflected in its high SD values. On the contrary, 28S had the least variable expression levels reflected in its low SD values. In addition, TATA (C_{tavg}=26.05) had the highest C\textsubscript{t} values and was therefore the least expressed among the gene candidates. 18S (C_{tavg}=10.72) had the lowest C\textsubscript{t} values and was therefore the mostly expressed among the gene candidates. (Figure 1).

Quantitative analysis of reference candidates based on geNorm

To determine the minimal number of genes required for normalization, we computed the V-value by geNorm. Starting with two genes, the software sequentially adds another gene and recalculates the normalization factor ratio. If the added gene does not increase the normalization factor ratio above the proposed 0.15 cut-off value, then the original pair of genes is enough for normalization. However, if the new ratio is above 0.15, then more genes should be included. The first V-value <0.15 was after V2/3 (Figure 2B). This means that two reference genes were enough for reliable normalization under the developmental stages and temperature conditions.

Determining the best reference candidates based on geNorm

GeNorm bases its ranking on the geometric mean of the SD of each transformed gene set of pair combinations (M-value). The lower the M-value is, the higher the ranking. SDHB and 16S were co-ranked as the most stable genes (M=0.382). The overall order based on geNorm from most stable to least stable reference genes was: SDHB=16S, NADH, 18S, RPL12, 28S, Tublin, Actin, EF1A, v-ATPase, TATA (Figure 2A, Table 2).
Determining the best reference candidates based on $\Delta C_t$ method

Gene ranking using the $\Delta C_t$ method relies on relative pair-wise comparisons. Using raw $C_t$ values, the average SD of each gene set is inversely proportional to gene stability. As shown in Tables S1 and 5, SDHB (0.57) was the top-ranked gene. The overall order from most stable to least stable reference genes based on the $\Delta C_t$ method was: SDHB, NADH,16S, RPL12,18S, Actin, EF1A,28S, Tublin, v-ATPase, TATA (Table 2).

Determining the best reference candidates based on NormFinder

SDHB (0.194) was the gene with the least variation in expression levels; thus SDHB would be the most reliable reference gene. The overall order from most stable to least stable reference genes based on NormFinder was: SDHB, NADH,16S, RPL12,18S, Actin, EF1A,28S, Tublin, v-ATPase, TATA (Table 2).

Determining the best reference candidates based on BestKeeper

BestKeeper provided a two-way ranking: Pearson’s correlation coefficient and BestKeeper computed SD values. The stability of a gene is directly proportional to the $|r|$ value, while it is inversely proportional to the SD value. SDHB($r=0.931$)
and 16S (r=0.852) had the highest [r] value, whereas 28S (SD=0.321) and EF1A (SD=0.371) had the least variable expression levels across all the samples (Table S2, 2).

Figure 2. Quantitative and qualitative analysis based on geNorm. (A) Ranking of the 11 housekeeping genes based on the stability value (M). A lower stability value indicates more stable expression. (B) Pairwise variation (V) analysis of the candidate reference genes. The geNorm first calculates an expression stability value (M) for each gene and then compares the pair-wise variation (V) of this gene with the others. A threshold of V<0.15 was suggested for valid normalization. geNorm starts by a gene pair, and tests whether the inclusion of a 3rd gene adds significant variation. The pair-wise variation (V_n/V_{n+1}) was analyzed between the normalization factors NF_n and NF_{n+1} by the geNorm software to determine the optimal number of references genes required for qRT-PCR data normalization.

doi:10.1371/journal.pone.0110454.g002

Comprehensive ranking of best reference genes using RefFinder
All software programs except the SD value based on BestKeeper indentified SDHB as the most stable gene (Table S2). According to RefFinder, the overall order from the most stable to the least stable reference genes was: SDHB, 16S, NADH, 28S, RPL12, 18S, EF1A, Actin, Tublin, v-ATPase, TATA. Among them, v-ATPase and
TATA both had GM values higher than 10.0 (Table 2), these two candidates had the lowest ranking and less suitable to serve as reliable reference genes for normalizing gene expression.

### Discussion

qRT-PCR quantification requires robust normalization by reference genes to offset confounding variations in experimental data. Most gene expression studies in the literature use a single endogenous control; this will profoundly influence the statistical outcome and may lead to inaccurate data interpretation [17]. Currently, the reference genes studies of many insects have been accomplished including whitefly, diamondback moth, brown planthopper, beet armyworm, oriental leafworm moth, Colorado potato beetle, and oriental fruit fly [4–8, 18, 19]. Here, the expression profiles of 11 candidate reference genes from the pea aphid were evaluated under different developmental stages and temperature conditions. Our finding is the first step toward establishing a standardized qRT-PCR analysis for this research model.

*BestKeeper* ranked the genes different from the other analysis methods used (Table 2). Unlike *Genorm* and *NormFinder, BestKeeper* is not specifically built to construct a hierarchy of reference genes. Instead, *BestKeeper* is intent to establish the best possible referencing point using an averaged expression of multiple housekeeping genes. Therefore, based on the needs, different analytical tools should be considered. There has been ongoing discussion about the optimal number of reference genes required for qRT-PCR analysis. The fact is that multiple reference genes are increasingly used to analyze gene expression under various experimental conditions in a given experiment, because one reference gene is usually insufficient to normalize the expression results of target genes [20].

---

**Table 2. Ranking of reference gene candidates using different algorithms***.

<table>
<thead>
<tr>
<th>RefFinder</th>
<th>geNorm</th>
<th>NormFider</th>
<th>ΔCt</th>
<th>BestKeeper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genes</td>
<td>Genes</td>
<td>Genes</td>
<td>Genes</td>
<td>Genes</td>
</tr>
<tr>
<td>SDHB</td>
<td>1.495</td>
<td>SDHB</td>
<td>0.382</td>
<td>SDHB</td>
</tr>
<tr>
<td>16S</td>
<td>2.449</td>
<td>16S</td>
<td>0.382</td>
<td>NADH</td>
</tr>
<tr>
<td>NADH</td>
<td>2.449</td>
<td>NADH</td>
<td>0.446</td>
<td>16S</td>
</tr>
<tr>
<td>28S</td>
<td>4.427</td>
<td>18S</td>
<td>0.504</td>
<td>RPL12</td>
</tr>
<tr>
<td>RPL12</td>
<td>4.681</td>
<td>RPL12</td>
<td>0.540</td>
<td>18S</td>
</tr>
<tr>
<td>18S</td>
<td>5.318</td>
<td>28S</td>
<td>0.566</td>
<td>Actin</td>
</tr>
<tr>
<td>EF1A</td>
<td>5.450</td>
<td>Tublin</td>
<td>0.585</td>
<td>EF1A</td>
</tr>
<tr>
<td>Actin</td>
<td>7.135</td>
<td>Actin</td>
<td>0.610</td>
<td>28S</td>
</tr>
<tr>
<td>Tublin</td>
<td>7.937</td>
<td>EF1A</td>
<td>0.624</td>
<td>Tublin</td>
</tr>
<tr>
<td>v-ATPase</td>
<td>10.241</td>
<td>v-ATPase</td>
<td>0.671</td>
<td>v-ATPase</td>
</tr>
<tr>
<td>TATA</td>
<td>10.741</td>
<td>TATA</td>
<td>0.720</td>
<td>TATA</td>
</tr>
</tbody>
</table>

***: Geometric mean (GM); Stability Value (SV); Pearson’s correlation coefficient ([r];) Standard Deviation (SD).

doi:10.1371/journal.pone.0110454.t002
This can decreased the probability of biased normalization. Our results demonstrated that the use of two reference genes can be sufficient to normalize the expression data and provides a more conservative estimation of target gene expression (Figure 2). As a result, we strongly suggest that two internal references are necessary for studying gene expression in pea aphid under different developmental stages and temperature conditions.

This is the first study to evaluate candidate reference genes for gene expression analyses in the pea aphid. Based on the comprehensive analysis, SDHB, 16S, and NADH were the three most stable house-keeping genes under different developmental stages and temperature conditions. This study not only sheds light on establishing a standardized qRT-PCR procedure in pea aphid, but also lays a solid foundation for the genomics and functional genomics research in this insect.

Supporting Information

Table S1. Summary of mean and SD values of gene pairwise comparison using the ΔCt method for 11 gene candidates.
doi:10.1371/journal.pone.0110454.s001 (DOCX)

Table S2. Ranking of 11 reference gene candidates based on BestKeeper. Two criteria are considered: Pearson's correlation coefficient and BestKeeper computed SD values. The stability of a gene is directly proportional to the [r] value, while it is inversely proportional to the SD value.
doi:10.1371/journal.pone.0110454.s002 (DOCX)

Acknowledgments

The authors are grateful to anonymous reviewers and the editor for their constructive criticisms. Special thanks go to Dr. Xun Zhu for his assistance with the data analysis. The information reported in this paper (No. 14-08-063) is part of a project of the Kentucky Agricultural Experiment Station and is published with the approval of the Director.

Author Contributions

Conceived and designed the experiments: HPP YL XGZ. Performed the experiments: CXY HPP. Analyzed the data: HPP. Contributed reagents/materials/analysis tools: XGZ. Wrote the paper: HPP CXY XGZ.

References


