Journal of Insect Science (*2016*) 16(1): 50; 1–8
doi: 10.1093/jisesa/iew034
Research article

OXFORD

Selection of Valid Reference Genes for Reverse Transcription Quantitative PCR Analysis in *Heliconius numata* (Lepidoptera: Nymphalidae)

Florence Piron Prunier,[1,2] Mathieu Chouteau,[1,3] Annabel Whibley,[1] Mathieu Joron,[1,3] and Violaine Llaurens[1]

[1]Institut de Systématique, Evolution, Biodiversité, ISYEB - UMR 7205, CNRS - Muséum National d'Histoire Naturelle - UPMC - EPHE, Sorbonne Universités, Paris, France, [2]Corresponding author, e-mail: fprunier@mnhn.fr, and [3]Centre d'Ecologie Fonctionnelle et Evolutive, CEFE UMR 5175, CNRS – Université de Montpellier – Université Paul Valéry Montpellier – EPHE, Montpellier 5, France

Subject Editor: Igor Sharakhov

Received 19 February 2016; Accepted 7 April 2016

Abstract

Identifying the genetic basis of adaptive variation is challenging in non-model organisms and quantitative real time PCR. is a useful tool for validating predictions regarding the expression of candidate genes. However, comparing expression levels in different conditions requires rigorous experimental design and statistical analyses. Here, we focused on the neotropical passion-vine butterflies *Heliconius,* non-model species studied in evolutionary biology for their adaptive variation in wing color patterns involved in mimicry and in the signaling of their toxicity to predators. We aimed at selecting stable reference genes to be used for normalization of gene expression data in RT-qPCR analyses from developing wing discs according to the minimal guidelines described in Minimum Information for publication of Quantitative Real-Time PCR Experiments (MIQE). To design internal RT-qPCR controls, we studied the stability of expression of nine candidate reference genes (*actin*, *annexin*, *eF1α*, *FK506BP*, *PolyABP*, *PolyUBQ*, *RpL3*, *RPS3A*, and *tubulin*) at two developmental stages (prepupal and pupal) using three widely used programs (GeNorm, NormFinder and BestKeeper). Results showed that, despite differences in statistical methods, genes *RpL3*, *eF1α*, *polyABP,* and *annexin* were stably expressed in wing discs in late larval and pupal stages of *Heliconius numata*. This combination of genes may be used as a reference for a reliable study of differential expression in wings for instance for genes involved in important phenotypic variation, such as wing color pattern variation. Through this example, we provide general useful technical recommendations as well as relevant statistical strategies for evolutionary biologists aiming to identify candidate-genes involved adaptive variation in non-model organisms.

Key words: RT-qPCR, housekeeping gene, butterfly, larval stage, wing disc

Introduction

Identifying the genetic basis of adaptive variation is challenging in non-model organisms. With the recent advance of next generation sequencing, we now have a facilitated access to genomic and transcriptomic datasets which often implicate a large number of candidate genes. Functional characterization is still required to pinpoint the causative genes among these candidates. Although genetic transformation methods provide the gold standard in linking genes to phenotypic effect, in systems where this remains difficult or impossible, indirect techniques such as quantitative reverse transcriptase PCR (RT-qPCR) permit the association of gene expression patterns or splicing variation with phenotypic variation (Kunte et al. 2014). RT-qPCR is a widely used method for fast, accurate, sensitive and cost-effective gene expression quantification when candidate-genes have been identified (Bustin 2002; Derveaux et al. 2010). This method presents several technical advantages including the possibility to investigate several target genes simultaneously. However, the simplicity of the technology itself makes it vulnerable to imprecision in experimental design, stressing the need of quality control throughout the entire procedure. Several technical insufficiencies could lead to unreliable assay performances such as for instance poor PCR primers or poor nucleic acid quality and quantity. For example, quantification of total RNA may be used for normalization but fluctuations in the efficiency of reverse transcription with different RNA samples may introduce misleading technical variations of cDNA quantities (Bustin 2002; Derveaux et al. 2010). To allow

more reliable and unequivocal interpretation of qPCR results, a MIQE guideline has been published in 2009 (Bustin et al. 2009) and is now a widely accepted standard, that we followed here.

Reference genes are used to control for intrinsic experimental variation introduced by differences in RNA extraction and reverse transcription yield (Bustin 2002) and allow a precise estimation of gene-specific expression variation. However, inadequate or non-validated reference genes may lead to erroneous interpretation of expression data (Bustin et al. 2009; Derveaux et al. 2010), stressing the need for a careful choice of reference genes. Reference genes should display invariant levels of expression across tissues, individuals, and experimental treatments. Housekeeping genes are generally thought to show relatively stable expression because of their role in the maintenance of basic cellular processes. Nevertheless, housekeeping genes may also participate in other processes (Petersen et al. 1990; Singh and Green 1993; Ishitani et al. 1996), and substantial variation may exist across individuals and populations in response to various factors such as cell type, cell metabolism, and culture conditions. Therefore, a given housekeeping gene cannot be universally taken as a reference gene in functional studies (Thellin et al. 1999). Furthermore, normalization using a single reference gene may potentially induce significant bias unless its expression is proven invariant under the experimental conditions described (Bustin et al. 2009). The use of a combination of reference genes is therefore recommended for an accurate comparison of candidate-gene levels of expression across samples (Bustin et al. 2009).

Here we report the optimization of reference genes for RT-qPCR in a non-model organism, the butterfly *Heliconius numata*, to study expression variations of candidate genes for wing pattern development according to the MIQE guidelines (Bustin et al. 2010). Neotropical butterflies in the genus *Heliconius* (Nymphalidae: Heliconiinae; passion-vine butterflies) are text-book examples of evolutionary convergence under strong natural selection (Benson 1972), and are prominent organisms for the study of ecological speciation (Brown 1981). These butterflies are toxic and display vivid wing colour patterns perceived as warning signals by predators (Merrill et al. 2015). Toxic species found within the same habitats tend to display similar warning patterns and resemble each-other sometimes with astonishing perfection. Such morphological convergence is referred as Müllerian mimicry and raises important questions on the genetic underpinnings of convergent traits, which requires a rigorous genetic dissection and the increasing use of modern genomic approaches (Brown and Benson 1974; Joron et al. 1999; Müller 1879). A reference genome was sequenced and assembled for a postman butterfly *Heliconius melpomene* (Heliconius Genome Consortium 2012) and is available (Altschul et al. 1997; Dasmahapatra et al. 2012; Heliconius Genome 2012). This reference genome is now used for comparative studies across the 45 species of the *Heliconius* genus, including our target species, *H. numata*, which diverged from *H. melpomene* about 4 Millions of years ago (Kozak et al. 2015).

H. numata, display a spectacular polymorphism, with up to seven mimetic morphs coexisting in local populations (Brown and Benson 1974; Joron et al. 2006), each being mimetic with other toxic butterflies sharing the same habitat. These wing pattern variations are controlled by a single genomic region, which contains at least 18 genes, called the supergene *P* (Joron et al. 2011). This supergene is characterized by a 400 kb chromosomal segment which is structured into distinct haplotype classes separated by 1–4% divergence. These classes are associated with different wing pattern alleles and with distinct chromosomal inversions (Joron et al. 2011). Reduced recombination in the polymorphic inversion (Joron et al. 2011) limits the power of association mapping to identify the genetic elements causing the developmental switch in color pattern, and stresses the need to document variation in expression levels for the different genes within the supergene during wing and scale development.

The supergene *P* is a positional homologue of two distinct loci in *H. melpomene*, *HmSb* and *HmYb,* which respectively control variations in the white margin and yellow bar on the hindwing, and the entire mapping region was fully sequenced and annotated (Ferguson et al. 2010). Furthermore, previous expression studies with microarray and quantitative PCR on transcripts in the genus *Heliconius* have pointed to genes involved in wing colour patterns and provide evidence for parallel gene expression on mimetic *Heliconius* butterfly wings (Reed et al. 2008; Ferguson and Jiggins 2009; Ferguson et al. 2011). Here we propose a reliable RT-qPCR design to test variations of expression of candidate color pattern genes in developing wing discs of late larval and pupal stages of *Heliconius*, focusing on the choice of suitable reference genes. This allows us to identify potential caveats and to provide useful recommendations for the choice of reference genes in non-model organisms.

Materials and Methods

Butterfly Material

In this study, the expression stability of putative reference genes was analysed in wing discs collected at two larval developmental stages from different morphs of *H. numata*. These larval stages, prepupae and 24 h after pupation, were targeted because they correspond to early development of wing color patterns (Nijhout 1991).

Wing discs from *H. numata* were obtained from dissection of larval individuals derived from controlled crosses between butterflies collected in PERU: Tarapoto, from September 2013 to March 2014. Adult wing color patterns of these larvae were inferred from the wing color patterns observed in the parents of the crosses and by genotyping assays at the supergene *P* as described in Le Poul et al. (2014). Six different genotypes were tested; five homozygous morphs (*H. n. silvana*, *H. n. tarapotensis*, *H. n. bicoloratus*, *H. n. aurora, H. n. arcuella*) and one heterozygous morph *H. numata [bicoloratus tarapotensis]*. Each morph was represented by two to four biological replicates. After dissection, material was immediately stabilized in RNA later at 4°C reagent according to the manufacturer's protocol (Qiagen, Hilden, Germany) and then stored at −20°C.

Total RNA Extraction and First Strand cDNA Synthesis

Tissue samples from one hindwing and one forewing for each sample were homogenised in 600 µl of RTL buffer with the Tissue Lyser (Qiagen, Hilden, Germany). Then, total RNA was extracted according to the manufacturer's protocol (RNeasy Mini kit, Qiagen, Hilden, Germany) and eluted in 30 µl of RNase-free water. To avoid genomic contamination, RNase free DNase treatment (Qiagen, Hilden, Germany) was performed during RNA extraction. The concentration and purity of the total RNA obtained was determined using Qubit 2.0 fluorometer (Invitrogen, Carlsbad, CA, USA). The RNA quality based on 28S and 18S integrity on electrophoresis gel, indicates a good quality for all the RNAs extracted

In this study, the priming method chosen was oligod(T) strategy to generate full length cDNA from poly(A)-tailed mRNA. So, for each sample, first strand cDNA synthesis was performed from 500 ng of total RNA using the SuperScript III and oligo$(dT)_{20}$ primers from Invitrogen (Carlsbad, CA, USA) according to the

manufacturer's instructions in a final volume of 20 µl. After synthesis, the total cDNA obtained was diluted 20-fold with ultrapure water.

To detect residual genomic DNA contamination, a control sample that does not include the RT enzyme (No template controls—NTCs) per condition was added. For a condition, these controls were constituted of a mix of RNA. The reaction mix was constituted of all components except the SuperScript III, in these cases an equal volume of nuclease-free water were added.

Identification of Candidate Reference Genes and Primers Designed

To select appropriate reference genes to normalize RT-qPCR data in *H. numata*, nine housekeeping genes were initially chosen and tested in this study, based on previous reports, mostly from insect systems: *elongation factor-1-alpha* (*eF1α*) and *ribosomal protein S3A* (*RPS3A*) genes in *H. melpomene* (Ferguson and Jiggins 2009; Ferguson et al. 2011), *annexin IX-B* gene in *Heliconius erato* (Reed et al. 2008), *FK506 binding protein 2* (*FK506BP*) gene in the butterfly *Bicyclus anynana* (Pijpe et al. 2011), *ribosomal protein L3* (*RpL3*) gene in the butterfly *Papilio polytes* (Nishikawa et al. 2013), *actin* gene in the moth *Spodoptera litura* (Lu et al. 2013), *tubulin* gene in the moth *Sesamia inferens* (Sun et al. 2015), *poly ubiquitin* (*polyUBQ*) in the bumblebees *Bombus terrestris* (Xiao et al. 2014), and *polyA-binding protein* (*polyABP*) gene in rat (Vesentini et al. 2012).

To find *Heliconius* orthologues, these sequences were blasted against the reference genome, *H. melpomene* primary v1.1 (Altschul et al. 1997; Dasmahapatra et al. 2012). *H. melpomene* orthologues were then blasted against *H. numata* sequence. *H. numata* sequences are available in the Sequence Read Archive under the project number PRJNA317526.

Specific DNA oligonucleotide primers were designed using sequences from *H. numata* coupled with *H. melpomene*'s genome annotation to identify exonic regions (Table 1). To distinguish cDNA amplicons from amplicons produced from residual traces of genomic DNA, we preferentially designed primers on either side of an intron. However, *eF1α, actin*, and *polyUBQ*, presented uncertainties in position of introns, or lacked introns, so primers were designed across a single exon.

We used PrimerQuest (Thornton and Basu 2015) to choose primer sequences using the following parameters for intercalating dye assays: melting temperature (Tm) of 60°C and GC content between 40 and 55%. Primers were purchased from Eurofins (Ebersberg, Germany) and are listed in Table 1. The specificity of each primer was checked using Basic Local Alignment Search Tool (23 November 2005, http://blast.ncbi.nlm.nih.gov/Blast.cgi) on *H. melpomene* genome and checking the absence of supplementary bands on electrophoresis gel.

Quantitative PCR

Quantitative PCR reactions were carried out in 96-well plates in a final volume of 10 µl with two technical replicates. Each 10 µl reaction contained 2 µl cDNA template (20-fold diluted), 5 µl FastStar Universal SYBR Green Master (ROX) (Roche, Bâle, Swiss), 0.5 µl of primer mix (final concentration of 0.25 µM for each primer) and 2.5 µl of ultrapure water. The negative control was a sample without template, replaced by nuclease free water only (NAC). The Master mix contained intercalating dye that fluoresces only when bound to double-stranded DNA, reporting on the PCR in real time by measuring fluorescence. Dyes do not fluoresce when free, as for instance when DNA is single-stranded. Quantitative PCR was carried out with a CFX96 Touch Real-Time PCR detection system (BIO-RAD, Hercules, CA, USA). The thermal cycling programme consisted of an initial activation of the FastStar Taq DNA polymerase during 10 min at 95°C, followed by 45 cycles of 95°C for 10 s and 60°C for 30 s. Finally, to confirm that a single amplicon product was produced by qPCR, a melt curve was generated as follows: at the end of the qPCR run, thermocycler was set to 65°C and fluorescence was measured. Temperature was then incrementally increased until 95°C (0.5°C each 5 s) whereas fluorescence kept being measured at each step. As temperature increases, double stranded DNA denaturation occurs, generating single-stranded DNA, and the associated fluorescent dye molecules dissociate, resulting in decreased fluorescence. The melt curve describes this fluorescence change during temperature decrease (dF/dT with F and T being fluorescence and temperature, respectively) for each genes studied. When several distinct PCR products are amplified within a qPCR, this generates multiple peaks in the melt curve, because different PCR products have different denaturation temperatures. For primer pairs located either side of

Table 1. Primer sequences for housekeeping genes

Gene name	Function	Primers Sequences	Size (bp)
actin	cytoskeletal protein	Fwd 5′ CTCCCTCGAGAAGTCCTATGAA 3′ Rev 5′ CCAAGAATGAGGGTTGGAAGAG 3′	101
annexin	Annexin IX-B	Fwd 5′ GAAACCCTTAGAAGACTGGATCG 3′ Rev 5′ GTTCTCAGTGTCGGCCATATT 3′	96
eF1α	Elongation factor 1α	Fwd 5′ GTGACATGAGGCAGACTGTAG 3′ Rev 5′ AGCGGCTTTGGTGACTTTA 3′	86
FK506BP	FK506 binding protein	Fwd 5′ GACCGTGGAAAGCCCTTCAAG 3′ Rev 5′ GTCCATAGGCATAGTCAGGAG 3′	130
polyABP	PolyA binding protein	Fwd 5′ GTTGAAATCTGAACGTTTGACACG 3′ Rev 5′ CTGCTTTAGTAGCTTCTTCAGG 3′	122
polyUBQ	Poly Ubiquitin	Fwd 5′CCTGACCAACAGAGGCTTATT 3′ Rev 5′ AGCACCAAGTGCAAAGTAGA 3′	101
RpL3	Ribosomal protein L3	Fwd 5′ GGGTTGTTGTATGGGACCTAAG 3′ Rev 5′ TCAGATTGATCTTCTCAAGAGCTG 3′	95
RPS3A	Ribosomal protein S3A	Fwd 5′ GAGATAAATTACCCGTGATGTGG 3′ Rev 5′ CTGGGTCTTTTCAGTACTTTTAC 3′	158
Tubulin	Microtubule structure	Fwd 5′ GACTTGGAGCCTACAGTAGTTG 3′ Rev 5′ CGCATCTTCCTTACCAGTGAT 3′	94

introns, melt curves are useful to check that only short PCR products were amplified from cDNA and that no larger size PCR product had been amplified from remaining gDNA.

Data Analysis

For each sample, the cycle in which fluorescence can first be detected, i.e., quantification cycle (*Cq* value), was estimated by the single threshold mode with the CFX 96 Touch Real-Time PCR detection system software (BIO-RAD, Hercules, CA, USA). This mode uses a single threshold value to calculate the *Cq* value based on the threshold crossing point of individual fluorescence curves.

Reaction Efficiency (*E*)

PCR amplification efficiency was established by means of calibration curves from the slope of the log-linear portion of the calibration curve. The reaction efficiency (*E*) describes the increase in amplicon per cycle. The theoretical maximum of 1.00 (100%) indicates that the amount of product doubles with each cycle. The value assigned to the efficiency of a PCR reaction is a measure of the overall performance of a real-time PCR assay. To evaluate and record the reaction efficiency of each primer pair, we generated a three-points standard curve using serial dilutions (1/10, 1/100, and 1/1,000) of a mix of cDNA.

The *Cq* value was then plotted against the log of the starting quantity (Pfaffl 2001; Vandesompele et al. 2002) and the efficiency (*E*) is calculated from the slope of the line derived from the standard curve with the following formula, $E = (10^{-1/\text{slope}} - 1) \times 100$.

Validation of Reference Genes

To evaluate whether the tested reference genes were expressed at similar relative level across different samples, we compared outputs obtained from three different tools commonly used for reference gene validation:

(1) The geNorm software (Vandesompele et al. 2002) calculates relative quantities from the *Cq* values and then gene expression stability measure (*M*) for a reference gene as the average pairwise variation, *V*, for that gene with all other tested reference genes in a given sample panel. Gene expression is considered stable when the stability parameter *M* is below 1.5. The lowest *M* values are produced by genes with the most stable expression; stepwise exclusion of the gene with the highest *M* value is then used to rank the tested genes according to their expression stability (Vandesompele et al. 2002).

(2) The Normfinder strategy (Andersen et al. 2004) is rooted in a mathematical model describing the expression values measured by RT-qPCR using separate analysis of sample subgroups (e.g., within a developmental stage and/or within color patterns). This algorithm then estimates both intra- and inter-group expression variations, and computes candidate gene stability values. This strategy provides a direct measure of the estimated expression variation of candidate reference genes across developmental stage or biological conditions. This prevents the introduction of an artefact error when using the candidate reference gene for comparing stages or conditions (Andersen et al. 2004).

(3) The Excel-based tool *BestKeeper* use raw data (*Cq* values) and PCR Efficiency (*E*) to determine the optimal housekeeping genes (Pfaffl et al. 2004). First, the estimation of expression stability is performed, based on the inspection of calculated variations, SD and coefficient of variance (CV) values. Then, the Excel-based tool can ordered genes from the most stably expressed, exhibiting the lowest variation, to the least stable one, exhibiting the highest variation. Furthermore, to estimate inter-gene relations of all possible candidate reference genes, pairwise correlation analyses are performed.

Results

We studied the stability of expression of nine candidate reference genes at two developmental stages using RT-qPCR, with the SYBR green as intercaling dye, and using three widely used programs.

Specificity of Amplification

The specificity of amplification was confirmed by the presence of a single band of the expected size for each primer pair on agarose gels following electrophoresis, and by visualizing the single-peak melt curves of the PCR products (Fig. 1). Additionally, C_qs ≥ 35 were observed for no template controls (NTC) and water controls (NAC). As long as the *Cq* value for the NTC is largely superior to the dynamic range of the DNA Standards (here *Cq* of NTC ≥ 35 comparing with *Cq* of cDNA samples ranging from 18 to 24), the NTC amplification can be ignored as it will have no significant effect on library quantitation. All these controls confirm the specificity of amplification and the absence of gDNA contamination, even in *eF1α, actin,* and *polyUBQ* genes whose amplification did not span over an intron.

Efficiency (*E*)

qPCR efficiency was then calculated from three-point standard curve with the cDNA mix for the eight candidate reference genes. Each point is the mean value obtained from the two technical replicates. PCR efficiency ranged from 57.1% (*polyUBQ*) to 110.8% (*tubulin*) (Table 2).

The PCR efficiency and the correlation coefficient R^2 , measuring how well the data fit the standard curve, reflects the linearity of the standard curve and the inhibition of PCR. Reference genes with at least 80% efficiency and $R^2 > 0.99$ are generally considered suitable for further RT-qPCR analysis (Bustin et al. 2009). We thus applied these thresholds to our candidate reference genes: the analysis of our calibration curves show that in seven out of our eight candidate reference genes, no significant inhibition of PCR was observed (Table 2). We thus excluded *FK506BP* and *polyUBQ* genes and considered the seven remaining genes (*actin*, *annexin*, *eF1α*, *polyABP*, *RpL3*, *RPS3A*, and *tubulin* genes) as robust and precise for RT-qPCR assays.

We used the *Cq*-value of the highest standard dilution (1000× diluted) as limit of detection (*LOD*), listed for all tested genes in Table 2. Across all tested genes, two samples had *Cq* values below the LOD values and were therefore removed from further analysis.

Estimation of Expression Level

Quantitative PCR was assayed on the seven remaining genes using SYBR Green to monitor levels of transcripts in wing tissue during the two developmental stages of different morphs of *H. numata*.

A suitable reference gene should display a constant expression level among tested samples; therefore we used the *Cq* values to compare the expression levels of the tested reference genes.

This reveals that the *annexin* gene was the least expressed, while *tubulin* was the most expressed. *Cq* values ranged from 15.33 to 29.01 cycles across samples. *eF1α* showed lowest levels of variation across the entire dataset, while *RPS3A* gene showed the highest level of variation (Fig. 2).

Selection of Reliable Reference Genes

The main criterion to choose a reference gene is its stability of expression across tissue types, developmental stages or physiological

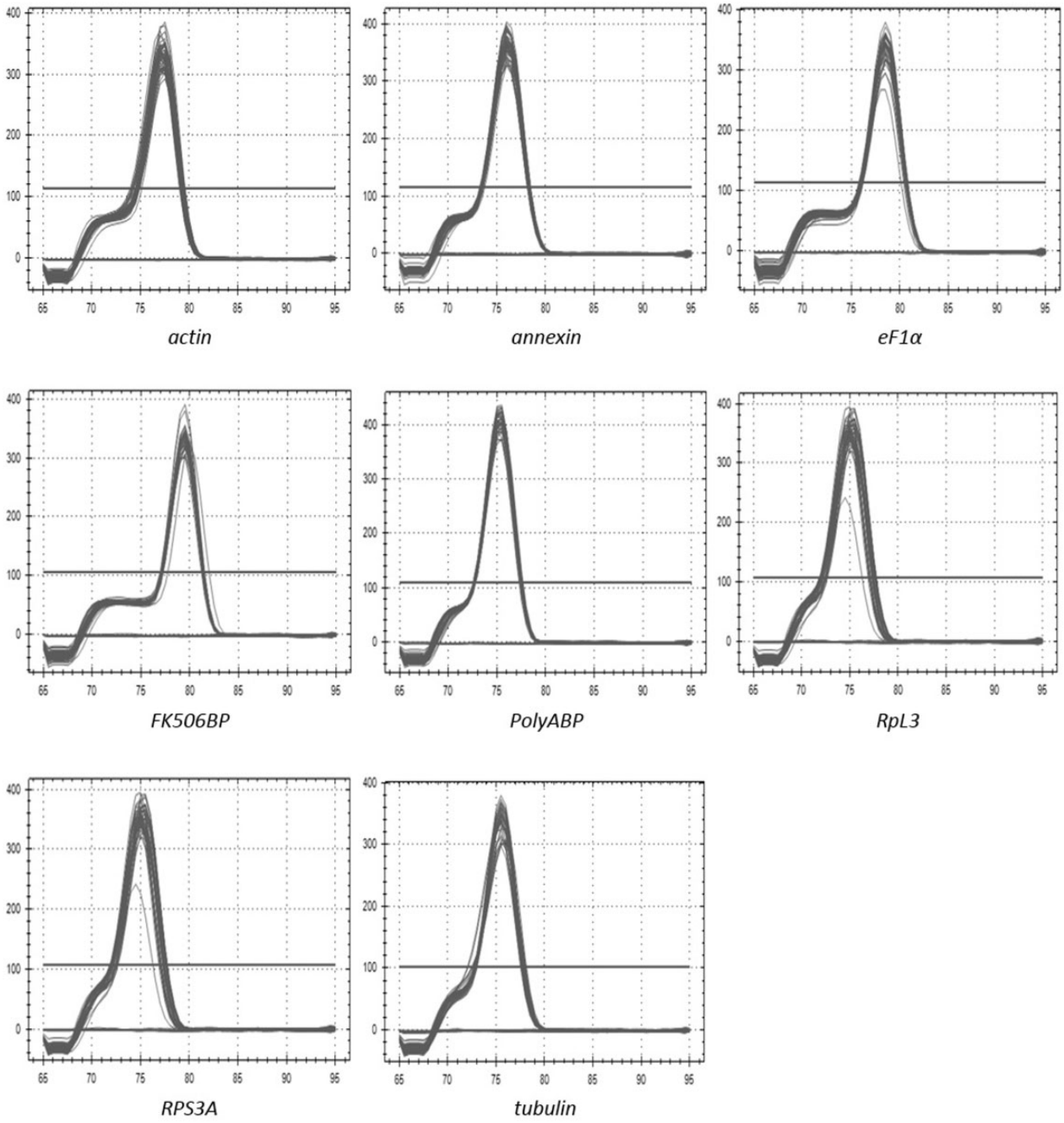

Fig 1. qPCR primer specificity. Melt curves, for all genes, are determined by plotting the negative first derivative of fluorescence versus the temperature in Celsius ($-d$(RFU)/dT).

conditions. We compared three different methods to infer expression stability.

g*eNorm*

In our samples, six out of seven genes had an *M* value below 1.0, which indicates high expression stability (Table 3). Comparison of pairwise variation in stability values among the seven genes indicated that *RpL3* and *eF1α* had the two most stable expressions in wing discs of *H. numata*. The *geNorm* program highlights the minimum number of reference genes for accurate normalization by calculating the pairwise variation $V_{n/n+1}$ between two sequential normalization factors ranked by stability. This tool is used to estimate the benefit from adding a supplementary reference gene from the list (Vandesompele et al. 2002). The lowest *V* value was found to be 0.080 for $V_{3/4}$ and is lower than the cut-off value of 0.15, suggesting that the first three reference genes are sufficient to accurately normalize gene expression data in these samples (*RpL3*, *eF1α*, and *PolyABP*) (Fig. 3). The two least stable candidates for *geNorm* software were *actin* (0.746) and *tubulin* (1.022).

N*ormFinder*

The expression stability of candidate reference genes was also calculated with *NormFinder* strategy. *NormFinder* analysis indicated that *eF1α* has the best stability value (stability value =

Table 2. Efficiency and expression of candidate reference genes.

Gene name	Efficiency	R^2	Slope	y-intercept	Average Cq	LOD
actinJTSS	92.9%	0.997	−3.504	11.896	18.16	22.34
annexin	96.4%	0.998	−3.412	18.160	24.22	28.41
eF1α	97.2%	1.000	−3.391	12.192	18.33	22.34
FK506BP	90.8%	**0.987**	−3.567	15.437	21.65	26.56
polyABP	87.1%	0.997	−3.677	15.045	21.98	26.19
polyUBQ	**57.1%**	**0.936**	−5.097	27.831	*ND*	*ND*
RpL3	90.2%	0.996	−3.582	13.441	19.65	24.15
RPS3A	96.2%	0.998	−3.416	12.887	19.21	23.27
tubulin	110.8%	0.994	−3.088	12.237	18.04	21.62

Values indicated in bold show values below thresholds, triggering an exclusion of the correspondent reference genes from further analysis. R^2, Correlation coefficient, Slope: slope of the line derived from the standard curve, y-intercept: theorical limit of reaction, LOD, limit of detection; ND, not determined.

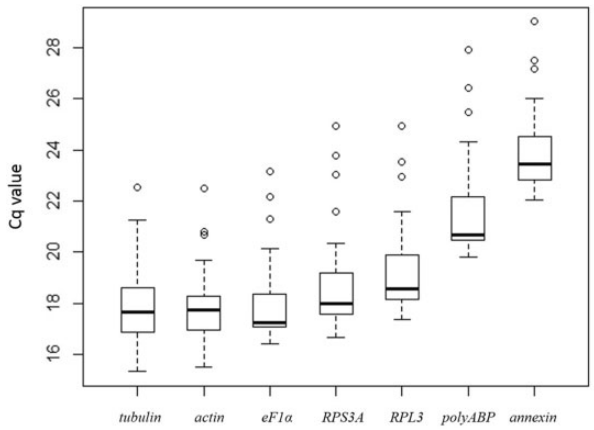

Fig. 2. Distribution of expression levels of candidate reference genes in the 21 wing discs. The expression level is shown through the *Cq* value.

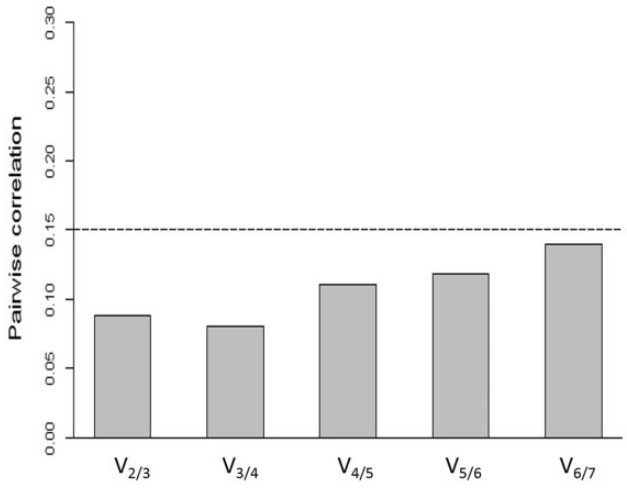

Fig. 3. geNorm pairwise variation (*V*) analysis to determine optimal number of reference genes for normalization in RT-qPCR reaction. The dotted line corresponds to the cut-off value of 0.15 for accurate normalization.

0.031). *Normfinder* analysis also indicated that the best combination of two genes was *eF1α* and *polyABP* with a stability value of 0.023. *Annexin gene* was ranked as the third most stable gene in *NormFinder* analysis (Stability value = 0.039). The two least stable genes were *actin* (Stability value = 0.053) and *tubulin* (Stability value = 0.084).

Table 3. Ranking of genes by expression stability estimated using *geNorm, NormFinder, BestKeeper*

Rank	geNorm (stability value)	NormFinder (stability value)	BestKeeper
1	*eF1α/RpL3* (0.250)	*eF1α/PolyABP (0.023)*	*polyABP*
2			*eF1α*
3	*polyABP* (0.292)	*annexin* (0.039)	*annexin*
4	*RPS3A* (0.355)	*RpL3* (0.048)	*RPS3A*
5	*annexin* (0.582)	*RPS3A* (0.048)	*RpL3*
6	*actin* (0.746)	*actin* (0.053)	*actin*
7	*tubulin* (1.022)	*tubulin* (0.084)	*tubulin*

B*estKeeper*

In our study, the analysis of *BestKeeper* results revealed that three housekeeping genes had an inacceptable variation in gene expression, higher than 1.00 (*RpL3*, *PolyABP*, and *RPS3A*). So, in our study, these genes cannot be correlated parametrically but only on their rank. These high SDs result from heterogeneous variance between groups of differently expressed genes; genes with low expression levels show different variance compared with genes with high expression. Nevertheless, the *BestKeeper index* ranked *polyABP*, *eF1α*, and *annexin* as the three most stably expressed genes.

Discussion

Here we showed substantial variation in expression levels across samples for some traditionally used housekeeping genes, highlighting the need for a proper testing of reference genes when transferring them across study species. We stress that accurate normalization of gene expression levels with robust reference genes is an absolute prerequisite for reliable RT-qPCR results. Moreover, when several reference genes are used simultaneously in a given experiment, the probability of a normalization bias decreases. Ferguson *et al.* (2010) already demonstrated that inappropriate reference gene selection for RT-qPCR normalization can have a profound influence on study conclusions influencing statistical power and perhaps leading to inaccurate data interpretation. Comparisons can be especially difficult when dealing with *in vivo* samples and comparing gene expression patterns between different individuals (Bustin, 2002). Even if RT-qPCR is widely used, there is no consensus as to which genes should be used for data normalization, and the selection of genes that are expressed at similar levels across all samples and different conditions of a study is still one of the challenges of this technique. Suzuki *et al.* (2000) have reviewed that *Glyceraldehyde-3-phosphate dehydrogenase* and *actin* genes usually used as reference genes are not perfect RNA quantitation controls because of their modulation under a variety of conditions in different cell types (Suzuki et al. 2000). The utility of reference genes must be experimentally validated for particular tissues or cell types and specific experimental designs (Bustin et al. 2009).

In our study, the *H. melpomene* genome (Altschul et al. 1997; Dasmahapatra et al. 2012) was used to select ours candidate reference genes. For the eight candidates housekeeping genes, the

specificity of amplification was experimentally validated with melt curves, NAC and NTC controls.

To determine the stability of gene expression and to select most appropriate reference genes, many statistical approaches exist and to date, there is no consensus on which approach gives valid results. The simultaneous use of more than one algorithm has been noted by other authors as producing highly correlated results and thus represents a good strategy for the selection of reference genes for RT-qPCR normalization (Reid et al. 2006; Paolacci et al. 2009).

The stability ranking obtained by *geNorm, NormFinder,* and *BestKeeper* was not identical for the genes with the most stable expression. In particular, the best normalization genes selected by the three programs were not the same. Nevertheless, there was substantial agreement when contrasting most stable *vs.* least stable groups of genes. Taken together, the most suitable reference genes appear to be *annexin*, *eF1α*, *polyABP,* and *RpL3* and three of these have thus been chosen for our further studies of expression in wing discs of *H. numata*. Here we detected some slight differences among the different algorithms tested. Likewise, results from other studies show slightly differences among distinct algorithms used (Demidenko et al. 2011; Wang et al. 2013; Yang et al. 2014) probably due to the differences in the statistical approaches. In our study, *BestKeeper* and *NormFinder* have ranked *polyABP, annexin,* and *eF1α* genes as the three most stable genes, whereas *geNorm* has ranked *RpL3*, *PolyABP*, and *eF1α* as optimal housekeeping genes. These differences in ranking may be attributed to the distinct underlying statistical algorithms. For example, the two most stable genes according to *geNorm* are the two genes that have the most identical expression pattern throughout the sample set. As discussed by different authors (Andersen et al. 2004; Bonefeld et al. 2008; Maccoux et al. 2007; Vandesompele et al. 2002), since *geNorm* determines stability by pairwise comparison of the variation of expression ratios, a careful choice of the tested candidate genes is required. In particular, candidate genes should not be co-regulated since this would lead to high similarity of expression between genes even if expression is not stable. In our study, special attention was paid to select genes which belong to different functional classes, significantly reducing the risk that our tested genes might be co-regulated. Moreover, *geNorm* provided an estimate of the ideal number of reference genes to be included for normalization which has a practical utility. In our study, *geNorm* advocates that the use of the two most stably expressed genes are sufficient for gene expression analyses ($V_{2/3} < 0.15$). However, to be congruent with MIQE guidelines [MIQE (Bustin et al. 2009)] which recommend the use of three reference genes for data normalization, we decided to select the three most stably reference genes for expression analyses.

The *NormFinder* algorithm can take into account experimental design covering various stages, such as various developmental stages or various morphs in our case. It should thus be preferentially used in experiments where the compared groups are distributed in a nested design, ensuring proper comparisons. This approach has been reported to perform in a more robust manner than *geNorm* and to be less sensitive to the presence of co-regulated genes (Andersen et al. 2004). The choice of candidate reference genes without functional relationships is thus less critical than in *geNorm*. *NormFinder* using a different statistical framework than *geNorm*, the *NormFinder* model attributes the three best stability values to *eF1α*, *polyABP,* and *annexin*.

For the *BestKeeper* tool, even if high standard deviations across samples prevent the use of correlation coefficient for gene expression analysis, the three best ranked candidate reference genes are *polyABP*, *eF1α,* and *annexin*.

Overall, the expression of the tested genes in wings dissected at the two developmental stages was sufficiently stable to use them as reference genes. The same reference genes can thus be reliably used to investigate candidate gene expression in both prepupal and pupal developmental stages.

Here, we report the validation of reference genes for the study of wing color pattern variation in *H. numata* using RT-qPCR, following MIQE recommendations. The aim was therefore to select genes with a stable expression across different color pattern variants from *H. numata*. This study is the first detailed evaluation of the expression stability of several candidate reference genes to be used for normalization in RT-qPCR studies in *Heliconius*. Based on the comparison of statistical methods analysing the stability of expression of the eight housekeeping genes tested, we decided to use three genes as reference for any studies of gene expression in developing wing disc of *H. numata*: *eF1α*, *polyABP* and *RpL3* providing a robust experimental framework to identify genes functionally involved in phenotypic variations. Despite the robustness of our determined reference genes, they will be used only in the particular case of wing discs and at particular stages. However, they should be good candidates for other tissues and developmental stages experiments in *Heliconius* and they should be tested on each particular case.

Acknowledgements

This work was supported by the French National Agency for Research (ANR) grant DOMEVOL attributed to V.L. (ANR-13-JSV7-0003-01). Molecular work was performed with the support of the BoEM laboratory at the Muséum National d'Histoire Naturelle, Paris, France. The 'Service de Systématique Moléculaire' (UMS7200) at the Muséum National d'Histoire Naturelle, Paris, France lent us the CFX96 Touch Real-Time PCR detection system (BIO-RAD) for experiments.

References Cited

Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25: 3389–3402.

Andersen, C. L., J. L. Jensen, and T. F. Orntoft. 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 Res. 64: 5245–5250.

Benson, W. W. 1972. Natural Selection for Miillerian Mimicry in Heliconius erato in Costa Rica. Science 176: 936–939.

Bonefeld, B. E., B. Elfving, and G. Wegener. 2008. Reference genes for normalization: a study of rat brain tissue. Synapse 62: 302–309.

Brown, K., & W. Benson. 1974. Adaptative polymorphism associated with multiple müllerian mimicry in Heliconius numata (Lepid. Nymph.). Biotropica 6: 205–228.

Brown, K. S. 1981. The biology of Heliconius and related genera. Annu. Rev. Entomol. 26: 427–457.

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

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

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

Dasmahapatra, K. K., J. R. Walters, A. D. Briscoe, J. W. Davey, A. Whibley, N. J. Nadeau, A. V. Zimin, D.S.T. Hughes, L. C. Ferguson, S. H. Martin,

et al. 2012. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. Nature 487: 94–98.

Demidenko, N. V., M. D. Logacheva, and A. A. Penin. 2011. Selection and validation of reference genes for quantitative real-time PCR in buckwheat (Fagopyrum esculentum) based on transcriptome sequence data. PLoS One 6: e19434.

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

Ferguson, L., S. F. Lee, N. Chamberlain, N. Nadeau, M. Joron, S. Baxter, P. Wilkinson, A. Papanicolaou, S. Kumar, T. J. Kee, et al. 2010. Characterization of a hotspot for mimicry: assembly of a butterfly wing transcriptome to genomic sequence at the HmYb/Sb locus. Mol Ecol 19: 240–254.

Ferguson, L. C., and C. D. Jiggins. 2009. Shared and divergent expression domains on mimetic Heliconius wings. Evol. Dev. 11: 498–512.

Ferguson, L. C., L. Maroja, and C. D. Jiggins. 2011. Convergent, modular expression of ebony and tan in the mimetic wing patterns of Heliconius butterflies. Dev. Genes Evol. 221: 297–308.

Heliconius Genome, C. 2012. Heliconius Genome Project (http://www.butterflygenome.org/).

Heliconius Genome Consortium. 2012. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. Nature 487: 94–98.

Ishitani, R., K. Sunaga, A. Hirano, P. Saunders, N. Katsube, and D. M. Chuang. 1996. Evidence that glyceraldehyde-3-phosphate dehydrogenase is involved in age-induced apoptosis in mature cerebellar neurons in culture. J. Neurochem. 66: 928–935.

Joron, M., L. Frezal, R. T. Jones, N. L. Chamberlain, S. F. Lee, C. R. Haag, A. Whibley, M. Becuwe, S. W. Baxter, L. Ferguson, et al. 2011. Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. Nature 477: 203–206.

Joron, M., C. D. Jiggins, A. Papanicolaou, and W. O. McMillan. 2006. Heliconius wing patterns: an evo-devo model for understanding phenotypic diversity. Heredity (Edinb) 97: 157–167.

Joron, M., I. Wynne, G. Lamas, and J. Mallet. 1999. Variable selection and the coexistence of multiple mimetic forms of the butterfly Heliconius numata. Evol. Ecol. 13: 721–754.

Kozak, K. M., N. Wahlberg, A. Neild, K. K. Dasmahapatra, J. Mallet, and C. D. Jiggins. 2015. Multilocus species trees show the recent adaptive radiation of the mimetic Heliconius butterflies. Syst. Biol. 64: 505–524.

Kunte, K., W. Zhang, A. Tenger-Trolander, D. H. Palmer, A. Martin, R. D. Reed, S. P. Mullen, and M. R. Kronforst. 2014. Doublesex is a mimicry supergene. Nature 507: 229–232.

Le Poul, Y., A. Whibley, M. Chouteau, F. Prunier, V. Llaurens, and M. Joron. 2014. Evolution of dominance mechanisms at a butterfly mimicry supergene. Nat. Commun. 5: 5644.

Lu, Y., M. Yuan, X. Gao, T. Kang, S. Zhan, H. Wan, and J. Li. 2013. Identification and validation of reference genes for gene expression analysis using quantitative PCR in Spodoptera litura (Lepidoptera: Noctuidae). PLoS One 8: e68059.

Maccoux, L. J., D. N. Clements, F. Salway, and P. J. Day. 2007. Identification of new reference genes for the normalisation of canine osteoarthritic joint tissue transcripts from microarray data. BMC Mol. Biol. 8: 62.

Merrill, R. M., K. K. Dasmahapatra, J. W. Davey, D. D. Dell'Aglio, J. J. Hanly, B. Huber, C. D. Jiggins, M. Joron, K. M. Kozak, V. Llaurens, et al. 2015. The diversification of Heliconius butterflies: what have we learned in 150 years? J. Evol. Biol. 28: 1417–1438.

Müller, F. 1879. Ituna and Thyridia; a remarkable case of mimicry in butterflies. Trans. Entomol. Soc. Lond. XX–XXIX.

Nijhout, H. F. 1991. The development and evolution of butterfly wing patterns, p. 297. Part of Smithsonian series in comparative evolutionary biology. Smithsonian Institution Press, Washington.

Nishikawa, H., M. Iga, J. Yamaguchi, K. Saito, H. Kataoka, Y. Suzuki, S. Sugano, and H. Fujiwara. 2013. Molecular basis of wing coloration in a Batesian mimic butterfly, Papilio polytes. Sci. Rep. 3: 3184.

Paolacci, A. R., O. A. Tanzarella, E. Porceddu, and M. Ciaffi. 2009. Identification and validation of reference genes for quantitative RT-PCR normalization in wheat. BMC Mol. Biol. 10: 11.

Petersen, B. H., R. Rapaport, D. P. Henry, C. Huseman, and M. W. Moore. 1990. Effect of treatment with biosynthetic human growth hormone (GH) on peripheral blood lymphocyte populations and function in growth hormone-deficient children. J. Clin. Endocrinol. Metab. 70: 1756–1760.

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

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

Pijpe, J., N. Pul, S. van Duijn, P. M. Brakefield, and B. J. Zwaan. 2011. Changed gene expression for candidate ageing genes in long-lived Bicyclus anynana butterflies. Exp. Gerontol. 46: 426–434.

Reed, R. D, W. O. McMillan, and L. M. Nagy. 2008. Gene expression underlying adaptive variation in Heliconius wing patterns: non-modular regulation of overlapping cinnabar and vermilion prepatterns. Proc. Biol. Sci. 275: 37–45.

Reid, K. E., N. Olsson, J. Schlosser, F. Peng, and S. T. Lund. 2006. An optimized grapevine RNA isolation procedure and statistical determination of reference genes for real-time RT-PCR during berry development. BMC Plant Biol. 6: 27.

Singh, R., and M. R, Green. 1993. Sequence-specific binding of transfer RNA by glyceraldehyde-3-phosphate dehydrogenase. Science 259: 365–368.

Sun, M., M. X. Lu, X. T. Tang, and Y. Z. Du. 2015. Exploring valid reference genes for quantitative real-time PCR analysis in Sesamia inferens (Lepidoptera: Noctuidae). PLoS One 10: e0115979.

Suzuki, T., P. J. Higgins, and D. R. Crawford. 2000. Control selection for RNA quantitation. Biotechniques 29: 332–337.

Thellin, O., W. Zorzi, B. Lakaye, B. De Borman, B. Coumans, G. Hennen, T. Grisar, A. Igout, and E. Heinen. 1999. Housekeeping genes as internal standards: use and limits. J. Biotechnol. 75: 291–295.

Thornton, B., and C. Basu. 2015. Rapid and simple method of qPCR primer design (http://eu.idtdna.com/Primerquest/Home/Index: PCR Primer Design, pp. 173–179. *In* C. Basu (ed.). Springer New York, New York.

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

Vesentini, N., C. Barsanti, A. Martino, C. Kusmic, A. Ripoli, A. Rossi, and A. L'Abbate. 2012. Selection of reference genes in different myocardial regions of an in vivo ischemia/reperfusion rat model for normalization of antioxidant gene expression. BMC Res. Notes 5: 124.

Wang, M., Q. Wang, and B. Zhang. 2013. Evaluation and selection of reliable reference genes for gene expression under abiotic stress in cotton (Gossypium hirsutum L.). Gene 530: 44–50.

Xiao, X., J. Ma, J. Wang, X. Wu, P. Li, and Y. Yao. 2014. Validation of suitable reference genes for gene expression analysis in the halophyte Salicornia europaea by real-time quantitative PCR. Front. Plant Sci. 5: 788.

Yang, H., J. Liu, S. Huang, T. Guo, L. Deng, and W. Hua. 2014. Selection and evaluation of novel reference genes for quantitative reverse transcription PCR (qRT-PCR) based on genome and transcriptome data in Brassica napus L. Gene 538: 113–122.