Validation of Reference Genes for Quantitative Expression Analysis by Real-Time RT-PCR in Four Lepidopteran Insects

Authors: Teng, Xiaolu, Zhang, Zan, He, Guiling, Yang, Liwen, and Li, Fei

Validation of reference genes for quantitative expression analysis by real-time RT-PCR in four lepidopteran insects

Xiaolu Teng[1,2a], Zan Zhang[1,2b], Guiling He[1,2,3c], Liwen Yang[1,2d], and Fei Li[1,2e]*

[1]Department of Entomology, College of Plant Protection, Nanjing Agricultural University, Nanjing, 210095, China
[2]Key Laboratory of Integrated Management of Crop Diseases and Pests, Nanjing Agricultural University, Ministry of Education, 210095, China
[3]Department of Chemistry, Jiaying University, Meizhou, Guangzhou, 514015, China

Abstract

Quantitative real-time polymerase chain reaction (qPCR) is an efficient and widely used technique to monitor gene expression. Housekeeping genes (HKGs) are often empirically selected as the reference genes for data normalization. However, the suitability of HKGs used as the reference genes has been seldom validated. Here, six HKGs were chosen (*actin A3*, *actin A1*, *GAPDH*, *G3PDH*, *E2F*, *rp49*) in four lepidopteran insects *Bombyx mori* L. (Lepidoptera: Bombycidae), *Plutella xylostella* L. (Plutellidae), *Chilo suppressalis* Walker (Crambidae), and *Spodoptera exigua* Hübner (Noctuidae) to study their expression stability. The algorithms of geNorm, NormFinder, stability index, and ΔCt analysis were used to evaluate these HKGs. Across different developmental stages, *actin A1* was the most stable in *P. xylostella* and *C. suppressalis*, but it was the least stable in *B. mori* and *S. exigua*. *Rp49* and *GAPDH* were the most stable in *B. mori* and *S. exigua*, respectively. In different tissues, *GAPDH*, *E2F*, and *Rp49* were the most stable in *B. mori*, *S. exigua*, and *C. suppressalis*, respectively. The relative abundances of *Siwi* genes estimated by $2^{-\Delta\Delta Ct}$ method were tested with different HKGs as the reference gene, proving the importance of internal controls in qPCR data analysis. The results not only presented a list of suitable reference genes in four lepidopteran insects, but also proved that the expression stabilities of HKGs were different among evolutionarily close species. There was no single universal reference gene that could be used in all situations. It is indispensable to validate the expression of HKGs before using them as the internal control in qPCR.

Cite this paper as:
Teng X, Zhang Z, He G, Yang L, Li F. 2012. Validation of reference genes for quantitative expression analysis by real-time RT-PCR in four lepidopteran insects. *Journal of Insect Science* 12:60 available online: insectscience.org/12.60

Introduction

Estimating transcript expression levels across different conditions is an important aspect of work in molecular biology. Many techniques such as northern blot, microarray, and quantitative real-time polymerase chain reaction (qPCR) have been developed to determine gene expression levels. Among these, qPCR is one of the best choices and is widely applied in monitoring transcript-level changes in gene expression.

In several steps of the qPCR protocol, including preparing material, RNA purification, cDNA synthesis, and PCR procedures, non-specific variations may be introduced. For example, the efficiencies of reverse transcription (RT) and PCR may be varied between samples. Therefore, selection of one or more reference genes, also called internal controls, is very important to eliminate the inter-sample variations in analyzing qPCR data. An ideal reference gene is constantly expressed in all samples under certain condition. Housekeeping genes (HKGs) are thought to have key roles in cellular processes and are expressed at a constant level, and thus are often used as reference genes in qPCR. There are two strategies to choose a reference gene. One is to use HKGs as reference genes that have already used in previous studies; another is to use HKGs that are homologous to a widely used reference gene in other model species.

These empirical choices often lack experimental support. Increasing evidence has suggested that HKGs are not always expressed stably in all experimental conditions. The α-*actin* gene, which has been widely used as a reference gene, was reported to be highly regulated by matrigel, and therefore unsuitable as an internal control in this condition (Selvey et al. 2001). Several housekeeping genes, such as α-tubulin and *GAPDH*, were also reported to have varied expression levels in some instances such as different tissues or different developmental stages (Brunner et al. 2004; Trivedi and Arasu 2005; Pei et al. 2007; Dhar et al. 2009; Dong et al. 2010; Lord et al. 2010; Wang et al. 2010). Some studies also proved that choosing unsuitable reference genes produced low precision or misleading results (Jian et al. 2008). Therefore, it is necessary to validate the expression stabilities of HKGs when choosing them as reference genes in qPCR.

Though qPCR has been widely used to estimate the gene expression in insects, very few studies have been performed to validate the expression stability of HKGs. In *Bombyx mori* L. (Lepidoptera: Bombycidae), translation initiation factor 4A, translation initiation factor 3 subunit 4, and translation initiation factor 3 subunit 5 were reported to be the most reliable reference genes during metamorphosis (Wang et al. 2008). However, the most commonly used reference gene, cytoplasmic actin, varied drastically throughout metamorphosis development (Wang et al. 2008). Ribosomal protein genes RPS3, RPS18, and RPL13a were the most stable genes in *Tribolium castaneum* exposed to *Beauveria bassiana* (Lord et al. 2010). Here, four economic and agriculturally important lepidopteran insects were chosen, including *B. mori*, *Plutella xylostella* L. (Plutellidae), *Chilo suppressalis* Walker (Pyralididae), and *Spodoptera exigua* Hübner (Noctuidae) to evaluate the expression stability of six HKGs (*actin A3*, *actin A1*, *GAPDH*, *G3PDH*, *E2F*, and *rp49*). Reliable reference genes are proposed in these four lepidopteran insects.

Materials and Methods

Sample collection

All insects were maintained at 25 ± 1 °C under a photoperiod of 16:8 L:D and 80% RH. The silkworms (*B. mori*) were fed on mulberry leaves, diamondback moths (*P. xylostella*) were reared on radishes, rice stem borers (*C. suppressalis*) were fed on rice seedling, and beet armyworms (*S. exigua*) were fed on an artificial diet (Huang et al. 2002; Merkx-Jacques and Bede 2005). The different developmental stages collected were egg, larvae (collected at the first day of each instar), pupa, and adult, for each insect. Tissues from the head, midgut, ovary, testis, fat body, malpighian tube, and epidermis were dissected from final instar larvae and kept at −70 °C for RNA purification. The silk glands of the *B. mori* were also collected for experiments. Two sets of samples were collected for experimental replication. About 30 insects were collected for each sample, and all experiments for each set of samples were repeated three times.

Collecting the sequences of candidate reference genes

A list of housekeeping genes (HKGs) was compiled that have been commonly used as the internal controls for normalization in qPCR data analysis in four insects by reference mining. The sequences of selected HKGs of *P. xylostella*, *C. suppressalis*, and *S. exigua* were obtained by searching transcriptome data. The sequences of *B. mori* genes were downloaded from the refseq database of NCBI. All the primers were designed using Beacon Designer 7 (Premier Biosoft, www.premierbiosoft.com). The sequences of primers are given in Table 1. *Siwi*-1 and *Siwi*-2 genes were chosen to study the impact of reference genes on the qPCR data analysis. The forward primer sequence of *Siwi*-1 was 5'-GTGCGGGTCTGCCTGAGTTG-3' and the reverse primer sequence of *Siwi*-1 was 5'-GGTGTGCTTGTGAATCTCCTGTTG-3'.

The forward primer sequence of *Siwi*-2 was 5'-AGCCTACCCTGACCTATGTTG-3' and the reverse primer sequence of *Siwi*-2 was 5'-ACCAGTCCCGTCTCGTTATAC-3'.

RNA purification

The samples were frozen with liquid nitrogen and homogenized in a tissue grinder. Then, 1 mL TRIzol reagent (GIBCO, www.invitrogen.com) was added to homogenized samples. Total RNA was isolated following the recommended procedures. Genomic DNA was removed from total RNAs by DNase I treatments following the protocol of the DNA-free kit (Ambion, www.invitrogen.com). The integrity of RNA was checked on a 1.5% agarose gel and visualized by ethidium bromide staining. The concentrations and quality of RNA were also measured with Nanodrop ND-1000 (Thermo Scientific, www.thermoscientific.com). The 260/280 nm absorbance ratios of all RNA samples were between 1.8 and 2.2. First strand cDNA was synthesized from 1 μg total RNA using M-MLV reverse transcriptase (Takara Bio Inc., www.takara-bio.com) and Oligo (dT18) as the anchor primer. The reaction mixtures were incubated at 70 °C for 10 min followed by 42 °C for one hour and 70 °C for 15 min.

Quantitative real-time PCR

The qPCR reactions were carried out with SYBR Premix Ex Taq (Takara Bio Inc.) following the manufacturer's protocol using an ABI Prism 7300 (Applied Biosystems, www.appliedbiosystems.com). The reaction mixture consisted of 2 μL of cDNA template in a final reaction volume of 20 μL. The qPCR

protocol included an initial step of 95 for 30 sec, followed by 40 cycles of 95 °C for five sec and then annealed at 60 °C for 31 sec, followed by one cycle of 95 °C for 15 sec, 60 °C for 60 sec, and 95 °C for 15 sec. Products were dissolved by curve analysis (60-95 °C) after 40 cycles. The specificity of the qPCR reactions was monitored with melting curve analysis using SDS software (version 1.4) and gel electrophoresis. Amplification efficiencies were determined by a series of template dilutions. All experiments were repeated in triplicate.

Data processing

The raw Ct values were obtained using the SDS software of ABI 7300 (version 1.4). The algorithms including geNorm (Vandesompele et al. 2002), Normfinder (Andersen et al. 2004), ΔCt approach (Silver et al. 2006), and Stability index (Brunner et al. 2004) were used to analyze the raw Ct values of selected HKGs. The analysis procedures strictly followed the manuals of the algorithms. The fold changes of *Siwi*-1 and *Siwi*-2 genes in different developmental stages or tissues were calculated using standard delta-delta-Ct method (Pfaffl 2001).

Results

Identification of candidate housekeeping genes

According to the mining of qPCR publications, six commonly-used candidate HKGs were selected to validate their expression stability, including cytoplasmic *actin* gene A3 (*actin A3*), cytoplasmic *actin* gene A1 (*actin A1*), glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*), glycerol-3-phosphate dehydrogenase-2 (*G3PDH*), *E2F* transcription factor 4 (*E2F*), and ribosomal protein 49 (*rp49*). The sequences of six HKGs in *B. mori* were

obtained by searching the refseq database of NCBI (Walter et al. 2009). The HKGs of three lepidopteran pests, *C. suppressalis*, *P. xylostella*, and *S. exigua* were searched from the transcriptome data using the BLASTX program (unpublished data). The sequences of *actin A3* in *B. mori* were used to search for homologs in the three lepidopteran pests, but homologs were not found and thus not used in further experiments. *Rp49* of *S. exigua* was not considered for further analysis because it had a very short sequence of poor quality.

Primer specificities were confirmed by melting curve analysis and all primer pairs amplified a single PCR product with the expected sizes. PCR products were confirmed by bi-direction sequencing. Amplification efficiencies of primers reached the standard requirements of conventional qPCR, which were determined by a serial dilution of cDNA template. All the experiments including RNA purification, cDNA synthesis, and qPCR were carried out by the same individual using the same reagent lot.

Expression levels of the reference gene candidates

The Cycle threshold (Ct) values in qPCR gave an overview of the gene expression variation in the samples. All selected HKGs had moderate abundance in different developmental stages of *C. suppressalis*. The mean Ct values of HKGs in four insect species were 15.20-29.56 cycles. Individual HKGs had different expression levels across all tested samples. According to the variations of Ct values, *GAPDH* showed the smallest gene expression variation (below five cycles) in *B. mori* and *S. exigua*, while *actin A1* had the highest expression variation (above eight cycles). However, *actin A1* was the most stable gene *C. suppressalis* and *P. xylostella*, whereas *GAPDH* was the least stable. The

Figure 1. The variance of raw Ct values of selected housekeeping genes (HKGs) in all tested samples. High quality figures are available online.
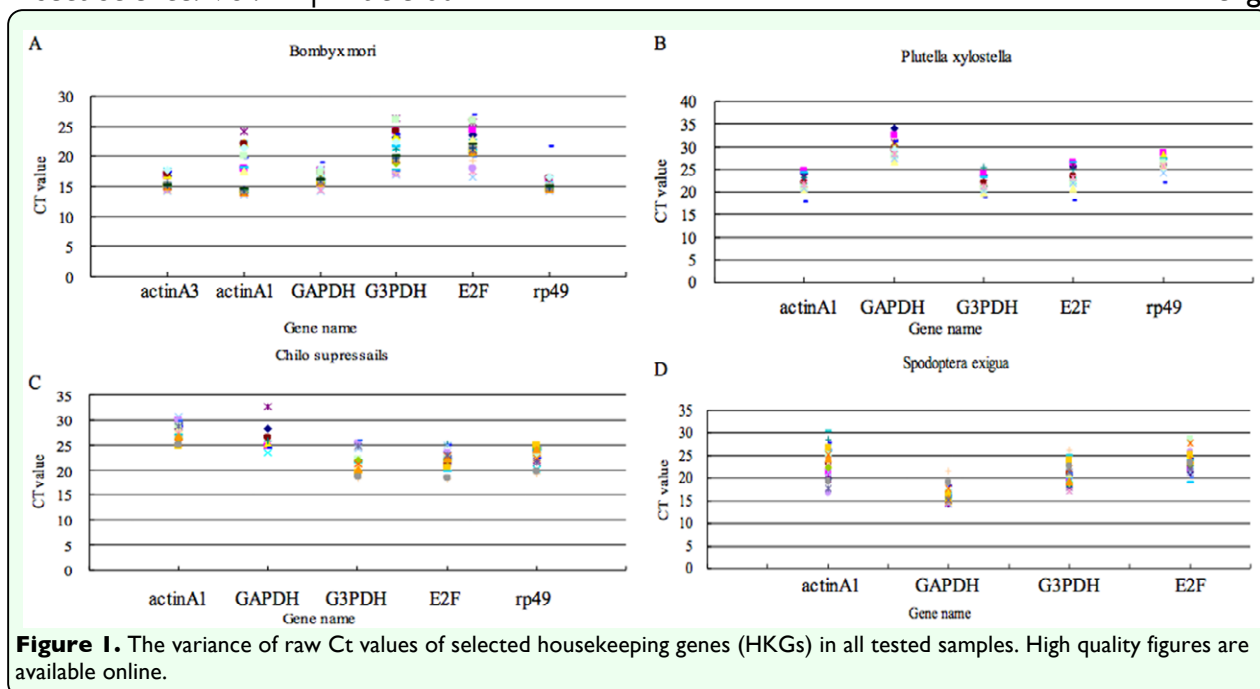
wide range of Ct values of six tested HKGs indicated that there was no HKG with constant expression under these conditions (Figure 1). Thus, determining a suitable HKG as a reference gene requires careful confirmation of expression stability.

Expression stability of selected HKGs across different developmental stages

Gene expression patterns across different developmental stages are frequently investigated. Insects were collected from developmental stages including egg, larvae collected at the first day of each instar, pupa, and adult, in order to study the expression stability of selected HKGs. Raw Ct values were first transformed to relative quantities using the ΔCt method. Then, geNorm and NormFinder were used to calculate the expression stability of selected HKGs. geNorm estimates the gene expression stability by calculating stability measure M for an HKG as the average pairwise variation for that gene with all other tested HKGs. The HKGs with low M values are the stably expressed genes. Based on this principle, geNorm recommend a pair of the most stable

HKGs as candidate reference genes (Vandesompele et al. 2002). According to the results of geNorm, the stabilities of selected HKGs were *actin* A3 = *rp49* > *GAPDH* > *G3PDH* > *E2F* > *actin A1* in *B. mori*. The HKG stabilities in *P. xylostella* were *Actin* A1 = *E2F* > *G3PDH* > *rp49* > *GAPDH*; in *C. suppressalis actin* A1 = *G3PDH* > *E2F* > *rp49*; in *S. exigua GAPDH* = *G3PDH* > *E2F* > *actin* A1. These results demonstrated that actin A1 was most stable in *P. xylostella* and *C. suppressalis*, whereas it was least stable in the other two lepidopteran insects across different developmental stages. *GAPDH* was the most stable HKG in *S. exigua*, whereas it was the least stable in *P. xylostella* (Table 2 and Figure S1). These results proved that the stabilities of HKGs are quite different in the four lepidopteran insects used in this study.

To confirm the results of geNorm, NormFinder was used to analyze the qPCR data. In general, the optimum reference genes found by NormFinder were similar to those of geNorm, suggesting the reliability of data analysis. Both geNorm and NormFinder ranked *actin* A1 as the most stable HKG in *P.*

Figure 2. ΔCt approach analysis of expression stability of selected housekeeping gene (HKGs) in the (A) *Bombyx mori*, (B) *Plutella xylostella*, (C) *Chilo suppressalis*, and (D) *Spodoptera exigua*. The qPCR data from all tested sample were put together for analysis. ΔCt variability of HKGs is shown as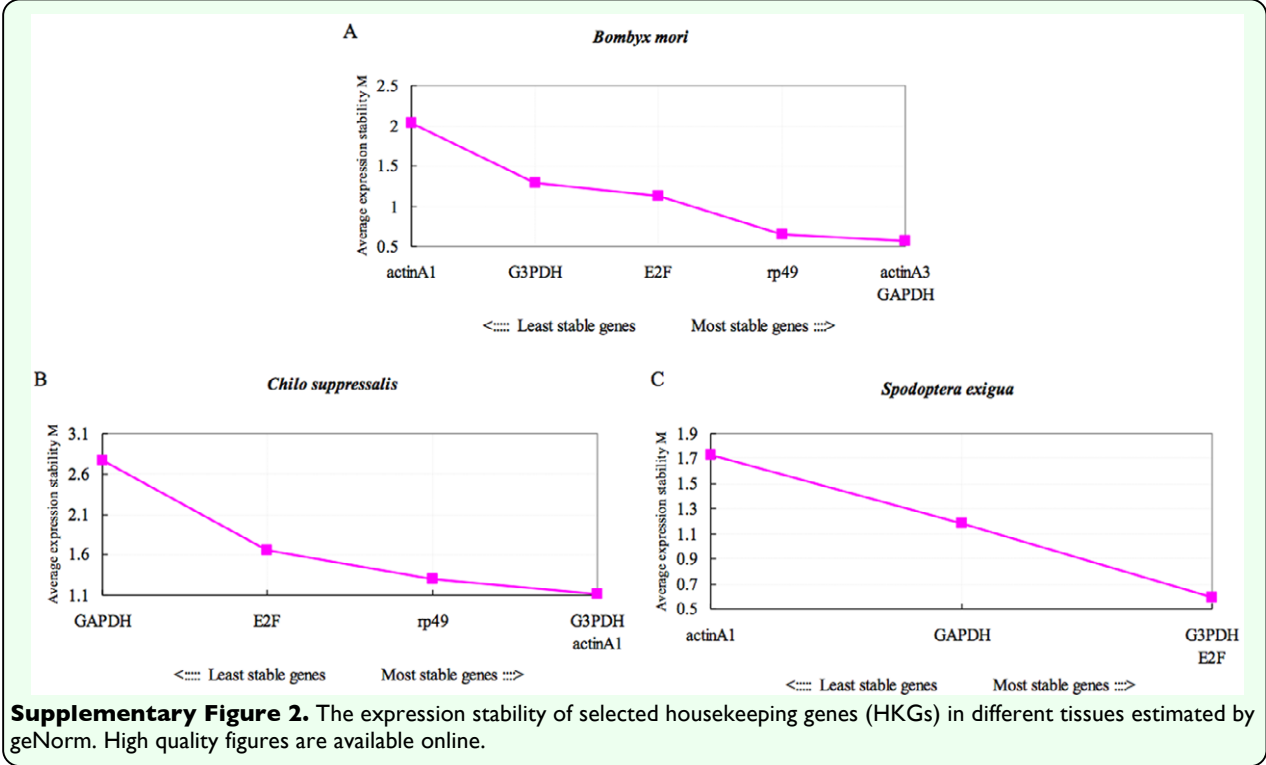 medians (lines), 25th percentile to the 75th percentile (boxes), and ranges (whiskers) estimated from the data of all samples. High quality figures are available online.

xylostella but the least stable in the *B. mori* and *S. exigua*. The least stable HKGs found by both two algorithms were the same. There was an inconsistency in determining the most stable gene in *B. mori* and *C. suppressalis* by two algorithms. In *B. mori*, *GAPDH* was evaluated as the most stable gene by NormFinder, but *rp49* and *actin A3* were the best candidates for reference genes by geNorm. In *C. suppressalis*, *E2F* was the most stable by NormFinder whereas *actin A1* and *G3PDH* were the most stable HKGs (Table 2). Taking all factors into account, we recommend that *rp49* in the *B. mori*, *actin A1* in *P. xylostella* and *C. suppressalis*, and *GAPDH* in *S. exigua* be used as the reference genes across different developmental stages.

In some cases, using the two HKGs in combination is more accurate than just using one reference gene. NormFinder not only finds the optimum reference genes out of a group of candidate genes, but also (in contrast to geNorm) takes information of groupings of samples into account (Andersen et al. 2004). Therefore, NormFinder is able to estimate the variation between sample groups, which can determine the best combination of two reference genes for normalization. The best combination of HKGs was *GAPDH* and *G3PDH* in *B. mori* with stability value of 0.421. The combinations of *actin* A1 and *G3PDH* were estimated as the best pair of reference genes in both *P. xylostella* and *S. exigua*. In *C. suppressalis*, *E2F* and *G3PDH* was the best combination with stability value of 0.454 (Table 3). It should be noticed that choosing the best combination considers the pairwise inter-variations and intra-variations in the grouping of samples. It does not rely on the stability value of an individual HKG. Actin A1 is the least stable HKG in the *S. exigua* when analyzed singly. However, it was estimated to be one of the best combinations (*actin* A1 and *G3PDH*) when considering multiple reference genes.

Supplementary Figure 1. The expression stability of selected housekeeping genes (HKGs) across different developmental stages estimated by geNorm. High quality figures are available online.

Supplementary Figure 2. The expression stability of selected housekeeping genes (HKGs) in different tissues estimated by geNorm. High quality figures are available online.
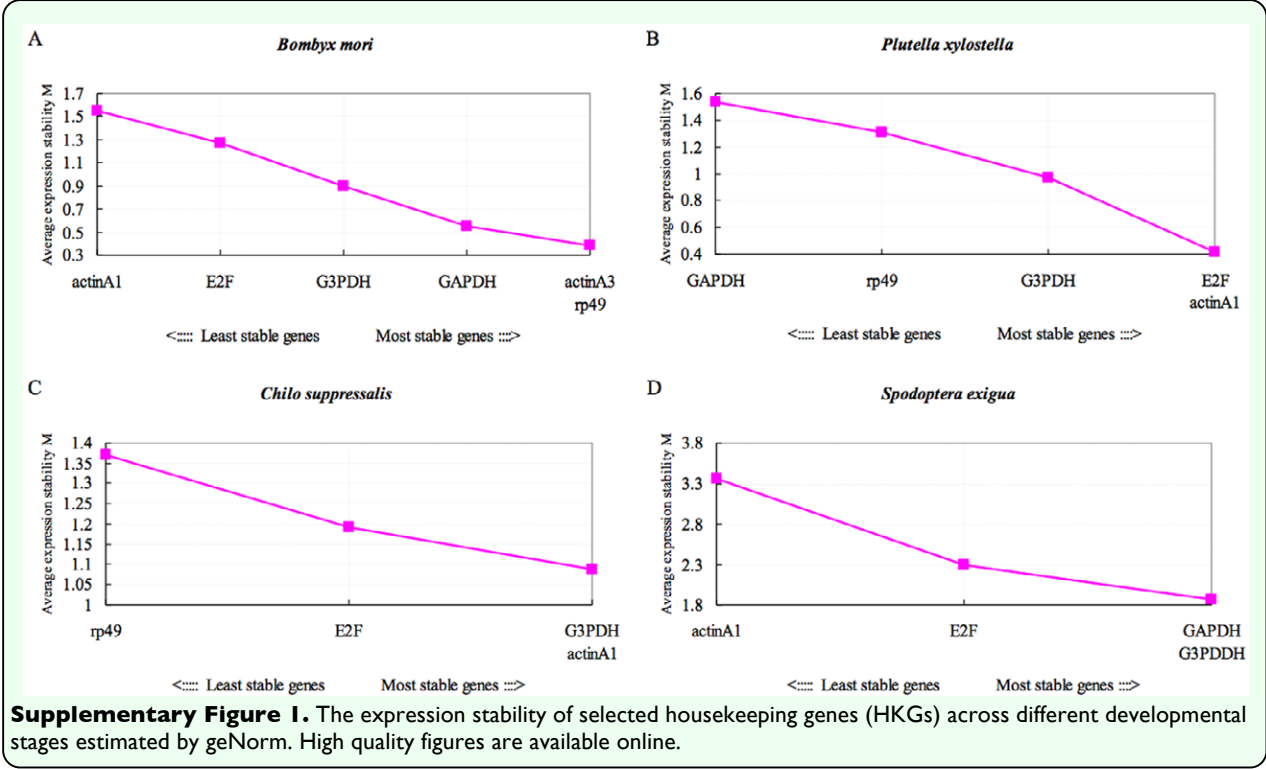
Expression stability of candidate reference genes in different tissues

The gene expression pattern in different tissues is also frequently studied to infer the gene functions. To screen suitable reference genes, different tissues were collected including head, midgut, ovary, testis, fat body, malpighian tube, epidermis, or silk gland of *B. mori*, *C. suppressalis*, and *S. exigua*. The

expression stabilities of selected HKGs were analyzed with geNorm and NormFinder. Though there were differences between the ranks of HKGs by two algorithms, the most stable and the least stable HKGs in different tissues of three insects found by NormFinder were identical with those found by geNorm. Considering all factors, *GAPDH* was the most stable of HKGs in *B. mori*, while *actin A1* in *C. suppressalis* and *E2F* in *S. exigua* were the best candidate reference genes (Table 4 and Figure S2). In *C. suppressalis*, the expression of *actin A1* was stable among different developmental stages and tissues. However, the most stable HKGs in the different tissues of *B. mori* and *S. exigua* were not same as those across different developmental stages, suggesting that HKGs have varied expression levels under different conditions.

For the best combinations of two HKGs, *GAPDH* and *G3PDH* were the most suitable candidates in different tissues of *B. mori*, which was identical with those across different developmental stages. *Rp49* and *actin A1* in *C. suppressalis* and *GAPDH* and *E2F* in *S. exigua* were the best combinations in different tissues, which was different from those across different developmental stages (Table 3).

Expression stability of candidate reference genes in all tested samples tissues

It is impossible to evaluate the expressions of HKGs under all kinds of experimental situations. To provide a general understanding of the expression stability of selected HKGs, the qPCR data from all tested samples was put together for evaluation. According to the results of geNorm and NormFinder, the most stable gene in *B. mori* was *GAPDH* followed by *actin A3*. *Actin A1* was the least stable HKG in *B. mori* and *S. exigua*, but it was the most suitable candidate of reference genes in

C. suppressalis and *P. xylostella*. *GAPDH* was the most stable HKG in *S. exigua*, followed by *G3PDH* (Table 5).

To validate the analysis results of geNorm and NormFinder, stability index assay and ΔCt analysis were also used to estimate the expression variations of HKGs in all tested samples. The stability index assay was first used to select suitable internal controls during the development of poplar (Brunner et al. 2004). According to the stability index assay, *rp49* had the lowest stability index in *B. mori*, *P. xylostella*, and *C. suppressalis*, while *GAPDH* was the most stable HKG in *S. exigua* (Table 5 and Table S1). In general, stability index assay had different results with those of geNorm and NormFinder in identifying the most stable HKGs. However, the most stable HKGs found by geNorm and NormFinder also had a low stability index.

The ΔCt approach was employed to select the most suitable HKG in reticulocytes (Silver et al. 2006). The principle of the ΔCt approach was to examine the ΔCt value between the two HKGs in different samples. If the ΔCt value remains constant, it suggests that either both HKGs are stably expressed among the tested samples or are co-regulated. However, the fluctuation of the ΔCt value indicates that at least one of them is variably expressed. After introducing more HKGs into the comparisons, an appropriate HKG can be found for a particular experimental system (Figure 2). The most stable HKGs selected by the ΔCt approach were identical with those by geNorm and NormFinder, which were *GAPDH in B. mori* and *S. exigua* and *actin A1* in *P. xylostella* and *C. suppressalis* (Table 5 and Table S2).

The impact of reference genes on qPCR data analysis

Based on the above analysis, we present evidence that the expression of HKGs in lepidopteran insects was not constant as expected. A stable HKG in one species did not mean that it also was expressed stably in another. Thus, it is our interest to study the impact of using "wrong" reference genes on the qPCR data analysis. The *Siwi*-1 and *Siwi*-2 genes in *B. mori* were chosen for evaluations of different HKGs. The samples from different developmental stages and six different tissues of the fifth instar larvae were used in the experiments.

For qPCR data analysis across different developmental stages, the egg stage was designated as the "calibrator sample" (abundance set to 1X). The relative expression levels of all other samples were then calculated relative to the calibrator sample. The results indicated that using different reference genes had apparent effects on the qPCR analysis. *Rp49* was recommended to be the most suitable candidate of reference genes across different developmental stages. The relative abundance of both *Siwi*-1 and *Siwi*-2 using *rp49* as the reference gene were in general similar to those obtained when using *actin A3*, the second most stable HKG (Table 6). However, the results normalized using *rp49* were significantly different from those using *actin A1*, the least stable HKG, as the internal control. In the third instar larvae, the relative abundance of *Siwi*-1 was 0.314 when using *rp49* as the reference gene, but it changed to 0.067 when using *actin A1*. A big difference could also be observed for the *Siwi*-2 gene. The relative abundance of *Siwi*-2 normalized using *actin A1* (about 10^{-4} level) was significantly lower than those using *rp49* (about 10^{-2} level) (Table 6).

For qPCR data analysis in different tissues, the head was selected as the "calibrator sample". In this case, *GAPDH* was the most stable HKG and was suggested to be the best reference gene. If using *GAPDH* as the reference gene, the relative mRNA abundances of *Siwi*-1 in the ovary and testis were only 2.406 and 1.769, respectively, but they changed to 26.438 and 46.709 when using *actin A1* as the internal control, which was the least stable HKG (Table 7). For *Siwi*-2 gene, if *GAPDH* was used as the reference gene, the relative expression levels in ovary and testis were 1.214 and 2.488, respectively. But they were 13.347 and 64.792 when using *actin A1* for normalization. On the contrary, the results of *GAPDH* were similar with those of *actin A3*, the second stable HKG (Table 7).

These surprising differences demonstrate the importance of choosing a proper reference gene. If an unsuitable reference gene were used in normalizing qPCR data, incorrect results might be obtained and lead to improper conclusions.

Discussion

qPCR has become a powerful tool for estimating gene expression levels because of its sensitivity and accuracy (Bustin 2000; Gachon et al. 2004; Wong and Medrano 2005; Nolan et al. 2006). The variations of qPCR will be unavoidably introduced by the procedures of RNA preparation, cDNA synthesis, or PCR processing. These variations can be controlled to some extent by carefully conducting the qRT-CPR experiments. The Ct values used for qPCR data analysis is determined mostly by the quantity of cDNA template (Walker 2002). The cDNA quantity is determined by the total RNA used in the first strand cDNA synthesis

and enzymatic efficiency of reverse transcriptase. Therefore, there is a close relationship between RNA integrity and expression quantitation. The quality and purity of the total RNA should be stringently analyzed before use in qPCR analysis (Bustin and Nolan 2004; Fleige and Pfaffl 2006; Fleige et al. 2006). However, it is impossible to eliminate all variations by these strategies. It is necessary to use HKGs as the proper reference genes for normalization.

To be a good reference gene, an HKG should meet three criteria: first, it should have amplification efficiency similar to the target genes; second, it should have moderate expression level; and third, its expression should be stable in all test samples. Unfortunately, it is unlikely that a single universal reference gene can be found in all species or in all experimental conditions of a species (Bustin 2000). Almost all genes including HKGs are regulated by other "regulators". It is unsafe to deduce the expression stability of an HKG based on known information of gene regulation. The *actin* gene is a widely used reference gene because it is thought to be stably expressed in all possible conditions. However, this gene has been challenged for its suitability as the internal control (Selvey et al. 2001). In our work, though *actin A1* was the most stable in *P. xylostella* and *C. suppressalis*, its expression was the least stable in *B. mori* and *S. exigua*.

Evaluation of the expression stability requires mathematical methods. Many algorithms such as geNorm (Vandesompele et al. 2002), Normfinder (Andersen et al. 2004), ΔCt approach (Silver et al. 2006), and stability index (Brunner et al. 2004) have been developed. The most stable HKG found by geNorm, NormFinder, and ΔCt approach were

the same in all four lepidopteran insects. But stability index had different results. We reasoned that this approach only considers the stability of HKGs individually but the other three algorithms evaluate the stability of HKGs by analyzing its variance comparing with other HKGs in pair. Thus, the reliability of stability index is not as high as others. Although the expression stability analyzed by ΔCt approach was similar to that by geNorm and Normfinder, it is not a good choice for reference gene selection because it relies on a strict principle that the expression variance of two HKGs should be identical in all samples in all experimental conditions or cell types. In addition, the computation procedures of ΔCt approach are cumbersome (Silver et al. 2006). geNorm and NormFinder use different mathematical methods to estimate the expression stability. The results of two algorithms can be used for cross validation. geNorm is one of the reliable algorithms to search for stably expressed HKGs that have low intra group variation and non-vanishing inter-group variation (Jain et al. 2006; Exposito-Rodriguez et al. 2008; Gutierrez et al. 2008; Jian et al. 2008). Compared with the geNorm, NormFinder finds not only the most stable gene but also the best combination of two reference genes if multiple internal controls are considered in qPCR data analysis (Oelkers et al. 2008).

In summary, our work presented a number of the most stable HKGs that are suitable to be used as the reference genes in four lepidopteran insects. The results demonstrated that no single universal reference gene could be used in all experiment conditions or treatments. The most reliable approach of choosing suitable HKGs as reference genes is to validate their expression stability using algorithms such as geNorm and NormFinder.

References

Andersen CL, Jensen JL, Orntoft TF. 2004. 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 Research* 64: 5245-5250.

Brunner AM, Yakovlev IA, Strauss SH. 2004. Validating internal controls for quantitative plant gene expression studies. *BMC Plant Biology* 4: 14.

Bustin SA. 2000. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *Journal of Molecular Endocrinology* 25: 169-193.

Bustin SA, Nolan T. 2004. Pitfalls of quantitative real-time reverse-transcription polymerase chain reaction. *Journal of Biomolecular Techniques* 15: 155-166.

Dhar AK, Bowers RM, Licon KS, Veazey G, Read B. 2009. Validation of reference genes for quantitative measurement of immune gene expression in shrimp. *Molecular Immunology* 46: 1688-1695.

Dong L, Sui C, Liu Y, Yang Y, Wei J. 2010. Validation and application of reference genes for quantitative gene expression analyses in various tissues of *Bupleurum chinense*. *Molecular Biology Reports* 38(8): 5017-5023

Exposito-Rodriguez M, Borges AA, Borges-Perez A, Perez JA. 2008. Selection of internal control genes for quantitative real-time RT-PCR studies during tomato development process. *BMC Plant Biology* 8: 131.

Fleige S, Pfaffl MW. 2006. RNA integrity and the effect on the real-time qRT-PCR performance. *Molecular Aspects of Medicine* 27: 126-139.

Fleige S, Walf V, Huch S, Prgomet C, Sehm J, Pfaffl MW. 2006. Comparison of relative mRNA quantification models and the impact of RNA integrity in quantitative real-time RT-PCR. *Biotechnology Letters* 28: 1601-1613.

Gachon C, Mingam A, Charrier B. 2004. Real-time PCR: what relevance to plant studies? *Journal of Experimental Botany* 55: 1445-1454.

Gutierrez L, Mauriat M, Guenin S, Pelloux J, Lefebvre JF, Louvet R, Rusterucci C, Moritz T, Guerineau F, Bellini C, Van Wuytswinkel O. 2008. The lack of a systematic validation of reference genes: a serious pitfall undervalued in reverse transcription-polymerase chain reaction (RT-PCR) analysis in plants. *Plant Biotechnology Journal* 6: 609-618.

Huang CX, Zhu LM, Ni YP, Chao XY. 2002. A method of rearing the beet armyworm, *Spodoptera exigua*. *Entomological Knowledge* 39: 229-231.

Jain M, Nijhawan A, Tyagi AK, Khurana JP. 2006. Validation of housekeeping genes as internal control for studying gene expression

in rice by quantitative real-time PCR. *Biochemical and Biophysical Research Communications* 345: 646-651.

Jian B, Liu B, Bi Y, Hou W, Wu C, Han T. 2008. Validation of internal control for gene expression study in soybean by quantitative real-time PCR. *BMC Molecular Biology* 9: 59.

Lord JC, Hartzer K, Toutges M, Oppert B. 2010. Evaluation of quantitative PCR reference genes for gene expression studies in *Tribolium castaneum* after fungal challenge. *Journal of Microbiological Methods* 80: 219-221.

Merkx-Jacques M, Bede JC. 2005. Influence of diet on the larval beet armyworm, *Spodoptera exigua*, glucose oxidase activity. *Journal of Insect Science* 5: 48. Available online, insectscience.org/5.48

Nolan T, Hands RE, Bustin SA. 2006. Quantification of mRNA using real-time RT-PCR. *Nature Protocols* 1: 1559-1582.

Oelkers K, Goffard N, Weiller GF, Gresshoff PM, Mathesius U, Frickey T. 2008. Bioinformatic analysis of the CLE signaling peptide family. *BMC Plant Biology* 8: 1.

Pei DS, Sun YH, Chen SP, Wang YP, Hu W, Zhu ZY. 2007. Zebrafish GAPDH can be used as a reference gene for expression analysis in cross-subfamily cloned embryos. *Analytical Biochemistry* 363: 291-293.

Pfaffl MW. 2001. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Research* 29: e45.

Selvey S, Thompson EW, Matthaei K, Lea RA, Irving MG, Griffiths LR. 2001. Beta-actin—an unsuitable internal control for RT-

PCR. *Molecular and Cellular Probes* 15: 307-311.

Silver N, Best S, Jiang J, Thein SL. 2006. Selection of housekeeping genes for gene expression studies in human reticulocytes using real-time PCR. *BMC Molecular Biology* 7: 33.

Trivedi S, Arasu P. 2005. Evaluation of endogenous reference genes for real-time PCR quantification of gene expression in *Ancylostoma caninum*. *Molecular and Biochemical Parasitology* 143: 241-4.

Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F. 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology* 3(7): RESEARCH0034.

Walker NJ. 2002. Tech.Sight. A technique whose time has come. *Science* 296: 557-559.

Walter MC, Rattei T, Arnold R, Guldener U, Munsterkotter M, Nenova K, Kastenmuller G, Tischler P, Wolling A, Volz A, Pongratz N, Jost R, Mewes HW, Frishman D. 2009. PEDANT covers all complete RefSeq genomes. *Nucleic Acids Research* 37: D408-11.

Wang GH, Xia QY, Cheng DJ, Duan J, Zhao P, Chen J, Zhu L. 2008. Reference genes identified in the silkworm *Bombyx mori* during metamorphism based on oligonucleotide microarray and confirmed by qRT-PCR. *Insect Science* 15: 405-413.

Wang Y, Han Z, Yan S, Mao A, Wang B, Ren H, Chi Y. 2010. Evaluation of suitable reference gene for real-time PCR in human umbilical cord mesenchymal stem cells with

long-term in vitro expansion. *In Vitro Cellular and Developmental Biology – Animal* 46: 595-599.

Wong ML, Medrano JF. 2005. Real-time PCR for mRNA quantitation. *Biotechniques* 39: 75-85.

Table 1. The primer sequences of selected housekeeping genes (HKGs) used in the experiments.

Species	Genes	Genbank Accession No.	Primers (from 5'to 3')	
			Sense	Antisense
Bombyx mori	*Actin A3*	X04507.1	TACGCACTCCCTCACGCCATCC	CCTCTGGGCAACGGAATCTTTCG
	actin A1	NM_001126252.1	GAGAGGGAAATCGTTCGTGACATC	AGGAAGGAAGGCTGGAAGAGAG
	GAPDH	AB262581.1	CGCTGGAATTTCTTTGAATGAC	CAATGACTCTGCTGGAATAACC
	G3PDH	AB164061.1	GCCTTCTAATGTTGTTGCTGTTCC	ACCACCTTCGGCTATATCAAATGC
	E2F	DQ311161.1	ATTCAATGGAGAGGAGCAGGTC	CATATCTTGGTTGTCTGGTTCGTC
	rp49	AB048205.1	AGGCATCAATCGGATCGCTATG	TTGTGAACTAGGACCTTACGGAATC
Plutella xylostella	actin A1	Transcriptome data	CACGGCATTGTCACAGATTGGAAC	CACTTCGGCGGACTTCTCACG
	GAPDH		CCGCATGGTATGTTCATTGAC	CAGGACTGTTCACTACTCTCG
	G3PDH		CCTACTGCTGCTGCGTTGTC	GCCACCTCAGAAGCGATGTTG
	E2F		TCTCAACTTCCTCGTCTTCCATAG	ACCAACAACCAAGTATTAGTGCTG
	rp49		CCTGGTTCCTGACGCTGTTC	GCATCGCCTGTCCATCTTCC
Chilo suppressalis	actin A1	Transcriptome data	GTCGCTTCCCAAATTACATC	CTCCATATCGTTCCAGTCG
	GAPDH		ATCTTGTGACCATCAATG	GCTGAATATGCTGCTTAC
	G3PDH		GTTGTGCCTCACCAATTTGTCAG	GCCACCTTCAGCGATGTCG
	E2F		ATTGCTGTGTGATAAAGAAGAAC	AGAAGGTGGTGGACTCAAC
	rp49		TGTTAGACACCATTCAGATAGG	GACCACAAGGAAGAAGATGC
Spodoptera exigua	actin A1	Transcriptome data	TTTCTCACGGTTCGCTTTGG	GTATCCTCACGCTCAAGTATCC
	GAPDH		AACATTTATCTCTACAACGCAATC	GTGACAACCACTCATCTATCTTC
	G3PDH		GTAGTTTCCCACCAGTTTGTCAG	CACCTCAGACGCAATGTTAGC
	E2F		GCCCAATCTGTTTCCTCAAAAG	GAACTTGCTCGCCGTAAGAC

Table 2. The stability values of selected housekeeping genes (HKGs) across the different developmental stages.

Species	Genes	NormFinder		geNorm	
		Stability values	Rank	Stability values	Rank
Bombyx mori	*Actin A3*	0.579	3	0.385	1
	actin A1	1.085	6	1.547	6
	GAPDH	0.459	1	0.548	3
	G3PDH	0.743	4	0.897	4
	E2F	0.752	5	1.27	5
	rp49	0.52	2	0.385	1
Plutella xylostella	actin A1	0.428	1	0.417	1
	GAPDH	1.081	5	1.532	5
	G3PDH	0.685	2	0.968	3
	E2F	0.751	3	0.417	1
	rp49	0.781	4	1.311	4
Chilo suppressalis	actin A1	0.647	2	1.087	1
	G3PDH	0.663	3	1.087	1
	E2F	0.505	1	1.192	3
	rp49	0.917	4	1.371	4
Spodoptera exigua	actin A1	2.012	4	3.36	4
	GAPDH	1.024	1	1.869	1
	G3PDH	1.19	2	1.869	1
	E2F	1.454	3	2.287	3

Table 3. The best combination of two housekeeping genes (HKGs) recommend by NormFinder, which can be used as the multiple references genes in different developmental stages or tissues.

Species	Developmental stages		Tissues	
	Best combination	Stability value	Best combination	Stability value
Bombyx mori	*GAPDH* and *G3PDH*	0.421	*GAPDH* and *G3PDH*	0.532
Plutella xylostella	actin A1 and *G3PDH*	0.474	N/A	N/A
Chilo suppressalis	*E2F* and *G3PDH*	0.454	rp49 and actin A1	0.722
Spodoptera exigua	actin A1 and *G3PDH*	1.185	*GAPDH* and *E2F*	0.621

Table 4. The stability values of selected housekeeping genes (HKGs) in different tissues.

Species	Gene name	NormFinder Stability values	NormFinder Rank	geNorm Stability values	geNorm Rank
Bombyx mori	actin A3	0.875	3	0.568	1
	actin A1	1.847	6	2.03	6
	GAPDH	0.364	1	0.568	1
	G3PDH	1.011	5	1.29	5
	E2F	0.923	4	1.123	4
	rp49	0.822	2	0.646	3
Chilo suppressalis	actin A1	0.475	1	1.114	1
	GAPDH	1.503	5	2.774	5
	G3PDH	1.061	2	1.114	1
	E2F	1.315	4	1.653	4
	rp49	1.136	3	1.3	3
Spodoptera exigua	actin A1	1.419	4	1.732	4
	GAPDH	0.673	1	1.182	3
	G3PDH	0.831	3	0.598	1
	E2F	0.673	1	0.598	1

Table 5. The stability values of selected housekeeping genes (HKGs) in all tested samples.

Species	Genes	Normfinder Stability values	Normfinder Rank	GeNorm Stability values	GeNorm Rank	Stability index Stability values	Stability index Rank	ΔCt SD	ΔCt Rank
Bombyx mori	Actin A3	0.646	2	0.595	1	0.364	2	2.202	3
	actin A1	1.587	6	1.89	6	3.451	6	3.439	6
	GAPDH	0.555	1	0.595	1	0.409	3	1.856	1
	G3PDH	0.91	4	1.56	5	1.803	5	2.325	5
	E2F	0.854	3	1.369	4	1.562	4	2.281	4
	rp49	1.042	5	0.877	3	0.083	1	2.095	2
Plutella xylostella	actin A1	0.428	1	0.649	1	3.565	2	1.202	1
	GAPDH	1.081	5	1.588	5	3.958	4	2.235	5
	G3PDH	0.685	2	0.923	3	3.704	3	1.498	3
	E2F	0.751	3	0.649	1	5.585	5	1.439	2
	rp49	0.781	4	1.156	4	2.232	1	1.6	4
Chilo suppressalis	actin A1	0.479	1	1.47	1	1.404	2	1.84	1
	GAPDH	1.838	5	2.339	5	11.79	5	3.378	5
	G3PDH	1.269	4	1.936	4	4.706	4	2.625	4
	E2F	0.846	2	1.47	1	1.799	3	2.361	3
	rp49	1.027	3	1.766	3	1.338	1	2.197	2
Spodoptera exigua	actin A1	2.426	4	2.844	4	6.454	4	2.592	4
	GAPDH	0.443	1	1.765	1	2.249	1	1.8	1
	G3PDH	1.058	2	1.765	1	3.962	2	1.913	2
	E2F	1.187	3	1.939	3	3.176	3	2.264	3

Table 6. The relative mRNA abundance of Siwi-1 and Siwi-2 genes across different developmental stages of the silkworm when using different housekeeping genes (HKGs) as the internal control. The rank of reference genes was determined by geNorm.

Genes	Rank	rp49 1	actin A3 2	GAPDH 3	G3PDH 4	E2F 5	actin A1 6
Siwi-1	Egg	1	1	1	1	1	1
	1st	0.169	0.272	0.202	0.316	0.952	0.037
	2nd	0.087	0.12	0.146	0.195	0.278	0.016
	3rd	0.314	0.327	0.292	0.084	0.407	0.067
	4th	0.027	0.025	0.041	0.128	0.155	0.293
	5th	0.048	0.063	0.083	0.088	0.179	0.01
	Pupa	0.773	1.55	0.498	0.379	0.604	0.157
	Adult	0.855	0.942	0.625	0.693	0.073	0.245
Siwi-2	Egg	1	1	1	1	1	1
	1st	0.004	0.031	0.022	0.175	0.577	0.00001
	2nd	0.173	0.166	0.879	0.02	2.703	0.00029
	3rd	0.019	0.021	0.066	0.011	0.479	0.00003
	4th	0.033	0.039	0.178	0.015	1.162	0.00005
	5th	0.028	0.032	0.087	0.031	0.937	0.00003
	Pupa	0.194	0.292	0.983	0.147	1.994	0.00044
	Adult	0.209	0.109	0.366	0.058	1.198	0.00015

Table 7. The relative mRNA abundance of *Siwi*-1 and *Siwi*-2 genes in different tissues of the silkworm when using different housekeeping genes (HKGs) as the internal control. The rank of reference genes was determined by geNorm.

Genes	Rank	*GAPDH* 1	*actin A3* 2	*rp49* 3	*E2F* 4	*G3PDH* 5	actin A1 6
Siwi-1	Head	1	1	1	1	1	1
	Midgut	0.498	0.151	5.312	0.815	0.218	1.962
	Ovary	2.406	1.599	2.037	0.532	2.513	26.438
	Testis	1.769	1.41	1.153	0.255	0.701	46.079
	Silk gland	0.788	0.176	0.256	1.977	5.955	11.864
	Malpighian tube	1.215	0.588	1.474	1.427	0.328	13.664
Siwi-2	Head	1	1	1	1	1	1
	Midgut	0.827	0.25	8.816	1.352	0.362	3.257
	Ovary	1.214	0.807	1.028	0.269	1.269	13.347
	Testis	2.488	1.983	1.621	0.358	0.985	64.792
	Silk gland	0.625	0.14	0.203	1.569	4.724	9.411
	Malpighian tube	1.415	0.685	1.718	1.663	0.383	15.923

Supplementary Table 1. Evaluation of the stability of selected housekeeping genes (HKGs) in all tested samples using stability index assay. HKGs with the lowest stability index are treated as the best control.

Species	Gene names	Mean	Stdev	Slope	R^2	CV%[A]	Stability index [B]	rank
Bombyx mori	actin A3	15.337	0.672	0.083	0.961	4.38	0.364	2
	actin A1	15.634	2.18	0.248	0.812	13.945	3.451	6
	GAPDH	16	0.738	0.089	0.907	4.613	0.409	3
	G3PDH	19.884	1.706	0.21	0.956	8.579	1.803	5
	E2F	21.208	1.709	0.194	0.81	8.06	1.562	4
	rp49	14.858	0.315	0.039	0.972	2.122	0.083	1
Plutella xylostella	actin A1	21.893	1.849	0.422	0.913	8.444	3.565	2
	GAPDH	29.561	2.153	0.544	0.935	7.282	3.958	4
	G3PDH	21.747	1.86	0.433	0.949	8.553	3.704	3
	E2F	23.352	2.381	0.548	0.925	10.198	5.585	5
	rp49	26.075	1.618	0.36	0.864	6.207	2.232	1
Chilo suppressalis	actin A1	27.854	1.671	0.234	0.98	6.001	1.404	2
	GAPDH	26.334	2.917	1.065	0.799	11.076	11.79	5
	G3PDH	21.795	2.256	0.455	0.921	10.349	4.706	4
	E2F	22.143	1.754	0.227	0.838	7.923	1.799	3
	rp49	23.045	1.548	0.199	0.827	6.719	1.338	1
Spodoptera exigua	actin A1	23.675	3.422	0.447	0.851	14.455	6.454	4
	GAPDH	16.119	1.705	0.213	0.777	10.578	2.249	1
	G3PDH	20.772	2.44	0.337	0.955	11.746	3.962	3
	E2F	23.523	2.348	0.318	0.919	9.981	3.176	2

Supplementary Table 2. The raw data of gene expression stability analysis by ΔCt approach.

Species	ΔCt	mean ΔCT	SD	Mean SD
Bombyx mori	actinA3 vs actin A1	-0.85	3.658	2.202
	actinA3 vs GAPDH	-0.201	1.285	
	actinA3 vs G3PDH	-4.918	2.286	
	actinA3 vs E2F	-6.259	2.197	
	actinA3 vs rp49	0.899	1.587	
	actin A1 vs actinA3	0.85	3.658	3.439
	actin A1 vs GAPDH	0.649	3.314	
	actin A1 vs G3PDH	-4.068	3.069	
	actin A1 vs E2F	-5.409	3.761	
	actin A1 vs rp49	1.749	3.395	
	GAPDH vs actinA3	0.201	1.285	1.856
	GAPDH vs actin A1	-0.649	3.314	
	GAPDH vs G3PDH	-4.717	2.035	
	GAPDH vs E2F	-6.058	1.66	
	GAPDH vs rp49	1.1	0.985	
	G3PDH vs actinA3	4.918	2.286	2.325
	G3PDH vs actin A1	4.068	3.069	
	G3PDH vs GAPDH	4.717	2.035	
	G3PDH vs E2F	-1.341	1.756	
	G3PDH vs rp49	5.817	2.477	
	E2F vs actin A3	6.259	2.197	2.281
	E2F vs actin A1	5.409	3.761	
	E2F vs GAPDH	6.058	1.66	
	E2F vs G3PDH	1.341	1.756	
	E2F vs rp49	7.158	2.029	
	rp49 vs actin A3	-0.899	1.587	2.095
	rp49 vs actin A1	-1.749	3.395	
	rp49 vs GAPDH	-1.1	0.985	
Plutella xylostella	GAPDH vs rp49	3.565	2.215	2.235
	GAPDH vs G3PDH	7.929	2.27	
	GAPDH vs E2F	6.45	2.329	
	GAPDH vs actin A1	7.849	2.126	
	rp49 vs GAPDH	-3.565	2.215	1.6
	rp49 vs G3PDH	4.328	1.561	
	rp49 vs E2F	2.723	1.522	
	rp49 vs actin A1	4.182	1.102	
	G3PDH vs GAPDH	-7.929	2.27	1.498
	G3PDH vs rp49	-4.328	1.561	
	G3PDH vs E2F	-1.605	1.244	
	G3PDH vs actin A1	-0.146	0.916	
	E2F vs GAPDH	-6.45	2.329	1.439
	E2F vs rp49	-2.723	1.522	
	E2F vs G3PDH	1.605	1.244	
	E2F vs actin A1	1.459	0.663	
	actin A1 vs GAPDH	-7.849	2.126	1.202
	actin A1 vs rp49	-4.182	1.102	
	actin A1 vs G3PDH	0.146	0.916	
	actin A1 vs E2F	-1.459	0.663	

	GAPDH vs rp49	3.52	3.101	3.378
	GAPDH vs E2F	3.507	3.624	
	GAPDH vs actin A1	-2.31	2.766	
	GAPDH vsG3PDH	2.486	4.02	
	rp49 vs GAPDH	-3.52	3.101	2.197
	rp49 vs G3PDH	0.123	2.167	
	rp49 vs E2F	0.57	2.014	
	rp49 vs actin A1	-5.298	1.505	
	G3PDH vs GAPDH	-2.486	4.02	2.625
Chilo	G3PDH vs rp49	-0.123	2.167	
suppressalis	G3PDH vs E2F	0.447	2.515	
	G3PDH vs actin A1	-5.421	1.8	
	E2F vs GAPDH	-3.507	3.624	2.361
	E2F vs rp49	-0.57	2.014	
	E2F vs G3PDH	-0.447	2.515	
	E2F vs actin A1	-5.868	1.289	
	actin A1 vs GAPDH	2.31	2.766	1.84
	actin A1 vs rp49	5.298	1.505	
	actin A1 vs G3PDH	5.421	1.8	
	actin A1 vs E2F	5.868	1.289	
	GAPDH vs E2F	-7.429	2.011	1.8
	GAPDH vs G3PDH	-4.04	1.45	
	GAPDH vs actin A1	-8.417	1.94	
	G3PDH vs GAPDH	4.04	1.45	1.913
	G3PDH vs E2F	-3.389	1.621	
Spodoptera	G3PDH vs actin A1	-4.377	2.669	
exigua	E2F vs GAPDH	7.429	2.011	2.264
	E2F vs G3PDH	3.389	1.621	
	E2F vs actin A1	-0.988	3.16	
	actin A1 vs GAPDH	8.417	1.94	2.592
	actin A1 vs G3PDH	4.569	2.676	
	actin A1 vs E2F	0.988	3.16	