

# Identification of suitable reference genes for quantitative gene expression analysis in clam *Cyclina sinensis* under salinity stress and *Vibrio* infection\*

Fengjuan JIANG, Qingyao WANG, Jingjing DU, Fu LÜ, Qing NIE\*\*, Weihong ZHAO\*\*

College of Marine and Biological Engineering, Yancheng Institute of Technology, Yancheng 224051, China

Received Oct. 13, 2021; accepted in principle Dec. 7, 2021; accepted for publication Dec. 21, 2021

© Chinese Society for Oceanology and Limnology, Science Press and Springer-Verlag GmbH Germany, part of Springer Nature 2023

**Abstract** The appropriate reference gene is a prerequisite for accurate normalization of gene expression level, and research on suitable reference genes in clam *Cyclina sinensis* is scarce. To improve the situation, we selected five commonly used housekeeping genes, including  $\beta$ -actin, Elongation factor 1- $\alpha$  (*EF1- $\alpha$* ), Glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*), 40S ribosomal protein S18 (*RPS18*), and Tubulin  $\alpha$  (*TUB- $\alpha$* ), then evaluated their expression stability in different adult tissues and under different experimental treatments (salinity stress and *Vibrio parahaemolyticus* infection). Their expression stability was analyzed by three frequently used programs, geNorm, NormFinder, and BestKeeper. This analysis indicated that multiple genes should be used for normalization, and we concluded that the reference gene combination *GAPDH-RPS18- $\beta$ -actin*, should be used for qRT-PCR analysis in different tissues of *C. sinensis* under normal physiological conditions. For the clams under salinity stress and *Vibrio* infection, *EF1- $\alpha$ -GAPDH-RPS18* was recommended as the gene combination for qRT-PCR normalization. *TUB- $\alpha$*  was generally poorly ranked by all programs, and should not be used in future studies. This study should provide fundamental support for accurate quantitative gene expression analysis of this species.

**Keyword:** *Cyclina sinensis*; reference gene; different tissues; salinity stress; *Vibrio* infection

## 1 INTRODUCTION

The clam *Cyclina sinensis* is naturally distributed along the coastline of China, Korea, and Japan. For the nutritious and economic importance of this species, *C. sinensis* has become one of the main maricultural shellfish in eastern China since it was developed in the 1980s (Liu et al., 2002). As an intertidal clam, salinity stress is one of the abiotic stresses frequently suffered by *C. sinensis*. Although *C. sinensis* belongs to the euryhaline organism, salinity changes can also affect the growth, survival, embryonic development, and energy metabolism in clams (Ni et al., 2020). Additionally, *Vibrio* infection is another main stress suffered by clams. The clams live in shallow water that is rich in pathogens, and research suggested that *Vibrio* is the main pathogen that causes the mass mortality of clams (Yue et al., 2010; Ni et al., 2020). In recent years, the adverse effects of salinity stress and *Vibrio* infection on the aquaculture industry of *C. sinensis* are

becoming serious. Therefore, researches on salinity tolerance and immune mechanism in *C. sinensis* have been increasing gradually (Lin et al., 2013; Ren et al., 2015; Ni et al., 2021; Sun et al., 2021). In most of these studies, gene expression analysis is one of the important ways to study gene function.

Quantitative real-time PCR (qRT-PCR) has been widely used to quantify gene expression levels owing to its high accuracy, sensitivity, and throughput. qRT-PCR including absolute and relative quantitative methods, of which relative quantitation has widely been used in the analysis of differentially expressed genes. In relative quantitation, the expression level of target gene is usually normalized according to stably expressed genes, known as reference genes.

\* Supported by the funding for school-level research projects of Yancheng Institute of Technology (No. xjr2019047) and the National Natural Science Foundation of China (No. 31902362)

\*\* Corresponding authors: nieqing0422@163.com; misszwh@163.com

The most frequently used endogenous references are the housekeeping genes, which are always expressed at the more or less constant level that is essential to the basic processes of cell living (Livak and Schmittgen, 2001; Stephenson, 2010). Numerous studies suggested that housekeeping genes were not consistently expressed under different tissue types (Jin et al., 2021), developmental stages (Huan et al., 2016), and experiment treatments (Chen et al., 2020). Therefore, the selection and validation of appropriate reference genes are critical for the accuracy of gene expression analysis by qRT-PCR.

To date, an increasing number of housekeeping genes have been evaluated to identify reference genes for qRT-PCR normalization under specific conditions in bivalve mollusks. For instance, the Elongation factor 1- $\alpha$  (*EF1- $\alpha$* ), Glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*), Tubulin- $\alpha$  (*TUB- $\alpha$* ), and 40S ribosomal protein S20 (*RPS20*), in the various combinations suggested by geNorm, NormFinder, and BestKeeper, were suitable for the normalization in clam *Ruditapes philippinarum* under the copper treated hemocytes (Volland et al., 2017). In oyster *Crassostrea gigas*, Ribosomal protein L7 (*RL7*) and Ribosomal protein S18 (*RPS18*) could be the best reference genes for normalization of qRT-PCR expression data in OsHV-1 infected larvae (Du et al., 2013). In king scallop *Pecten maximus*, NADH dehydrogenase (ubiquinone) 1 alpha subcomplex subunit 7 (*ndufa7*), *RPS $\alpha$* , and *EF1- $\alpha$*  were the most stable genes in the ovary, and the *18s rRNA*, *ndufa7*, and *GAPDH* were the best-ranked reference genes in testis (Mauriz et al., 2012). Therefore, there was no perfect reference gene that was appropriate for use in all experimental conditions, and the stability of reference genes should be validated for the conditions of studies.

For clam *C. sinensis*, systematic evaluation of reference genes that can be used among different tissues or experiment treatments is lacking, and  $\beta$ -actin is most frequently used as a qRT-PCR reference gene without validation (Ni et al., 2021; Sun et al., 2021). To improve this situation, we selected five commonly used housekeeping genes, including  $\beta$ -actin, *EF1- $\alpha$* , *GAPDH*, *RPS18*, and *TUB- $\alpha$* , and evaluated their expression stability in different adult tissues and under low salinity stress and *Vibrio parahaemolyticus* infection using three frequently used programs, geNorm, NormFinder, and BestKeeper. To our knowledge, this is the first systematic analysis of reference genes for normalization of qRT-PCR data in

*C. sinensis*. These results will facilitate more accurate and reliable gene expression studies of this species.

## 2 MATERIAL AND METHOD

### 2.1 Sample collection

The one-year-old clams *C. sinensis* with a mean weight of  $8.23 \pm 0.89$  g used in this study were obtained from a local market in Yancheng, China. All clams were acclimated at 26 °C in continuous aerated artificial seawater (salinity 25) for one week before the experiments. Five tissues from four clams under normal physiological conditions, including hepatopancreas, mantle, gill, foot, and adductor muscle were dissected, immediately frozen in liquid nitrogen, and then stored at -80 °C until tissue distribution analysis of candidate reference genes could be conducted.

As an intertidal clam, salinity stress and *Vibrio* infection are the main environmental stresses suffered by *C. sinensis*. Therefore, two experimental treatments were conducted in this study, including low salinity stress and *V. parahaemolyticus* immersion challenge. For the low salinity stress, clams were transferred from the seawater with salinity 25 to 5. To obtain a global view of the expression of candidate reference genes during salinity stress, we collected samples at 0 h, 12 h, 1 d, 3 d, 5 d, and 7 d. Five clams were randomly collected at each time point, the hepatopancreases of clams were dissected and preserved in liquid nitrogen for RNA extraction.

For the *Vibrio* challenge, clams were immersed in the seawater (salinity 25) with  $\sim 1 \times 10^7$  colony forming units (CFU)/mL of *V. parahaemolyticus*. The *Vibrio* challenge test was conducted as described in Jiang et al. (2018) with minor modification. During the *Vibrio* challenge, very low mortality was observed during the first 5 days, while the mortality increased drastically around 10 d, exceeding 50%. Five clams were randomly collected at 0 h, 12 h, 1 d, 2 d, 3 d, and 5 d post-infection respectively, the gills of clams were dissected and preserved in liquid nitrogen for RNA extraction.

### 2.2 RNA extraction and reverse transcription

Total RNA was extracted from samples using an E.Z.N.A Total RNA Kit II (Omega Bio-Teck, USA) according to the manufacturer's instructions. RNA quality was verified by agarose gel electrophoresis and spectrophotometry. About 1- $\mu$ g total RNA was reverse-transcribed in a 20- $\mu$ L reaction using the PrimeScript

**Table 1 The five candidate housekeeping genes in this study**

Gene ID	Gene name (abbreviation)	Functional class
Cluster-38134.8968	Beta-actin ( <i>β-actin</i> )	Cytoskeletal protein
Evm.TU.Hic_asm_7.1527	Elongation factor 1-alpha ( <i>EF1-α</i> )	Translation
Evm.TU.Hic_asm_4.304	Glyceraldehyde-3-phosphate dehydrogenase ( <i>GAPDH</i> )	Glycolysis
Evm.TU.Hic_asm_5.958	40S ribosomal protein S18 ( <i>RPS18</i> )	Translation
Evm.TU.Hic_asm_3.89	Tubulin alpha-1 ( <i>TUB-α</i> )	Cytoskeletal protein

**Table 2 Primers used in this study**

Gene	Primer (5'→3')	Product length (bp)	<i>E</i>	Coefficient ( <i>R</i> <sup>2</sup> )
<i>β-actin</i>	F: CAAACATGGCAGATGACGATAAT	165	0.99	0.996 7
	R: TGTCCCTTCTGACCCATACCAACC			
<i>EF1-α</i>	F: CTGGAAGAGTTACCAAAGAT	128	1.04	0.995 5
	R: ATGTCACGCACAGCAAAAACG			
<i>GAPDH</i>	F: CCAGTTTGGTATCAAGGAAGGAC	142	1.00	0.995 9
	R: AGCACCAGTTGAGGATGGAATAA			
<i>RPS18</i>	F: GTGTAGGTCGTCGCTATGCCA	227	1.00	0.997 5
	R: GTCTTCACGGAGCTTGTGTGC			
<i>TUB-α</i>	F: CTACAGGCAGTTGTCCACCCC	166	1.07	0.996 2
	R: GATGAGGAATCCCTGGAGACCG			

RT reagent kit with gDNA Eraser (TaKaRa, Japan) according to the manufacturer's protocol. To detect residual genomic DNA contamination in the cDNA samples, RNA samples were pooled to prepare an RT-minus negative control in which reverse transcriptase was excluded from the reaction mixture. No influence from genomic DNA was confirmed by the results that there was no amplification or high enough cycle threshold value (*C<sub>t</sub>*) values (e.g., >35) for all primer pairs when the RT-minus negative control was used as a template in the subsequent qRT-PCR assay.

### 2.3 Candidate reference genes selection

We collected five housekeeping genes as candidate reference genes (Table 1), including beta-actin (*β-actin*), Elongation factor 1-alpha (*EF1-α*), Glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*), 40S ribosomal protein S18 (*RPS18*), and Tubulin alpha-1 (*TUB-α*). The gene sequences of candidate reference genes were obtained from the genome and transcriptome databases of the clam *C. sinensis* (Wei et al., 2020). The fragments amplified by primers of these sequences were verified and then were provided in Supplementary File 1.

### 2.4 Primer specificity and amplification efficiency

Primers were designed using Primer premier 5.0 with the following criteria: primer *T<sub>m</sub>*, 60±5 °C;

primer length, 20–25 bp; expected products, 100–250 bp. Primer specificity and amplification efficiency of each pair primer were examined before they were used to estimate gene expression level. Amplicons of each primer pair were tested by 1.5% agarose gel electrophoresis to verify the product's size and specificity. Amplification specificity was also certified when only a single peak was observed in melting curve analysis in a qRT-PCR assay.

The amplification efficiency (*E*) of each primer pair was determined by qRT-PCR, using a serial dilution of pooled cDNA samples (×1, ×4, ×16, ×64, and ×256 dilutions). Amplification efficiency was calculated based on the slope of a standard curve generated by plotting *C<sub>t</sub>* values against the relative quantity of templates, the calculation formula is  $E=4^{-1/\text{slope}}$ . Primer sequences were adjusted until all primers had good amplification specificity and had amplification efficiencies of 0.90–1.10 (Stephenson, 2010). Primer information is provided in Table 2.

### 2.5 Quantitative real-time PCR

The qRT-PCR was conducted in a 96-well plate using CFX Connect Real-Time PCR Detection System (Bio-Rad, USA). Each reaction was performed in triplicate and a 10-μL volume containing 2×TB Green Premix Ex Taq II (TaKaRa, Japan), 0.4 μmol/L of each primer, and 0.5 μL of cDNA. The qRT-PCR

program was 95 °C for 30 s, and 40 cycles of 95 °C for 5 s, 55 °C for 30 s, and 72 °C for 30 s. A melting curve analysis was performed after each reaction to confirm specificity.

## 2.6 Gene expression stability analysis

Gene expression stability was evaluated using three frequently used programs, BestKeeper V1, NormFinder, and geNorm V3.5. Raw  $C_t$  values and amplification efficiency were directly submitted to BestKeeper. BestKeeper calculates a BestKeeper Index and evaluates gene expression stability based on correlation analysis between a gene and the BestKeeper Index, genes with high Pearson's correlation coefficient ( $r$ ) and low  $P$  values are considered to be stably expressed (Pfaffl et al., 2004). For NormFinder and geNorm, data were transformed to relative quantification data by normalizing to the sample with the highest  $C_t$  value using the formula: relative quantification data =  $(1+E)^{\Delta C_t}$ , where  $\Delta C_t$  = the highest  $C_t$  value -  $C_t$  value of the sample, and  $E$  represents the amplification efficiency. After transforming the raw  $C_t$  values into relative quantification data, the NormFinder calculates intra- and intergroup variations which are then combined into a stability value, the gene with the lowest stability value, was the best reference gene (Andersen et al., 2004). The geNorm calculates the value of gene expression stability ( $M$ ) for all candidate genes, and a lower  $M$  value indicated a more stable expression. For determining the optimal number of internal parameters required for qRT-PCR, geNorm calculates the pairwise variation  $V_n/V_{n+1}$  which represents the variation between using  $n$  most stable genes and using  $n+1$  most stable genes. This evaluation uses 0.15 as the threshold value, when  $V_n/V_{n+1} < 0.15$ ,  $n$  reference genes are sufficient for accurate normalization, otherwise,  $n+1$  internal reference genes are required (Vandesompele et al., 2002).

## 2.7 Quantification of target genes using different normalization strategies

The AMP-activated protein kinase  $\alpha 2$  gene (*AMPK $\alpha 2$* , F: 5'-GAGTTCCAAATCAAGGAGCCG-3', R: 5'-GGCATTCTTTCAGGGTGGG-3') and Interleukin17-1 gene (*IL17-1*, F: 5'-TCGTTGAGACTTTTGACACCAGC-3', R: 5'-GGACACGAGGACATTTCACTTGC-3') were selected as the target genes to test the effect of using different normalization strategies. The expression levels of *AMPK $\alpha 2$*  and *IL17-1* were determined in clams under

salinity stress and *Vibrio* infection respectively. The relative expressions of the target genes were calculated through the  $2^{-\Delta\Delta C_t}$  method using either commonly used  $\beta$ -actin or the optimal gene combination as reference (Livak and Schmittgen, 2001). When multiple reference genes were recommended for normalization, the geometric mean of their  $C_t$  values was applied (Vandesompele et al., 2002). One-way analysis of variance (ANOVA) was used to determine whether the relative expression patterns derived from the two normalization strategies were different.

## 3 RESULT

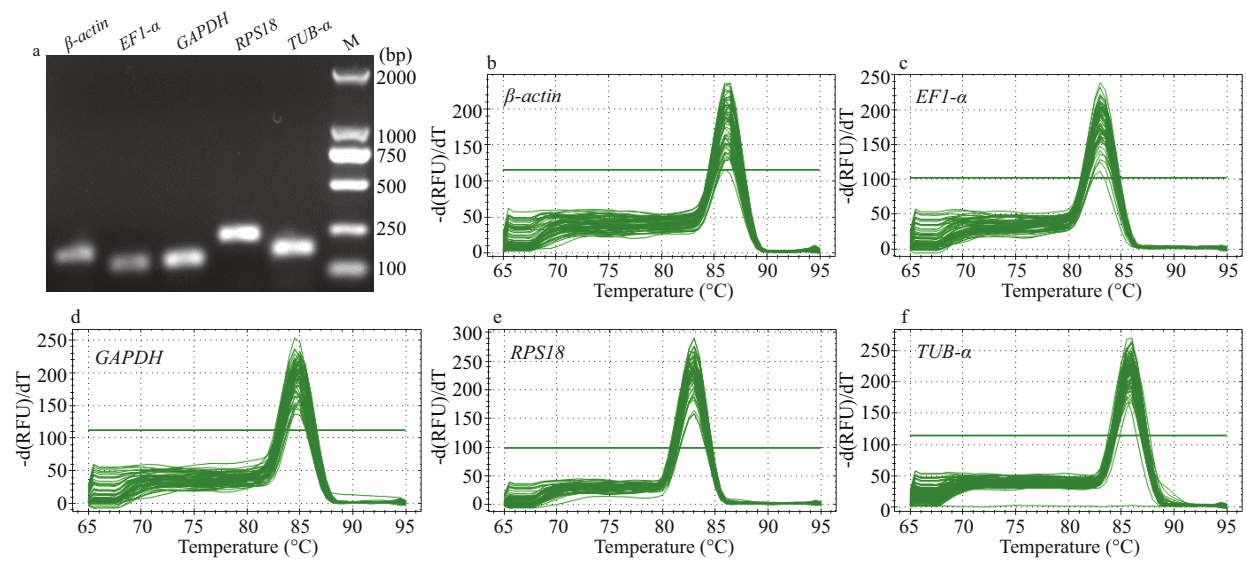
### 3.1 Primer specificity and amplification efficiency

The performance of each primer pair was validated for both amplification specificity and efficiency. The amplification products of each primer pair were confirmed by agarose gel electrophoresis, appeared as a single band with the expected size (Fig. 1a). Then the specificity of the products from each primer pair was also validated as only a single peak was observed in melting curve analysis (Fig. 1b-f). The amplification efficiency of all the primer pairs varied from 0.99 for  $\beta$ -actin to 1.07 for *TUB- $\alpha$*  (Table 2). Therefore, the primers designed were acceptable for further qRT-PCR assays.

### 3.2 The expression of candidate reference genes

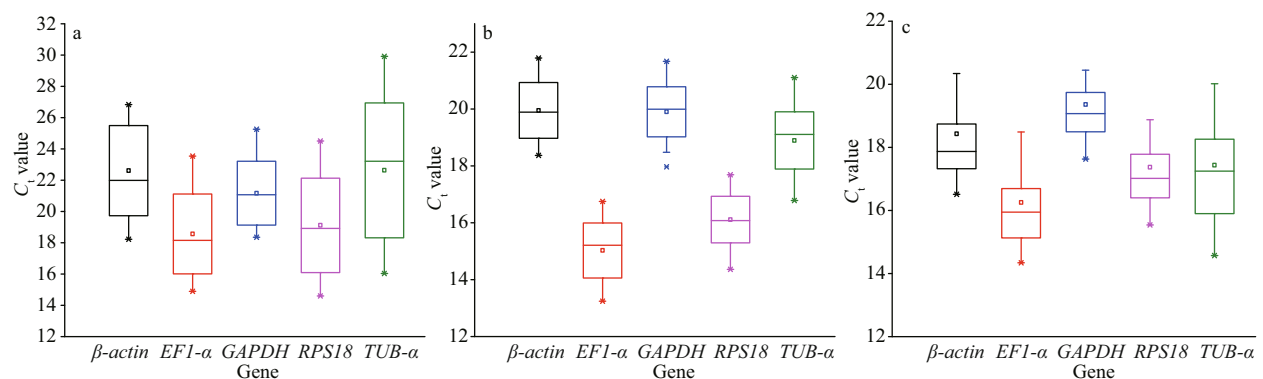
The boxplot graphs depicting the distributions of  $C_t$  values were shown in Fig. 2, and all  $C_t$  values are provided in Supplementary Tables S1-S3. In the different tissue samples tested, the  $C_t$  values of the five genes ranged from 14.60 to 29.92 (Supplementary Table S1). As shown in Fig. 2a, the  $C_t$  value of *EF1- $\alpha$*  was lower than those of the other four genes, indicating that *EF1- $\alpha$*  has a relatively higher expression level than the other genes. Among these five candidate reference genes, the  $C_t$  value distribution of *TUB- $\alpha$*  was the most discrete, indicating that the expression of *TUB- $\alpha$*  was unstable in different tissues of *C. sinensis* (Fig. 2a).

In the low salinity stress, the  $C_t$  values of the five genes in the hepatopancreas ranged from 13.24 to 21.79 (Supplementary Table S2). As shown in Fig. 2b, the gene expression levels of *GAPDH* and  $\beta$ -actin were relatively lower than the others, and the most highly expressed gene was *EF1- $\alpha$* . In the *Vibrio* immersion challenge, the  $C_t$  values of the candidate genes in gills ranged from 14.34 to 27.05 (Supplementary Table S3). As shown in Fig. 2c, the most highly expressed



**Fig.1** Primer specificity of five candidate reference genes in *C. sinensis*

a. PCR amplification products of candidate reference genes shown by agarose gel electrophoresis; b–f. melting curve of candidate reference genes in qRT-PCR.



**Fig.2** Boxplot graph depicting the distributions of  $C_t$  values of the five candidate reference genes in *C. sinensis*

a. the  $C_t$  values of different adult tissues under normal physiological conditions; b. the  $C_t$  values of the hepatopancreas under low salinity stress; c. the  $C_t$  values of the gills under *Vibrio* infection. The outlier is indicated with \*.

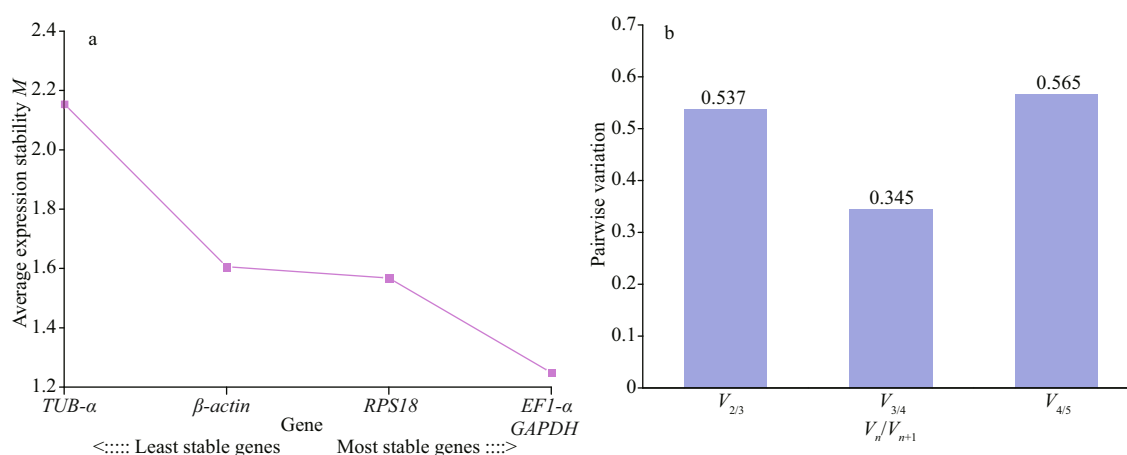
gene was also *EF1- $\alpha$* , and the gene expression level of *GAPDH* was lower than the others. Therefore, the expression levels of candidate reference genes might be more affected by tissue types and experimental treatments, a simple comparison of the raw  $C_t$  values for the candidate reference genes could not provide sufficient information. So, the following analysis was conducted using three different statistical algorithms for reference genes identification.

### 3.3 Expression stability analysis of candidate genes in different tissues

The expression level data of the candidate reference genes were analyzed using three programs, geNorm, NormFinder, and BestKeeper. For different tissues of *C. sinensis* under normal physiological conditions, geNorm analysis showed that *EF1- $\alpha$*  and *GAPDH* were the most stable genes, which were followed

by *RPS18*,  *$\beta$ -actin*, and *TUB- $\alpha$*  (Table 3, Fig.3a). The optimal number of reference genes required for accurate normalization in qRT-PCR was also calculated by geNorm, the result showed that there was no  $V_{n/n+1}$  value below 0.15, the lowest value was  $V_{3/4}$  (0.345), indicating that using three genes could be preferred for normalization (Fig.3b). NormFinder calculated gene expression stability values, which revealed that *RPS18* was the most stably expressed gene, which was followed by  *$\beta$ -actin*, *GAPDH*, *EF1- $\alpha$* , and *TUB- $\alpha$* . The five candidate reference genes were further evaluated using BestKeeper. The top three ranked genes were *RPS18*,  *$\beta$ -actin*, and *GAPDH*, these results were similar to those from NormFinder (Table 3).

While the results from different programs varied, they share many similarities, especially for NormFinder and BestKeeper. *TUB- $\alpha$*  was generally



**Fig.3 Expression stability of candidate reference genes analyzed by geNorm in different tissues of *C. sinensis***

a. the  $M$  values of candidate reference genes are calculated by geNorm. The lower  $M$  value indicates a more stable expression; b. the optimal number of reference genes required for normalization was calculated using pairwise variation analysis ( $V_n/V_{n+1}$ ).

**Table 3 Ranks of the candidate reference genes obtained from the three programs**

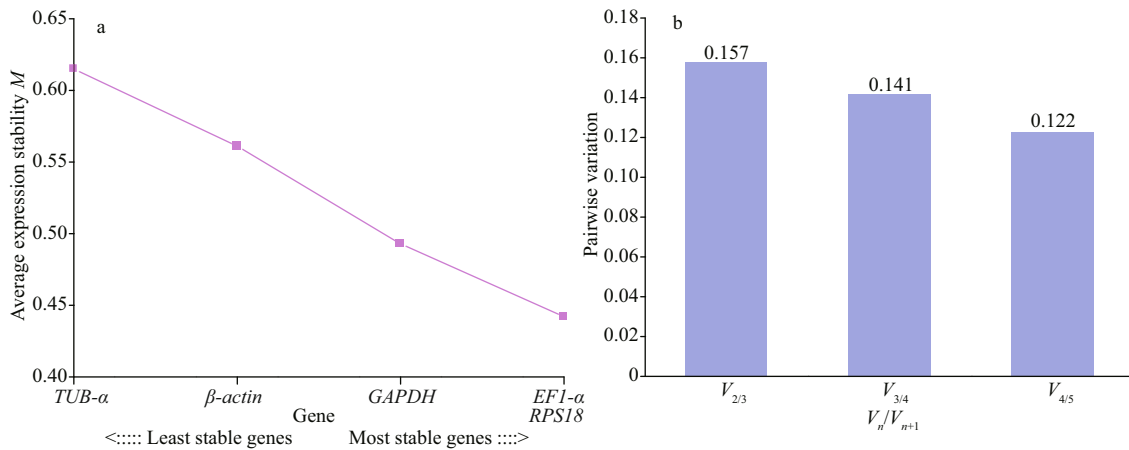
Sample type	Gene	geNorm		NormFinder		BestKeeper		
		$M$ value	Rank	Stability value	Rank	Coeff. of corr ( $r$ )	$P$ value	Rank
Different tissues	<i>β-actin</i>	1.606	4	0.699	2	0.927	0.001	2
	<i>EF1-α</i>	1.249	1	0.978	4	0.896	0.001	5
	<i>GAPDH</i>	1.249	1	0.866	3	0.924	0.001	3
	<i>RPS18</i>	1.568	3	0.574	1	0.951	0.001	1
	<i>TUB-α</i>	2.155	5	1.916	5	0.908	0.001	4
Salinity stress	<i>β-actin</i>	0.561	4	0.310	3	0.904	0.001	3
	<i>EF1-α</i>	0.442	1	0.257	2	0.949	0.001	1
	<i>GAPDH</i>	0.493	3	0.194	1	0.941	0.001	2
	<i>RPS18</i>	0.442	1	0.332	4	0.890	0.001	4
	<i>TUB-α</i>	0.615	5	0.400	5	0.878	0.001	5
<i>Vibrio</i> challenge	<i>β-actin</i>	0.837	4	0.829	4	0.844	0.001	5
	<i>EF1-α</i>	0.403	1	0.092	1	0.980	0.001	2
	<i>GAPDH</i>	0.539	3	0.287	3	0.977	0.001	3
	<i>RPS18</i>	0.403	1	0.140	2	0.982	0.001	1
	<i>TUB-α</i>	1.079	5	0.924	5	0.932	0.001	4

poorly ranked by all programs, ranked fifth by NormFinder and BestKeeper, and ranked fourth by BestKeeper. *EF1-α* was highly ranked by geNorm but poorly ranked by NormFinder and BestKeeper (Table 3). As a result, *GAPDH*, *RPS18*, and *β-actin* were identified as the optimal gene combination for qRT-PCR analysis in different tissues of clam *C. sinensis* under normal physiological conditions.

### 3.4 Expression stability analysis of candidate genes under low salinity stress

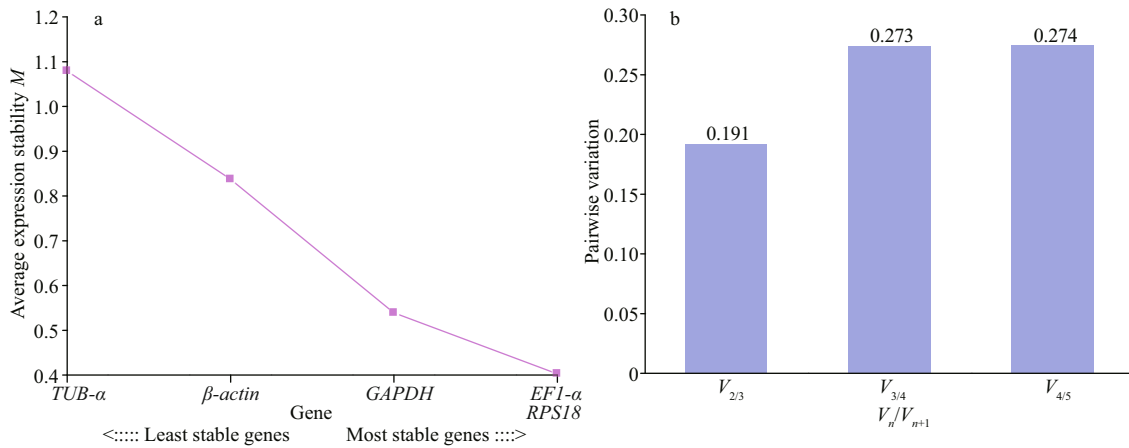
Salinity stress is one of the environmental stresses often suffered by *C. sinensis* in tidal flats. To find out a suitable reference genes combination,

the expression stability of the candidate reference genes was assessed in hepatopancreas under low salinity stress. As shown in Fig.4a, geNorm analysis indicated that *EF1-α* and *RPS18* were the most stable genes, which were followed by *GAPDH*, *β-actin*, and *TUB-α*. The NormFinder recommended *GAPDH* as the most stable reference gene, the ranking of the other four candidate genes were *EF1-α*, *β-actin*, *RPS18*, and *TUB-α*. The BestKeeper identified *EF1-α* as the most stable gene, which was followed by *GAPDH*, *β-actin*, *RPS18*, and *TUB-α* (Table 3). For the number of reference genes required for data normalization, geNorm analysis showed that when the reference genes number increased from two to



**Fig.4 Expression stability of candidate reference genes analyzed by geNorm in the hepatopancreas of *C. sinensis* under low salinity stress**

a. the *M* values of candidate reference genes; b the optimal number of reference genes is required for normalization.



**Fig.5 Expression stability of candidate reference genes analyzed by geNorm in the gills of *C. sinensis* after *Vibrio* challenge**

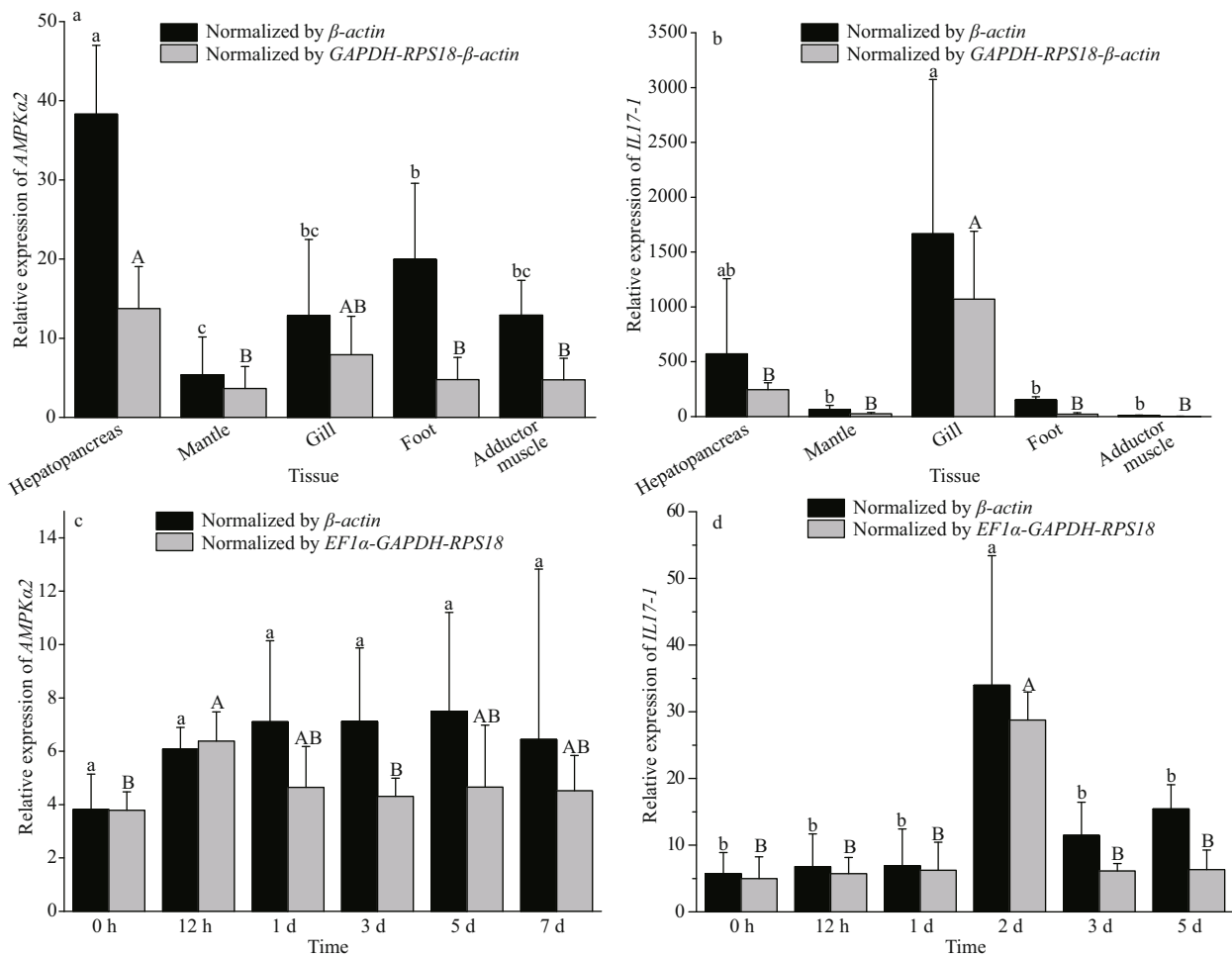
a. the *M* values of candidate reference genes; b the optimal number of reference genes is required for normalization.

four, the  $V_{n/n+1}$  value decreased from 0.157 to 0.122. The  $V_{3/4}$  value (0.141) was less than the cut-off value of 0.15, thus the most three stable genes were required for accurate normalization (Fig.4b). According to the comprehensive rank of these five candidate reference genes, *EF1-α*, *GAPDH*, and *RPS18* could be recommended as the reference gene combination in the qRT-PCR analysis of clam *C. sinensis* under salinity stress.

**3.5 Expression stability analysis of candidate genes after *Vibrio* challenge**

*Vibrio* infection is one of the main biotic stresses suffered by *C. sinensis*. To find out appropriate reference genes, we tested five candidate housekeeping genes for expression stability in the gill of *Vibrio* infected clams. As shown in Fig.5a, geNorm analysis showed that *EF1-α* and *RPS18* were the most stable genes. Meanwhile, both NormFinder and BestKeeper showed that *EF1-α* and *RPS18* were the

most stable genes, although the ranking was different. For *GAPDH*, the ranking orders generated by the three programs were the same, ranking third. Both geNorm and NormFinder showed that *TUB-α* was the least stable gene, while the BestKeeper recommended *β-actin* as the least (Table 3). All the three  $V_n/V_{n+1}$  values for the five candidate genes generated by geNorm were higher than the cut-off value of 0.15, shown that no optimal gene combination could be provided by geNorm (Fig.5b). Here the  $V_{2/3}$  value (0.191) was the lowest and a little bit higher than 0.15, indicating that using two stable genes could be preferred. Still, considering that the two most stable genes had the same function and were both involved in the translation process, we included *GAPDH* in the gene combination for normalization to avoid errors caused by gene co-regulation. Thus *EF1-α*, *RPS18*, and *GAPDH* could be recommended as the reference genes combination in the qRT-PCR analysis of clam *C. sinensis* after *Vibrio* challenge.



**Fig.6 Comparison of different normalization strategies**

a. the tissue expression distribution of *AMPKα2*; b. the tissue expression distribution of *IL17-1*; c. the expression patterns of *AMPKα2* in the hepatopancreas of *C. sinensis* under salinity stress at different time points; d. the expression patterns of *IL17-1* in the gill of *C. sinensis* after *Vibrio* challenge at different time points. These gene expression data were expressed as the mean  $\pm$  SD, the statistical results with the two normalization strategies were represented by capital and lowercase letters respectively. Values with different letters differ significantly, significance was set at  $P < 0.05$ , as determined by one-way ANOVA followed by Duncan's test.

### 3.6 Comparison of different normalization strategies

To test whether different normalization strategies might influence the quantification results of gene expression, we analyzed the expression patterns of *AMPKα2* and *IL17-1* genes in different tissues under normal physiological conditions, salinity stress, and *Vibrio* infection of *C. sinensis* respectively. The results of the above reference genes validation indicated that the gene combination of three housekeeping genes should be used for normalization. Therefore, we used either  $\beta$ -actin or the gene combination *GAPDH-RPS18-β-actin* for normalization in different tissues, another gene combination *EF1α-GAPDH-RPS18* for normalization in specific tissue under salinity and *Vibrio* stress. The results showed that the expression patterns of target genes mostly varied using different normalization strategies.

For *AMPKα2*, the expression patterns obtained from the two normalization strategies were various. In detail, the expression analysis normalized by  $\beta$ -actin showed that the expression of *AMPKα2* in hepatopancreas was significantly higher than that in the other four tissues. On the contrary, the expression analysis normalized by gene combination showed that *AMPKα2* also had relatively high expression in the gill, and no significant difference was found between hepatopancreas and gill (Fig.6a). Then, the expression pattern of *AMPKα2* was further analyzed in the hepatopancreas of clams under salinity stress. As shown in Fig.6c, the expression of *AMPKα2* was significantly up-regulated at 12 h, and then gradually recuperated to the basal level. However, no significant change was observed in the expression of the *AMPKα2* gene normalized by a single  $\beta$ -actin gene.

For *IL17-1*, the expression patterns of *IL17-1* in



different tissues obtained from the two normalization strategies were slightly different. The result normalized by the gene combination *GAPDH-RPS18-β-actin* showed that the expression level of *IL17-1* in gill was the highest in comparison with the other tissues. However, the expression analysis normalized by *β-actin* showed that *IL17-1* also had relatively high expression in the hepatopancreas, and no significant difference was found between the hepatopancreas and gill (Fig.6b). The expression patterns of *IL17-1* in clams post-*Vibrio* infection obtained from the two normalization strategies were similar, expression significantly increased at 2 d post-treatment (Fig.6d). Besides, it should be noted that most of the standard deviation of normalization by a single reference was larger than that of normalization by multiple genes. Hence, we recommend that multiple reference genes should be used for the normalization of gene expression analysis.

#### 4 DISCUSSION

qRT-PCR is one of the most widely used techniques for gene expression analysis in biological research. In relative quantitation, the stable reference gene is a prerequisite for accurate normalization of gene expression levels (Guénin et al., 2009). Previous studies have shown that there is no ideal reference gene that is appropriate for use in all experimental conditions (Vandesompele et al., 2002; Chapman and Waldenström, 2015). As a consequence, it is necessary to verify the expression stability of reference genes under specific experimental conditions. To date, numerous researches have been conducted on reference gene selection in aquatic animals, such as fishes (Li et al., 2020; Dharmaratnam et al., 2021; Szczygieł et al., 2021), crustaceans (Hu et al., 2018; Zhou et al., 2018), and mollusks (Huan et al., 2016; Volland et al., 2017; Zhao et al., 2018), and research on reference gene selection in clam *C. sinensis* is vacant. For the economic and ecological importance, the first genome sequencing for *C. sinensis* was performed in 2020, and 27 344 protein-coding genes were annotated, providing massive gene resources for gene expression studies and systematic reference gene validation in this clam (Wei et al., 2020). In the present study, five commonly used reference genes were selected from the genome and transcriptome datasets of *C. sinensis*, and their expression stabilities were evaluated. We revealed that the expression stabilities of these housekeeping genes were not consistent in different tissues and experimental treatments, and hence highlighted the need to verify the expression

stability of reference genes for normalization.

There is no consensus on how many control genes were required for reliable normalization. In the past, the vast majority of studies used a single reference gene for normalization. As exact normalization is such a vital component of gene expression analysis, it has been strongly advocated that at least two validated reference genes be employed for normalization (Chapman and Waldenström, 2015). In this study, we selected the algorithm of geNorm to define the optimal number of reference genes required for accurate normalization in qRT-PCR. Among the five different tissues, the result showed that all  $V_{n/n+1}$  values were higher than 0.15. This suggested that all the five candidate reference genes were not sufficient for accurate normalization according to the cut-off value of 0.15 proposed by Vandesompele et al. (2002). Similar to our results, all  $V_{n/n+1}$  values over 0.15 were also observed in different tissues of other species, such as the prawn *Macrobrachium olfersii* (Jaramillo et al., 2017) and the octopus *Octopus minor* (Whang et al., 2020). This may be different tissues have different cell types and perform different functions, so various housekeeping genes showed variation in expression. Thus, it is not surprising that all  $V_{n/n+1}$  values exceeded 0.15. The number of reference genes used for normalization is a trade-off between practical considerations and accuracy. When only a few target genes need to be studied or minimal amounts of cDNA are obtained, it is impractical to use many reference genes for normalization. Literature showed that, in most cases, three proper control genes are sufficient for a valid normalization strategy (Vandesompele et al., 2002). We noticed that the lowest  $V_{n/n+1}$  value of different adult tissues was  $V_{3/4}$  (0.345), indicating that using three selected control genes (*GAPDH-RPS18-β-actin*) for normalization could be acceptable.

The optimal reference genes should not be regulated through common upstream effectors or should not directly regulate each other, that is, co-regulated genes should not be used as reference genes (Kadegowda et al., 2009). This criterion is particularly important when the multiple reference genes were used for normalization. When evaluating gene expression stability, co-regulated genes would lead to distorted results (Huan et al., 2016). In this study, the five selected housekeeping genes were involved in three functional classes, including cytoskeletal protein (*β-actin*, *TUB-α*), translation (*EFl-α*, *RPS18*), and glycolysis (*GAPDH*). The geNorm analysis showed that the  $V_{2/3}$  value (0.191) was the lowest, indicating

that using two stable genes for normalization could be accepted in clams after *Vibrio* infection. Nevertheless, the top two genes ranked by all the three programs (*EF1- $\alpha$* , *RPS18*) belong to the same functional class, and they might be co-regulated. In oyster *C. gigas*, Similarly, *EF1- $\alpha$*  and *RPS18* were found to be stable during the development of oyster *C. gigas*, and these two reference genes were unlikely to be co-regulated (Huan et al., 2016). Even so, to be cautious, we also included *GAPDH* in the gene combination (*EF1 $\alpha$ -GAPDH-RPS18*) for normalization in clam's post-*Vibrio* challenge. Consequently, if it is not determined to exclude the co-regulated genes, the selected multiple reference genes should belong to at least two functional categories to avoid errors caused by gene co-regulation.

Among these selected housekeeping genes in this study,  *$\beta$ -actin* is the most commonly used reference gene in aquatic organisms. Although some studies have shown that the expression level of  *$\beta$ -actin* is stable and could be used as an internal reference gene (Hu et al., 2018; Geng et al., 2019; Li et al., 2020), an increasing number of studies have shown that  *$\beta$ -actin* is a less stable housekeeping gene (Morga et al., 2010; Feng et al., 2013; Zhou et al., 2018). In this study,  *$\beta$ -actin* was highly ranked by NormFinder and BestKeeper among different adult tissues but poorly ranked in clams under salinity stress and *Vibrio* infection. Hence, it is likely that  *$\beta$ -actin* is stably expressed in normal tissues and has varied expression under experimental stress. In contrast to  *$\beta$ -actin*, *EF1- $\alpha$*  was poorly ranked by NormFinder and BestKeeper in different tissues but highly ranked in clams under salinity stress and *Vibrio* infection. *EF1- $\alpha$*  encodes the  $\alpha$  subunit of eukaryotic elongation factor 1 and has been commonly used to study gene expression levels in bivalves (Morga et al., 2010; Moreira et al., 2014; Huan et al., 2016; Yan et al., 2017). Similar to our results, *EF1- $\alpha$*  was a less stable reference gene when analyzing adult tissues of *C. gigas* (Dheilly et al., 2011). *RPS18* takes part in ribosome biogenesis in all cell types and was used as a reference gene for normalization in bivalve species (Mateo et al., 2010; Du et al., 2013). Our study validated that *RPS18* could be selected as a control gene to study target gene expression in clams. *GAPDH* encodes a key enzyme in glucose metabolism and is a frequently used reference gene for normalization. We found the expression of *GAPDH* was relatively stable in our test and could be used as a reference gene for normalization in clams. Likewise, *GAPDH* and

*EF1- $\alpha$*  have been combined as reference genes when studying expression levels in hemocytes of flat oyster *Ostrea edulis* (Morga et al., 2010). Taken together, the gene combinations *GAPDH-RPS18- $\beta$ -actin* and *EF1 $\alpha$ -GAPDH-RPS18* were recommended to use for normalization in future studies for normal tissues and experimental stress, respectively.

To test the effect of reference gene choice on gene expression analysis, we analyzed the expression patterns of *AMPK $\alpha$ 2* and *IL17-1* genes using two normalization strategies. *AMPK $\alpha$ 2* is a highly conserved protein kinase that plays an important role in cellular stress response. Several studies have shown that salinity stress can cause significant changes in *AMPK $\alpha$ 2* expression level (Xu et al., 2016; Zeng et al., 2016; Xiao et al., 2018). In the present study, the expression patterns in clams under salinity stress obtained from the two normalization strategies were different. The significant change was observed only in the expression of *AMPK $\alpha$ 2* normalized by reference gene combination. *IL17-1* is a proinflammatory cytokine that plays a key role in innate immune responses. In the Pacific oyster *C. gigas*, *IL17-1* was highly expressed in the gill tissues, and the expression of the *IL17-1* was significantly up-regulated post pathogen challenge (Li et al., 2014). Similar results were also found in this study, although the expression patterns of *IL17-1* in different tissues obtained from the two normalization strategies were slightly different. As a result, this study indicated that the normalization strategy could influence quantification results, especially when it comes to which genes are subtle changes in expression. Combined with the results of gene stability analysis, we recommend that the geometric mean of three reference genes be used for normalization in future gene expression study of *C. sinensis*.

## 5 CONCLUSION

Using three stable reference genes with different functional classes was most appropriate for the normalization strategy. We proposed to use the gene combination *GAPDH-RPS18- $\beta$ -actin* for normalization among normal adult tissues, and another gene combination *EF1 $\alpha$ -GAPDH-RPS18* for normalization under salinity and *Vibrio* stress. This is the first study that provides the validated reference gene combinations for accurate normalization for specific experimental conditions of *C. sinensis*, which is the key to functional studies of target genes in this marine bivalve in the future.

## 6 DATA AVAILABILITY STATEMENT

All data generated and/or analyzed during this study are included in this published article and its supplementary information files.

### References

- Andersen C L, Jensen J L, Ørntoft T F. 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**(15): 5245-5250, <https://doi.org/10.1158/0008-5472.CAN-04-0496>.
- Chapman J R, Waldenström J. 2015. With reference to reference genes: a systematic review of endogenous controls in gene expression studies. *PLoS One*, **10**(11): e0141853, <https://doi.org/10.1371/journal.pone.0141853>.
- Chen X J, Zhang X Q, Sun Y et al. 2020. Assessment of internal controls for data normalization of gene expression after different bacterial stimulation by quantitative real-time PCR in golden pompano *Trachinotus blochii*. *Journal of Oceanology and Limnology*, **38**(2): 480-489, <https://doi.org/10.1007/s00343-019-9053-5>.
- Dharmaratnam A, Sudhagar A, Nithianantham S R et al. 2021. Evaluation of candidate reference genes for quantitative RTqPCR analysis in goldfish (*Carassius auratus* L.) in healthy and CyHV-2 infected fish. *Veterinary Immunology and Immunopathology*, **237**: 110270, <https://doi.org/10.1016/j.vetimm.2021.110270>.
- Dheilly N M, Lelong C, Huvet A et al. 2011. Development of a Pacific oyster (*Crassostrea gigas*) 31, 918-feature microarray: identification of reference genes and tissue-enriched expression patterns. *BMC Genomics*, **12**(1): 468, <https://doi.org/10.1186/1471-2164-12-468>.
- Du Y S, Zhang L L, Xu F et al. 2013. Validation of housekeeping genes as internal controls for studying gene expression during Pacific oyster (*Crassostrea gigas*) development by quantitative real-time PCR. *Fish & Shellfish Immunology*, **34**(3): 939-945, <https://doi.org/10.1016/j.fsi.2012.12.007>.
- Feng L Y, Yu Q, Li X et al. 2013. Identification of reference genes for qRT-PCR analysis in Yesso scallop *Patinoptecten yessoensis*. *PLoS One*, **8**(9): e75609, <https://doi.org/10.1371/journal.pone.0075609>.
- Geng W Y, Yao F J, Tang T et al. 2019. Evaluation of the expression stability of  $\beta$ -actin under bacterial infection in *Macrobrachium nipponense*. *Molecular Biology Reports*, **46**(1): 309-315, <https://doi.org/10.1007/s11033-018-4473-4>.
- Guénin S, Mauriat M, Pelloux J et al. 2009. Normalization of qRT-PCR data: the necessity of adopting a systematic, experimental conditions-specific, validation of references. *Journal of Experimental Botany*, **60**(2): 487-493, <https://doi.org/10.1093/jxb/ern305>.
- Hu Y N, Fu H T, Qiao H et al. 2018. Validation and evaluation of reference genes for quantitative real-time PCR in *Macrobrachium nipponense*. *International Journal of Molecular Sciences*, **19**(8): 2258, <https://doi.org/10.3390/ijms19082258>.
- Huan P, Wang H X, Liu B Z. 2016. Assessment of housekeeping genes as internal references in quantitative expression analysis during early development of oyster. *Genes & Genetic Systems*, **91**(5): 257-265, <https://doi.org/10.1266/ggs.16-00007>.
- Jaramillo M L, Ammar D, Quispe R L et al. 2017. Identification and evaluation of reference genes for expression studies by RT-qPCR during embryonic development of the emerging model organism, *Macrobrachium olfersii*. *Gene*, **598**: 97-106, <https://doi.org/10.1016/j.gene.2016.11.001>.
- Jiang F J, Wang H X, Yue X et al. 2018. Integrating the *Vibrio*-resistance phenotype and gene expression data for discovery of markers used for resistance evaluation in the clam *Meretrix petechialis*. *Aquaculture*, **482**: 130-136, <https://doi.org/10.1016/j.aquaculture.2017.09.033>.
- Jin C F, Song W H, Wang M Y et al. 2021. Transcriptome-wide identification and validation of reference genes in black rockfish (*Sebastes schlegelii*). *Journal of Ocean University of China*, **20**(3): 654-660, <https://doi.org/10.1007/s11802-021-4588-4>.
- Kadegowda A K G, Bionaz M, Thering B et al. 2009. Identification of internal control genes for quantitative polymerase chain reaction in mammary tissue of lactating cows receiving lipid supplements. *Journal of Dairy Science*, **92**(5): 2007-2019, <https://doi.org/10.3168/jds.2008-1655>.
- Li J, Zhang Y, Zhang Y H et al. 2014. Genomic characterization and expression analysis of five novel IL-17 genes in the Pacific oyster, *Crassostrea gigas*. *Fish & Shellfish Immunology*, **40**(2): 455-465, <https://doi.org/10.1016/j.fsi.2014.07.026>.
- Li Y K, Han J B, Wu J Y et al. 2020. Transcriptome-based evaluation and validation of suitable housekeeping gene for quantification real-time PCR under specific experiment condition in teleost fishes. *Fish & Shellfish Immunology*, **98**: 218-223, <https://doi.org/10.1016/j.fsi.2020.01.018>.
- Lin T T, Lai Q F, Yao Z L et al. 2013. Combined effects of carbonate alkalinity and pH on survival, growth and haemocyte parameters of the Venus clam *Cyclina sinensis*. *Fish & Shellfish Immunology*, **35**(2): 525-531, <https://doi.org/10.1016/j.fsi.2013.05.006>.
- Liu W S, Ma Y H, Hu S Y et al. 2002. Rearing Venus clam seeds, *Cyclina sinensis* (Gmelin), on a commercial scale. *Aquaculture*, **211**(1-4): 109-114, [https://doi.org/10.1016/S0044-8486\(01\)00859-6](https://doi.org/10.1016/S0044-8486(01)00859-6).
- Livak K J, Schmittgen T D. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the  $2^{-\Delta\Delta C_t}$  method. *Methods*, **25**(4): 402-408, <https://doi.org/10.1006/meth.2001.1262>.
- Mateo D R, Greenwood S J, Araya M T et al. 2010. Differential gene expression of  $\gamma$ -actin, Toll-like receptor 2 (TLR-2) and interleukin-1 receptor-associated kinase 4 (IRAK-4) in *Mya arenaria* haemocytes induced by *in vivo* infections with two *Vibrio splendidus* strains. *Developmental & Comparative Immunology*, **34**(7): 710-714, <https://doi.org/10.1016/j.dci.2010.02.006>.
- Mauriz O, Maneiro V, Pérez-Parallé M L et al. 2012. Selection of reference genes for quantitative RT-PCR studies on

- the gonad of the bivalve mollusc *Pecten maximus* L. *Aquaculture*, **370-371**: 158-165, <https://doi.org/10.1016/j.aquaculture.2012.10.020>.
- Moreira R, Pereiro P, Costa M M et al. 2014. Evaluation of reference genes of *Mytilus galloprovincialis* and *Ruditapes philippinarum* infected with three bacteria strains for gene expression analysis. *Aquatic Living Resources*, **27**(3-4): 147-152, <https://doi.org/10.1051/alr/2014015>.
- Morga B, Arzul I, Faury N et al. 2010. Identification of genes from flat oyster *Ostrea edulis* as suitable housekeeping genes for quantitative real time PCR. *Fish & Shellfish Immunology*, **29**(6): 937-945, <https://doi.org/10.1016/j.fsi.2010.07.028>.
- Ni Q, Li W Q, Jia X W et al. 2020. Effect of salinity on growth performance and resistance of the clam *Cyclina sinensis* against *Vibrio parahaemolyticus* infection. *The Israeli Journal of Aquaculture-Bamidgeh*, **73**: 1124924, <https://doi.org/10.46989/001c.21693>.
- Ni Q, Li W Q, Liang X F et al. 2021. Gill transcriptome analysis reveals the molecular response to the acute low-salinity stress in *Cyclina sinensis*. *Aquaculture Reports*, **19**: 100564, <https://doi.org/10.1016/j.aqrep.2020.100564>.
- Pfaffl M W, Tichopad A, Prgomet C et al. 2004. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: bestkeeper-Excel-based tool using pair-wise correlations. *Biotechnology Letters*, **26**(6): 509-515, <https://doi.org/10.1023/b:bile.0000019559.84305.47>.
- Ren Y P, Zhang H, Pan B P et al. 2015. A Kazal-type serine proteinase inhibitor from *Cyclina sinensis* is involved in immune response and signal pathway initiation. *Fish & Shellfish Immunology*, **47**(1): 110-116, <https://doi.org/10.1016/j.fsi.2015.08.026>.
- Stephenson F H. 2010. The real-time polymerase chain reaction (RT-PCR). In: Stephenson F H ed. *Calculations for Molecular Biology and Biotechnology: A Guide to Mathematics in the Laboratory*. 2<sup>nd</sup> edn. Academic Press, Elsevier. p.211-311, <https://doi.org/10.1016/b978-0-12-375690-9.00009-7>.
- Sun Z Y, Sun W W, Pan B P et al. 2021. Molecular characterization of a novel p38 MAPK cDNA from *Cyclina sinensis* and its potential immune-related function under the threat of *Vibrio anguillarum*. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, **255**: 110599, <https://doi.org/10.1016/j.cbpb.2021.110599>.
- Szczygieł J, Kamińska-Gibas T, Petit J et al. 2021. Re-evaluation of common carp (*Cyprinus carpio* L.) housekeeping genes for gene expression studies-considering duplicated genes. *Fish & Shellfish Immunology*, **115**: 58-69, <https://doi.org/10.1016/j.fsi.2021.05.013>.
- Vandesompele J, De Preter K, Pattyn F et al. 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology*, **3**(7): research0034.1, <https://doi.org/10.1186/gb-2002-3-7-research0034>.
- Volland M, Blasco J, Hampel M. 2017. Validation of reference genes for RT-qPCR in marine bivalve ecotoxicology: systematic review and case study using copper treated primary *Ruditapes philippinarum* hemocytes. *Aquatic Toxicology*, **185**: 86-94, <https://doi.org/10.1016/j.aquatox.2017.01.003>.
- Wei M, Ge H X, Shao C W et al. 2020. Chromosome-level clam genome helps elucidate the molecular basis of adaptation to a buried lifestyle. *iScience*, **23**(6): 101148, <https://doi.org/10.1016/j.isci.2020.101148>.
- Whang I, Kang H S, Kim Y. 2020. Validation of reference genes for quantitative gene expression studies in *Octopus minor*. *Ocean Science Journal*, **55**(1): 183-191, <https://doi.org/10.1007/s12601-020-0007-9>.
- Xiao S, Wong N K, Li J et al. 2018. Analysis of *in situ* transcriptomes reveals divergent adaptive response to hyper- and hypo-salinity in the Hong Kong oyster, *Crassostrea hongkongensis*. *Frontiers in Physiology*, **9**: 1491, <https://doi.org/10.3389/fphys.2018.01491>.
- Xu C, Li E C, Xu Z X et al. 2016. Molecular characterization and expression of AMP-activated protein kinase in response to low-salinity stress in the Pacific white shrimp *Litopenaeus vannamei*. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, **198**: 79-90, <https://doi.org/10.1016/j.cbpb.2016.04.005>.
- Yan L L, Su J Q, Wang Z P et al. 2017. Selection of reference genes for expression analysis of Kumamoto and Portuguese oysters and their hybrid. *Journal of Ocean University of China*, **16**(6): 1139-1147, <https://doi.org/10.1007/s11802-017-3339-z>.
- Yue X, Liu B Z, Xiang J H, et al. 2010. Identification and characterization of the pathogenic effect of a *Vibrio parahaemolyticus*-related bacterium isolated from clam *Meretrix meretrix* with mass mortality. *Journal of Invertebrate Pathology*, **103**(2): 109-115, <https://doi.org/10.1016/j.jip.2009.11.008>.
- Zeng L, Liu B, Wu C W et al. 2016. Molecular characterization and expression analysis of AMPK  $\alpha$  subunit isoform genes from *Scophthalmus maximus* responding to salinity stress. *Fish Physiology and Biochemistry*, **42**(6): 1595-1607, <https://doi.org/10.1007/s10695-016-0243-1>.
- Zhao X L, Fu J P, Jiang L T et al. 2018. Transcriptome-based identification of the optimal reference genes as internal controls for quantitative RT-PCR in razor clam (*Sinonovacula constricta*). *Genes & Genomics*, **40**(6): 603-613, <https://doi.org/10.1007/s13258-018-0661-9>.
- Zhou S M, Tao Z, Shen C et al. 2018.  $\beta$ -actin gene expression is variable among individuals and not suitable for normalizing mRNA levels in *Portunus trituberculatus*. *Fish & Shellfish Immunology*, **81**: 338-342, <https://doi.org/10.1016/j.fsi.2018.07.021>.

### Electronic supplementary material

Supplementary material (Supplementary File 1 and Supplementary Tables S1–S3) is available in the online version of this article at <https://doi.org/10.1007/s00343-021-1335-z>.