



OPEN

## Screening of reference genes for microRNA analysis in the study of soldier caste differentiation of Formosan subterranean termite *Coptotermes formosanus* Shiraki

He Du<sup>1</sup>, Runmei Huang<sup>1</sup>, Dasong Chen<sup>1</sup>, Chaofu Huang<sup>2</sup>, Huan Zhang<sup>3</sup> & Zhiqiang Lia<sup>1</sup>✉

The soldier caste differentiation is a complex process that is governed by the transcriptional regulation and post-transcriptional regulation. microRNAs (miRNAs) are noncoding RNAs that control a wide range of activities. However, their roles in soldier caste differentiation are barely studied. RT-qPCR is a powerful tool to study the function of genes. A reference gene is required for normalization for the relative quantification method. However, no reference gene is available for miRNA quantification in the study of soldier caste differentiation of *Coptotermes formosanus* Shiraki. In this research, in order to screen the suitable reference genes for the study of the roles of miRNAs in soldier caste differentiation, the expression levels of 8 candidate miRNA genes were quantified in the head and thorax + abdomen during soldier differentiation. The qPCR data were analyzed using geNorm, NormFinder, BestKeeper,  $\Delta C_t$  method and RefFinder. The normalization effect of the reference genes was evaluated using the *let-7-3p*. Our study showed that *novel-m0649-3p* was the most stable reference gene, while *U6* was the least stable reference gene. Our study has selected the most stable reference gene, and has paved the way for functional analysis of miRNAs in soldier caste differentiation.

Termites are eusocial insects with division of labor as one of their distinct characteristics. This feature of dividing labor among colony members allows termites to form a rather complex social system, and enables them to improve the efficiency of operating the termite colony, thus facilitating termites' evolutionary and ecological success<sup>1</sup>. The termite society typically consists of different castes, such as reproductives, workers and soldiers. In particular, the soldier caste is the defensive caste inside the society<sup>2</sup>. Normally, the colony could maintain a relative stable ratio of soldier caste<sup>3</sup>, but when the soldier ratio is below the regular ratio for the species, new soldier could form<sup>4</sup>. The formation of soldier is induced by high juvenile hormone (JH) titer in the hemolymph<sup>5-7</sup>. Juvenile hormone analog (JHA) could also artificially induce formation of soldiers. JH drives the differential expression of the genes, and triggers a complex of gene cascade during the soldier morphogenesis. For example, hexamerin is synthesized massively during the process and identified as the key regulation protein to buffer the function of JH<sup>8</sup>. In addition, the classical signaling pathways, such as JH signaling pathway, TGF $\beta$  signaling pathway, and insulin signaling pathway, are involved in soldier formation regulation<sup>9-11</sup>. However, the mechanism of soldier caste differentiation is mainly focused at the transcriptional level, the roles of microRNAs (miRNAs) are barely studied.

miRNAs are noncoding RNAs with the length of ~21–24 nt. They play important roles in RNA silencing and the post-transcriptional regulation through repression of gene translation or degradation of mRNA transcripts<sup>12</sup>. miRNAs regulate behavior, development, reproduction and immunity in insects<sup>13-15</sup>. In particular, miRNAs mediate the social immunity of termites to pathogens<sup>16</sup>. The caste differentiation of termites involves a dynamic change of gene cascade. At a certain stage of the transformation, the high expressed genes need to be cleaned up

<sup>1</sup>Guangdong Key Laboratory of Integrated Pest Management in Agriculture, Guangdong Public Laboratory of Wild Animal Conservation and Utilization, Institute of Zoology, Guangdong Academy of Sciences, Guangzhou 510260, China. <sup>2</sup>Nanning Institute of Termite Control, Nanning 530023, China. <sup>3</sup>State Key Laboratory of Integrated Management of Pest Insects and Rodents, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China. ✉email: lizhiqiang61@163.com

or maintained at relative low expression. Considering the significance of miRNAs in a wide range of physiological activities, they may serve functions in soldier caste differentiation. Up to date, only two pieces of research on the roles of miRNAs in caste differentiation in termites were completed. Itakura et al.<sup>17</sup> sequenced the miRNAs in worker caste of *Coptotermes formosanus* Shiraki and *Reticulitermes speratus* (Kolbe). Matsunami et al.<sup>18</sup> reported the differential expression of miRNAs in the workers and soldiers of *R. speratus*, indicating the possible involvement of miRNAs in caste differentiation in termites. However, the roles of miRNAs in soldier differentiation are not quite clear, and have to be confirmed with gene expression analysis.

RT-qPCR is a standard technique to quantify gene expression level, either using the absolute quantification or the relative quantification method. The latter needs a reference gene to normalize the qPCR data. Unproper choice of reference gene has a negative effect on data analysis, such as increasing the error or even not detecting of gene expression pattern<sup>19–21</sup>. Although the housekeeping genes, such as 18S,  $\beta$ -actin, 5S rRNA and U6 snRNA, have been widely used for qPCR data normalization, their expression levels vary considerably in certain experimental conditions<sup>22–25</sup>. Since choice of reference genes depends on the conditions of the experiment, there are no universal reference genes for all the qPCR data analysis. The ideal way to choose reference gene is to screen the reference genes for the specific experiment. Hence, many studies have been conducted to screen the suitable reference genes for different species and under various experimental conditions<sup>26–34</sup>. However, there is no reference gene available for miRNA quantification in the study of soldier caste differentiation of *C. formosanus*.

We have sequenced the miRNAs of *C. formosanus*, and through checking the miRNome data, we identified 8 candidate reference genes which had the middle expression level and were least influenced by the stimulation of a JHA, methoprene. In order to test the suitability of the reference genes, the expression profile of a conserved miRNA, *let-7*, was quantified. The objective of this project is to screen the suitable reference gene for miRNA quantification in the study of the soldier caste differentiation.

## Materials and methods

**Bioassay and sample preparation.** Termites were collected from Shangchong Fruit Park, South China University of Technology, Guangdong Tree Park, and Dafushan Forest Park, Guangzhou, China. Termites were kept in the laboratory at room temperature and used within two weeks.

In order to induce soldier formation, workers were treated with methoprene<sup>35</sup>. Acetone solution of methoprene was dripped evenly onto filter paper to make the concentration of methoprene as 1000 ppm. Two pieces of filter paper (32 mm in diameter) were placed in a Petri dish (35 mm in diameter). The paired filter paper was then wet by adding 240  $\mu$ L distilled water. Twenty workers were added to each Petri dish. The Petri dishes were placed in a sealed container. A piece of wet paper towel was placed in the container to maintain the moisture level inside. The experimental set was placed in an incubator at 28 °C. Four colonies were used.

Termites were collected after feeding on methoprene for the following days: 1st day, 5th day, 10th day, and 15th day. Twenty workers were also collected just before the methoprene bioassay (which was designated as 0D). So, there were five samples for each colony (0D, 1D, 5D, 10D, 15D). The heads of the termites were cut with a scalpel. The head and thorax + abdomen were put in 100  $\mu$ L of QIAzol Lysis Reagent (Qiagen, Hilden, Germany) separately. The tubes containing the termite were immersed into liquid nitrogen immediately after collection, and then stored at –80 °C. The head is the place where the JH is synthesized. The thorax + abdomen is where the fat body is located, as the fat body responses most obviously after JHA treatment<sup>44</sup>. So, we split the termites into two different body parts, and screened the reference genes separately.

**miRNA extraction and quantification.** miRNA was extracted using the RNeasy Mini Kit (Qiagen, Hilden, Germany) following the manufacturer's protocol. The concentration of the RNA was measured with the TGen Spectrophotometer Plus (Tiangen, Beijing, China). The integrity of the RNA was confirmed using agarose gel electrophoresis. RNA was reverse-transcribed with stem-loop RT primers using M-MuLV Reverse Transcriptase (NEB, Ipswich, Massachusetts, United States). The stem-loop RT primers were designed according to Chen et al.<sup>36</sup> (Table 1). The 10  $\mu$ L transcription system contained the following ingredients: 2.5  $\mu$ L M-MuLV Reverse Transcriptase (200 U/ $\mu$ L), 1  $\mu$ L M-MuLV buffer, 0.7 mM dNTP, 30 nM stem-loop RT primer, 2 U RNasin<sup>®</sup> Ribonuclease Inhibitor (Promega, Madison, Wisconsin, United States), and 1  $\mu$ g RNA. Probe-based method was used to conduct the qPCR. The probes were synthesized with 5'-FAM and 3'-MGB modifications (TsingKe, Beijing, China) (Table 1). qPCR was conducted using Luna<sup>®</sup> Universal Probe qPCR Master Mix (NEB, Ipswich, Massachusetts, United States). The 10  $\mu$ L qPCR system was as follows: 5  $\mu$ L Luna universal probe qPCR master mix, 1.5  $\mu$ M forward primer, 0.7  $\mu$ M reverse primer, 0.2  $\mu$ M probe, and 0.7  $\mu$ L cDNA. The qPCR was conducted on BIO-RAD CFX Connect (Bio-Rad, California, USA). The program for amplification: denaturation at 95°C for 1 min, with 40 cycles of denaturation at 95°C 15 s, annealing/extension at 60°C for 30 s.

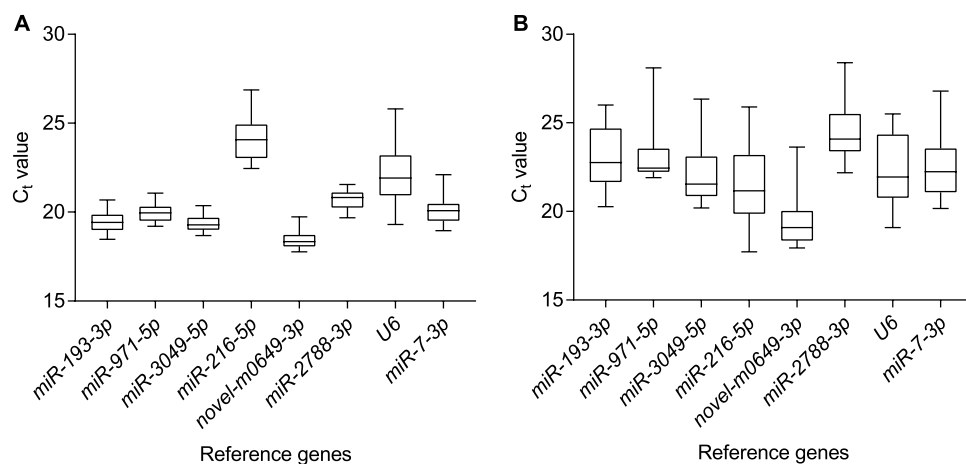
Eight genes were chosen as candidate reference genes: *miR-193-3p*, *miR-971-5p*, *miR-3049-5p*, *miR-216-5p*, *miR-7-3p*, *novel-m0649-3p*, *miR-2788-3p*, *U6*. Those genes were chosen from the miRNome data of *C. formosanus*. They were screened because they had middle expression levels and were least influenced by the methoprene treatment according to the analysis of the miRNome data. A conserved miRNA, *let-7-3p*, was chosen as the target gene to evaluate normalization results of the candidate reference genes. A combination of 9 genes  $\times$  5 sampling points (0D, 1D, 5D, 10D, 15D) was tested in one qPCR run. So, 4 colonies  $\times$  2 body parts = 8 qPCR runs were conducted in total.

**Data analysis.** To test the effect of body part and sampling time on  $C_t$  values, two-way analysis of variance (ANOVA) was used to analyze  $C_t$  values for each candidate reference gene, with body part and sampling time as two factors. The gene whose  $C_t$  value is influenced by the bioassay date will be screened out for further analysis, because using this kind of gene as reference gene may skew the normalization results. To test the suitability of

RNA	RNA sequence	Stem loop primer	Probe	Forward primer	E (%)	R <sup>2</sup>
miR-193-3p	UACUGGCCUGCUAAGUCC CAAG	GTCGTATCCAGTGCAGGGTCC GAGGTATTCGCACTGGATACG ACCTTGGG	(FAM)ATACGACCTTGG GACT(MGB)	CGTGTACTGGCCTGCTAAGT	113.3	0.997
miR-971-5p	CACUCUAAGCUCGAACAC CAAGC	GTCGTATCCAGTGCAGGGTCC GAGGTATTCGCACTGGATACG ACGCTTGG	(FAM)ATACGACGCTTGG GTG(MGB)	GCGTCACTCTAAGCTCGAACA	118.0	0.992
miR-3049-5p	UCGGAAGGCAGUUGCGG CGGACU	GTCGTATCCAGTGCAGGGTCC GAGGTATTCGCACTGGATACG ACAGTCCG	(FAM)ATACGACAGTCC GCC(MGB)	TCGGAAGGCAGTTGCG	100.5	0.991
miR-216-5p	AAAUAUCAGCUGGUAUU CUGA	GTCGTATCCAGTGCAGGGTCC GAGGTATTCGCACTGGATACG ACTCAGAA	(FAM)TGGATACGACTC AGAAT(MGB)	CCGAGCGAAATATCAGCT GGTAA	114.5	0.999
miR-7-3p	CAAGAAAUCACUCAUU CCU	GTCGTATCCAGTGCAGGGTCC GAGGTATTCGCACTGGATACG ACAGGAAG	(FAM)TGGATACGACAG GAAG(MGB)	CGACGCCAAGAAATCACTCAT	106.4	0.995
novel-m0649-3p	CCUGUAUUGGUCACUCU CCU	GTCGTATCCAGTGCAGGGTCC GAGGTATTCGCACTGGATACG ACAGGAGA	(FAM)TGGATACGACAG GAGAG(MGB)	CGTGGAGCCTGTATATGG TCAC	110.2	0.995
miR-2788-3p	CAAUGCCCUUGGAAUCCCA	GTCGTATCCAGTGCAGGGTCC GAGGTATTCGCACTGGATACG ACTGGGAT	(FAM)TGGATACGACTG GGATT(MGB)	CGAGGTCAATGCCCTTGAA	112.2	0.985
U6	ACUAAAAUUGGAACGAUA CAGAGAAGAUAGCAUGG CCCUGCGCAAGGAUGAC ACGCAAAUCGUGAAGCG UCCACAUUUUU	GTCGTATCCAGTGCAGGGTCC GAGGTATTCGCACTGGATACG ACAAAAAT	(FAM)CTGGATACGACA AAAATG(MGB)	GCAAAATCGTGAAGCGTT CCA	104.0	0.995
let-7-3p	CUGUACAACUUGCUAACU UCC	GTCGTATCCAGTGCAGGGTCC GAGGTATTCGCACTGGATACG ACGGAAG	(FAM)TGGATACGACGG AAAGT(MGB)	TGCGACCTGTACAACCTT CTAA	105.6	0.995

**Table 1.** The candidate reference genes, the target gene, the stem loop primers, the probes, the forward primers, and the PCR efficiency.

the genes as the candidate reference genes, the  $C_t$  values of the candidate reference genes were analyzed using the four methods: geNorm<sup>37</sup>, NormFinder<sup>38</sup>, BestKeeper<sup>39</sup>, and  $\Delta C_t$  method<sup>40</sup>. The  $C_t$  value was also analyzed by RefFinder to get a comprehensive stability value<sup>41</sup>. After confirming the suitable genes, correlation was conducted between most stable reference genes and other reference genes to find out the suitable combination to normalize the target gene<sup>19</sup>. To validate the screened reference genes, the expression levels of *let-7-3p* were compared among different sampling dates through normalizing to difference reference genes. The  $C_t$  value of the target gene was normalized to the  $C_t$  value of the most stable reference gene, the least stable reference gene and the combination of two reference genes screened out, respectively<sup>37,42,43</sup>. The relative expression levels were log<sub>2</sub> transformed, and then analyzed with one-way ANOVA.



**Figure 1.** Expression levels of the 8 candidate reference genes in the methoprene bioassay in *Coptotermes formosanus* Shiraki. The median, the upper and lower quartile are shown in the box plots. The expression levels of two body parts are presented separately. (A) The expression levels of the 8 candidate reference genes in the head, (B) The expression levels of the 8 candidate reference genes in the thorax + abdomen.

## Results

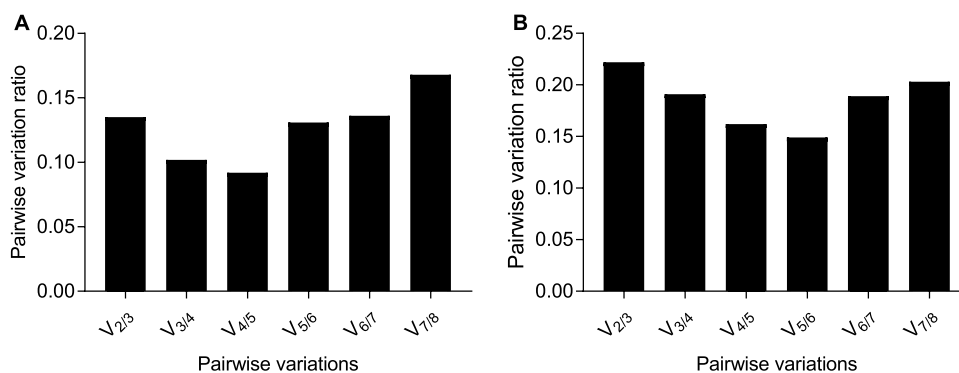
**Expression profiles of candidate reference genes.** The efficiency of the primers is summarized in Table 1. The  $C_t$  values of the 8 candidate reference genes are shown in Fig. 1. Both the  $C_t$  values from the head and those from the abdomen + thorax had similar range (Fig. 1). However, more variation was observed in the  $C_t$  values in the thorax + abdomen compared to that in the head. According to the two-way ANOVA analysis of the  $C_t$  value for each gene, no effect of sampling time on the  $C_t$  values was observed for all the genes (Table S1.). Therefore, no genes were excluded for further analysis.

**Expression stability of candidate reference genes in the head.** Different algorithms returned the results slightly differently (Table 2). According to the comprehensive analysis tool RefFinder, *novel-m0649-3p* was the most stable gene in the head.  $\Delta C_t$  method, NormFinder and geNorm also ranked *novel-m0649-3p* as the most stable gene. However, the result of BestKeeper was slightly different. BestKeeper reports two criteria: SD and [r] as the stability values. *Novel-m0649-3p* ranked as the most stable genes according to SD criterion. However, *novel-m0649-3p* only ranked as the third most stable gene based on the value of [r]. *miR-216-5p* was the most stable gene according to [r]. Although the most stable gene was different, *U6* ranked as the least stable gene according to all of the five algorithms. geNorm calculates the optimal number of genes required for normalization of gene expression. According to pairwise variation analysis, *novel-m0649-3p* and *miR-193-3p* was enough for normalization in the head, as  $V_{2/3} = 0.135$ , which was less than the threshold line (0.15) required by the algorithm (Fig. 2A). As suggested by Babion et al.<sup>19</sup>, if two genes are highly correlated ( $r > 0.9$ ), i.e. having similar expression pattern, using the combination of two genes to conduct normalization will not result in less technical variation of the target gene than using single gene. The practicability of using the combination of *novel-m0649-3p* and *miR-193-3p* for normalization was also demonstrated by the correlation coefficient between those two genes ( $r = 0.76$ ), which was less than 0.9 (Fig. 3A).

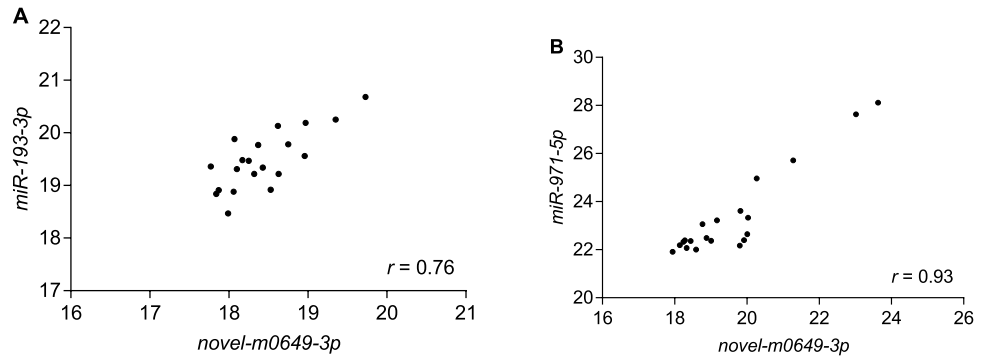
**Expression stability of candidate reference genes in the thorax + abdomen.** The selection of the most stable gene was consistent in all algorithms: all the five methods selected *novel-m0649-3p* as the most stable genes (Table 3). However, the second most stable gene resulted from different algorithms differed slightly from each other.  $\Delta C_t$  method and NormFinder reported *miR-7-3p*, geNorm reported *miR-971-5p* as the second most stable gene, respectively. The comprehensive method, RefFinder, reported *miR-971-5p* as the second most stable

RefFinder		$\Delta C_t$		NormFinder		geNorm		BestKeeper				
Gene	GM	Gene	SV	Gene	SV	Gene	SV	Gene	SD	Gene	[r]	p value
<i>novel-m0649-3p</i>	1.00	<i>novel-m0649-3p</i>	0.68	<i>novel-m0649-3p</i>	0.178	<i>novel-m0649-3p</i>	0.374	<i>novel-m0649-3p</i>	0.40	<i>miR-216-5p</i>	0.922	0.001
<i>miR-193-3p</i>	2.11	<i>miR-193-3p</i>	0.74	<i>miR-193-3p</i>	0.429	<i>miR-193-3p</i>	0.374	<i>miR-971-5p</i>	0.41	<i>U6</i>	0.918	0.001
<i>miR-971-5p</i>	2.91	<i>miR-971-5p</i>	0.78	<i>miR-7-3p</i>	0.443	<i>miR-971-5p</i>	0.420	<i>miR-3049-5p</i>	0.42	<i>novel-m0649-3p</i>	0.913	0.001
<i>miR-2788-3p</i>	4.47	<i>miR-7-3p</i>	0.82	<i>miR-971-5p</i>	0.550	<i>miR-2788-3p</i>	0.447	<i>miR-2788-3p</i>	0.43	<i>miR-7-3p</i>	0.857	0.001
<i>miR-7-3p</i>	4.56	<i>miR-2788-3p</i>	0.84	<i>miR-2788-3p</i>	0.642	<i>miR-3049-5p</i>	0.479	<i>miR-193-3p</i>	0.44	<i>miR-193-3p</i>	0.762	0.001
<i>miR-3049-5p</i>	4.82	<i>miR-3049-5p</i>	0.85	<i>miR-3049-5p</i>	0.645	<i>miR-7-3p</i>	0.596	<i>miR-7-3p</i>	0.68	<i>miR-971-5p</i>	0.633	0.003
<i>miR-216-5p</i>	7.00	<i>miR-216-5p</i>	0.97	<i>miR-216-5p</i>	0.719	<i>miR-216-5p</i>	0.714	<i>miR-216-5p</i>	1.01	<i>miR-2788-3p</i>	0.523	0.018
<i>U6</i>	8.00	<i>U6</i>	1.39	<i>U6</i>	1.328	<i>U6</i>	0.883	<i>U6</i>	1.46	<i>miR-3049-5p</i>	0.512	0.021

**Table 2.** Ranking of the stability values of the 8 candidate reference genes by the five algorithms in the head.



**Figure 2.** Pairwise variation analysis to determine the optimal number of reference genes using geNorm. (A) The pairwise variation analysis result in the head, (B) The pairwise analysis result in the thorax + abdomen.



**Figure 3.** Correlation of the  $C_t$  values of the two most stable reference genes. **(A)** Correlation of the  $C_t$  values of the two most stable reference genes in the head. **(B)** Correlation of the  $C_t$  values of the two most stable reference genes in the thorax + abdomen.

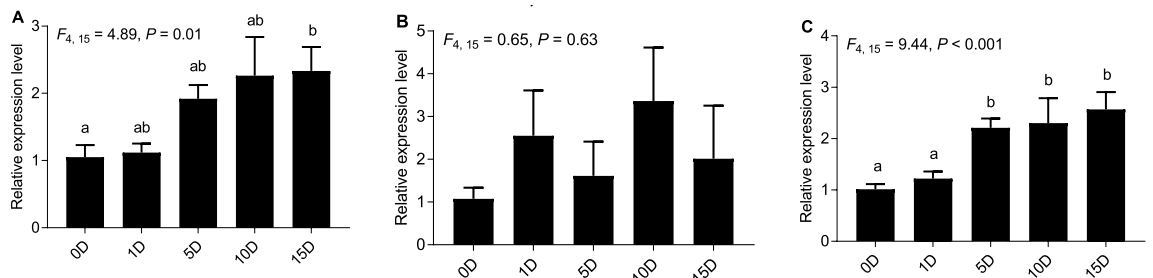
RefFinder		$\Delta C_t$		NormFinder		geNorm		BestKeeper				
Gene	GM	Gene	SV	Gene	SV	Gene	SV	Gene	SD	Gene	[r]	p value
<i>novel-m0649-3p</i>	1.00	<i>novel-m0649-3p</i>	0.95	<i>novel-m0649-3p</i>	0.111	<i>novel-m0649-3p</i>	0.675	<i>novel-m0649-3p</i>	1.16	<i>novel-m0649-3p</i>	0.977	0.001
<i>miR-971-5p</i>	2.45	<i>miR-7-3p</i>	1.04	<i>miR-7-3p</i>	0.446	<i>miR-971-5p</i>	0.675	<i>miR-2788-3p</i>	1.28	<i>miR-7-3p</i>	0.972	0.001
<i>miR-7-3p</i>	3.13	<i>miR-971-5p</i>	1.12	<i>miR-971-5p</i>	0.765	<i>miR-2788-3p</i>	0.722	<i>miR-3049-5p</i>	1.31	<i>miR-971-5p</i>	0.916	0.001
<i>miR-2788-3p</i>	3.13	<i>miR-2788-3p</i>	1.14	<i>miR-2788-3p</i>	0.767	<i>miR-7-3p</i>	0.791	<i>miR-971-5p</i>	1.33	<i>miR-216-5p</i>	0.898	0.001
<i>miR-193-3p</i>	5.00	<i>miR-193-3p</i>	1.18	<i>miR-193-3p</i>	0.792	<i>miR-193-3p</i>	0.853	<i>miR-193-3p</i>	1.33	<i>miR-2788-3p</i>	0.892	0.001
<i>miR-3049-5p</i>	5.05	<i>miR-3049-5p</i>	1.29	<i>miR-3049-5p</i>	1.003	<i>miR-3049-5p</i>	0.915	<i>miR-7-3p</i>	1.50	<i>miR-193-3p</i>	0.888	0.001
<i>miR-216-5p</i>	7.24	<i>miR-216-5p</i>	1.41	<i>miR-216-5p</i>	1.118	<i>miR-216-5p</i>	1.063	U6	1.66	<i>miR-3049-5p</i>	0.846	0.001
U6	7.74	U6	1.75	U6	1.597	U6	1.236	<i>miR-216-5p</i>	1.79	U6	0.697	0.001

**Table 3.** Ranking of the stability values of the 8 candidate reference genes by the five algorithms in the abdomen + thorax.

gene. Similar to the situation in the head, U6 also ranked as the least stable genes by all the algorithms except the SD criteria of BestKeeper. According to the SD criteria of BestKeeper, *miR-216-5p* ranked as the least stable gene.

The number of genes required for normalization in the thorax + abdomen was more than that in the head. Five genes were recommended for normalization, as  $V_{5/6}$  was the first number which was less than 0.15 ( $V_{5/6} = 0.149$ ) (Fig. 2B). However, it is not convenient in practice to use five genes for normalization. The correlation coefficient between *novel-m0649-3p* and *miR-971-5p* was more than 0.9 ( $r = 0.93$ ) (Fig. 3B). Therefore, the combination of *novel-m0649-3p* and *miR-971-5p* for normalization was refused. After the correlation coefficient between *novel-m0649-3p* and other genes was checked, correlation coefficient between *novel-m0649-3p* and *miR-2788-3p* was less than 0.9 ( $r = 0.899$ ). So, combination of *novel-m0649-3p* and *miR-2788-3p* was suggested for normalization.

**Effect of different normalization strategies on *let-7-3p* quantification in the head.** Using different normalization method gave rise to different gene expression results. Significant difference among different assay dates was found using *novel-m0649-3p* as the reference gene ( $F_{4, 15} = 4.89, P = 0.01$ ), while no significant



**Figure 4.** Expression profiles of *let-7-3p* using different reference genes in the head after treating workers with methoprene. **(A)** Expression profile of *let-7-3p* normalized to the most stable reference gene (*novel-m0649-3p*) in the head. **(B)** Expression profile of *let-7-3p* normalized to the least stable reference gene (U6) in the head. **(C)** Expression profile of *let-7-3p* normalized to the combination of *novel-m0649-3p* and *miR-193-3p*.

difference was found using *U6* as the reference gene ( $F_{4,15} = 0.65$ ,  $P = 0.63$ ) in the head (Fig. 4A, B). Smaller  $P$  value was obtained using the combination of *novel-m0649-3p* and *miR-193-3p* for normalization than that using *novel-m0649-3p* only in the head (Fig. 4C). According to Fig. 4, *let-7-3p* exhibited a clear rising pattern when normalized to the most stable reference gene, or the combination of the two most stable reference genes in comparison with when normalized to the least stable reference gene *U6*. No clear expression pattern of *let-7-3p* was observed using the *U6* as the normalization gene.

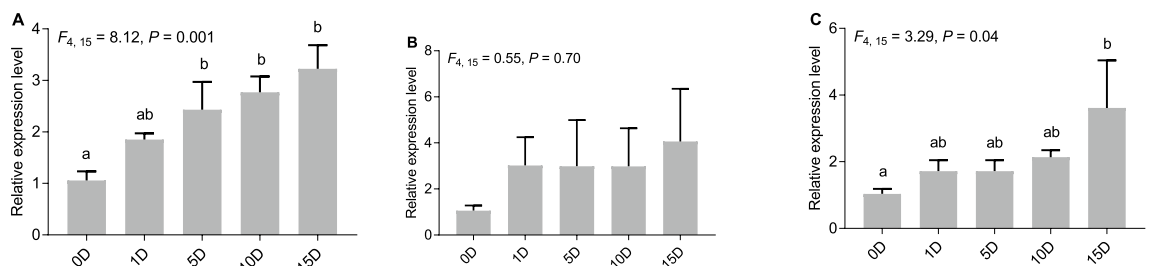
**Effect of different normalization strategies on *let-7-3p* quantification in the thorax + abdomen.** There were similar results in the thorax + abdomen using *novel-m0649-3p* as the reference gene: significant difference was found ( $F_{4,15} = 8.12$ ,  $P = 0.001$ ), while no significant difference was found using *U6* as the reference gene ( $F_{4,15} = 0.55$ ,  $P = 0.70$ ) (Fig. 5A, B). However, larger  $P$  value was observed for the combination of *novel-m0649-3p* and *miR-2788-3p* than that for *novel-m0649-3p* only in the thorax + abdomen (Fig. 5C).

## Discussion

RT-qPCR is a powerful and sensitive tool for gene quantification. Normalization to a reference gene which is stably expressed is a popular method to eliminate the error in the relative quantification method. However, there are no genes whose expression is stable in all the experimental conditions. As more attention was paid to the influence of reference gene on the quantification results, a range of reference genes screening researches were conducted both for the mRNA and miRNA quantification. Here, we screened the reference genes for miRNA expression analysis in the study of soldier caste differentiation. Our data showed that *novel-m0649-3p* was the most stable reference gene both in the head and thorax + abdomen. *U6*, which is widely used for normalization of miRNA quantification data, was not a good reference gene. In addition, normalization to two reference genes was better than that to one reference gene in the head. More variation of the Ct values was observed in the thorax + abdomen than that in the head. The most obvious changes after treatment of workers with methoprene happen in the abdomen, such as enlargement of fat body<sup>44</sup>, massive expression of the hexamerin<sup>8</sup> etc. Therefore, more response genes are expected to be observed in the thorax + abdomen. Colony response variability was observed in *C. formosanus* after methoprene treatment<sup>35</sup>, implying gene expression variability at the colony level. Therefore, this may be one reason for more variation in the thorax + abdomen.

*Let-7* is first discovered in *Caenorhabditis elegans* (Maupas) as a gene switcher. Its function is conserved across species<sup>45</sup>. The expression levels of *let-7-3p* was analyzed using different reference genes in our study. *U6* is used widely as reference gene. However, our analysis recognizes *U6* as the least stable reference gene. When our data were normalized to *U6*, the data became more scattered in comparison to normalizing to *novel-m0649-3p* gene, and no significant difference was found. In contrast, using *novel-m0649-3p* as the reference gene produced a clear rising pattern of *let-7-3p*. The pattern of expression profile of *let-7-3p* was different using the most stable and the least stable reference genes to normalize the data, which demonstrates the importance of choosing the suitable reference genes for normalization of the qPCR data. So, it is suggested to screen the reference gene for the specific purpose.

Although all the algorithms identified *novel-m0649-3p* as the most stable reference genes, the ranking of other genes was slightly different by different algorithms. Different algorithms returned slightly different results, as the underlying principles vary among the five algorithms. geNorm calculates the gene stability value  $M$  to represent the stability of the expression of the gene. The less the value of  $M$ , the more stable of the gene. The pairwise variation value ( $V$  value) is introduced to determine the number of the reference genes needed. If  $V_n/V_{n+1}$  is less than 0.15, number of  $n$  genes is suggested for normalization. In the head,  $V_{2/3}$  was less than 0.15, so a combination of 2 genes was enough for normalization. However, combination of 5 genes was needed for normalization in the thorax + abdomen, as  $V_{5/6} = 0.149$ . Nevertheless, it is not convenient to use 5 genes for normalization. The combination of two genes was still recommended for normalization in the thorax + abdomen. As suggested by Babion et al.<sup>19</sup>, the combination of two genes for normalization should not be highly correlated. So, in the thorax + abdomen, combination of *novel-m0649-3p* and *miR-2788-3p* was suggested for normalization. The combination of two genes for normalization gave rise to smaller  $P$  value compared to normalization to one gene in the head. However, because the two reference genes in the combination were closely correlated with each other in the abdomen+thorax, using two reference genes did not lead to better results than using the



**Figure 5.** Expression profiles of *let-7-3p* using different reference genes in the thorax + abdomen after treating workers with methoprene. (A) Expression profile of *let-7-3p* normalized to the most stable reference gene (*novel-m0649-3p*) in the thorax + abdomen. (B) Expression profile of *let-7-3p* normalized to the least stable reference gene (*U6*) in the thorax + abdomen. (C) Expression profile of *let-7-3p* normalized to the combination of *novel-m0649-3p* and *miR-2788-3p*.

most stable reference gene. Normfinder estimates the stability value of a candidate reference genes by calculating the intragroup and intergroup variations of the candidate reference genes. The stability value represents the systematic error using a certain gene for normalization. However, Normfinder requires larger sample size than geNorm. Bestkeeper calculates two parameters as the gene stability criteria. One is the standard deviation of the  $C_t$  value; another is the correlation coefficient between the candidate reference gene and the BestKeeper Index. Although difference in the gene stability ranking was observed for the five algorithms, the results obtained by different algorithms were somewhat similar, which indicates that the five algorithms have the capability to differentiate the relative stability of the candidate reference genes.

There are two kinds of methods for qPCR: the dye-based method and the probe-based method. The probe-based method is more sensitive and more specific compared with the dye-based method. So, the probe-based method was adopted in our project. In order to acquire the template for qPCR, reverse transcription of miRNAs is needed. There are two kinds of primers for reverse transcription of miRNAs: stem-loop RT primer and poly(A)-tailing RT primer. According to our preliminary results, poly(A)-tailing method was likely to produce negative results, because there was unspecific amplification according to the melting curve and the gel image of the agarose gel electrophoresis of the qPCR products. Therefore, the stem-loop RT primer was used for the reverse transcription of miRNAs in our project. Chen et al.<sup>36</sup> suggested using 50 nM stem-loop RT primers in the probe-based method, while Kramer<sup>46</sup> suggested using 1 nM of the primer. According to our preliminary experiment, the concentration of 30 nM stem-loop RT primers led to more sensitive and more efficient amplification during qPCR. As the RT primer has the binding site of the reverse primer of qPCR, too high a concentration of the RT primer may lead to the binding of the reverse primer with the RT primer, which results in less efficient amplification.

Our study has established normalization strategies for expression analysis of miRNAs in the study of soldier caste differentiation in *C. formosanus*. It is suggested to use *novel-m0649-3p* or the combination of *novel-m0649-3p* and *miR-193-3p* for normalization in the head, and to use *novel-m0649-3p* as the reference gene in the thorax + abdomen. The study of the differentiation mechanism is mainly focused at the mRNA and proteins level. miRNAs are important post-transcriptional regulators. There is still a gap in the studies of the roles of miRNAs in soldier caste regulation. Screening of the suitable reference genes establishes the first step to study the miRNA expression level during soldier caste differentiation. Further studies of the roles of miRNAs in soldier caste differentiation will widen our understanding of the caste differentiation in termites.

## Data availability

The miRNome data were submitted to GenBank Sequence Read Archive databases under the BioProject PRJNA922245.

Received: 16 January 2023; Accepted: 25 May 2023

Published online: 09 June 2023

## References

- Oster, G. F. & Wilson E. O. Caste and ecology in the social insects. (Princeton Univ. Press, 1978).
- Brian, M. V. Caste differentiation and division of labor. In *Social Insects*, Vol. 1, 121–222 (Academic Press Inc., 1979).
- Haverty, M. I. The proportion of soldiers in termite colonies: a list and a bibliography. *Sociobiology* **2**, 199–216 (1977).
- Lefeuvre, P. & Bordereau, C. Soldier formation regulated by a primer pheromone from the soldier frontal gland in a higher termite, *Nasutitermes lujae*. *Proc. Natl. Acad. Sci. USA* **81**, 7665–7668 (1984).
- Lüscher, M. & Springhetti, A. Untersuchungen über die Bedeutung der Corpora allata für die Differenzierung der Kasten bei der Termiten *Kaloterms flavicollis* F. *J. Insect Physiol.* **5**, 190–212 (1960).
- Liischer, M. Die Bedeutung des Juvenilhormons für die Differenzierung der Soldaten bei Termiten *Kaloterms flavicollis*. In *Proceedings, 6th Congress of the International Union for Study of Social Insects*, Beme, Switzerland, 165–170 (1969).
- Korb, J. Juvenile hormone: a central regulator of termite caste polyphenism. In *Advances in Insect Physiology* **48**, 131–161 (Academic Press Inc., 2015).
- Zhou, X., Oi, F. M. & Scharf, M. E. Social exploitation of hexamerin: RNAi reveals a major caste-regulatory factor in termites. *Proc. Natl. Acad. Sci. USA* **103**, 4499–4504 (2006).
- Hattori, A. et al. Soldier morphogenesis in the damp-wood termite is regulated by the insulin signaling pathway. *J. Exp. Zool. Part B* **320**, 295–306 (2013).
- Masuoka, Y., Yaguchi, H., Suzuki, R. & Maekawa, K. Knockdown of the juvenile hormone receptor gene inhibits soldier-specific morphogenesis in the damp-wood termite *Zootermopsis nevadensis* (Isoptera: Archotermsidae). *Insect Biochem. Mol. Biol.* **64**, 25–31 (2015).
- Masuoka, Y., Yaguchi, H., Toga, K., Shigenobu, S. & Maekawa, K. TGFβ signaling related genes are involved in hormonal mediation during termite soldier differentiation. *PLoS Genet.* **14**, e1007338 (2018).
- Huntzinger, E. & Izaurralde, E. Gene silencing by microRNAs: contributions of translational repression and mRNA decay. *Nat. Rev. Genet.* **12**, 99–110 (2011).
- Belles, X. MicroRNAs and the evolution of insect metamorphosis. *Annu. Rev. Entomol.* **62**, 111–125 (2017).
- Lucas, K. J., Zhao, B., Liu, S. & Raikhel, A. S. Regulation of physiological processes by microRNAs in insects. *Curr. Opin. Insect Sci.* **11**, 1–7 (2015).
- Song, J. & Zhou, S. Post-transcriptional regulation of insect metamorphosis and oogenesis. *Cell. Mol. Life Sci.* **77**, 1893–1909 (2020).
- Liu, L. et al. microRNAs shape social immunity: a potential target for biological control of the termite *Reticulitermes chinensis*. *J. Pest Sci.*, 1–15 (2022).
- Itakura, S., Hattori, K. & Umezawa, K. Identification and expression analysis of microRNAs in worker caste termites of *Coptotermes formosanus* and *Reticulitermes speratus*. *J. Asia-Pac. Entomol.* **21**, 388–393 (2018).
- Matsunami, M. et al. Caste-specific microRNA expression in termites: insights into soldier differentiation. *Insect Mol. Biol.* **28**, 86–98 (2019).
- Babion, I., Snoek, B. C., van de Wiel, M. A., Wilting, S. M. & Steenbergen, R. D. A strategy to find suitable reference genes for miRNA quantitative PCR analysis and its application to cervical specimens. *J. Mol. Diagn.* **19**, 625–637 (2017).
- Masé, M. et al. Selection of reference genes is critical for miRNA expression analysis in human cardiac tissue: A focus on atrial fibrillation. *Sci. Rep.* **7**, 41127 (2017).

21. Tricarico, C. *et al.* Quantitative real-time reverse transcription polymerase chain reaction: normalization to rRNA or single house-keeping genes is inappropriate for human tissue biopsies. *Anal. Biochem.* **309**, 293–300 (2002).
22. Bas, A., Forsberg, G., Hammarström, S. & Hammarström, M. L. Utility of the housekeeping genes 18S rRNA,  $\beta$ -actin and glyceraldehyde-3-phosphate-dehydrogenase for normalization in real-time quantitative reverse transcriptase-polymerase chain reaction analysis of gene expression in human T lymphocytes. *Scand. J. Immunol.* **59**, 566–573 (2004).
23. Du, H., Wu, W., Huang, X. & Li, Z. Screening of reference genes for expression analysis in the study of soldier caste differentiation of Formosan subterranean termite *Coptotermes formosanus* Shiraki. *PeerJ* **7**, e7981 (2019).
24. Peltier, H. J. & Latham, G. J. Normalization of microRNA expression levels in quantitative RT-PCR assays: identification of suitable reference RNA targets in normal and cancerous human solid tissues. *RNA* **14**, 844–852 (2008).
25. Zavala, E., Reyes, D., Deerenberg, R. & Vidal, R. Selection of reference genes for microRNA analysis associated to early stress response to handling and confinement in *Salmo salar*. *Sci. Rep.* **7**, 1756 (2017).
26. Broggin, C., Abril, N., Carranza, J. & Membrillo, A. Evaluation of candidate reference genes for quantitative real-time PCR normalization in blood from red deer developing antlers. *Sci. Rep.* **12**, 16264 (2022).
27. Dong, J., Li, J., Huang, J. & Wu, J. Identification of suitable reference genes for miRNA quantitation in bumblebee (Hymenoptera: Apidae) response to reproduction. *Apidologie* **50**, 40–50 (2019).
28. He, F. *et al.* Validation of reference genes for gene expression analysis in fruit development of *Vaccinium bracteatum* Thunb. using quantitative real-time PCR. *Sci. Rep.* **12**, 16946 (2022).
29. Lü, J., Yang, C., Zhang, Y. & Pan, H. Selection of reference genes for the normalization of RT-qPCR data in gene expression studies in insects: A systematic review. *Front. Physiol.* **9**, 1560 (2018).
30. Sellamuthu, G., Bily, J., Joga, M. R., Synek, J. & Roy, A. Identifying optimal reference genes for gene expression studies in Eurasian spruce bark beetle, *Ips typographus* (Coleoptera: Curculionidae: Scolytinae). *Sci. Rep.* **12**, 4671 (2022).
31. Wang, M. *et al.* Identification of RT-qPCR reference genes suitable for gene function studies in the pitaya canker disease pathogen *Neoscytalidium dimidiatum*. *Sci. Rep.* **12**, 22357 (2022).
32. Yang, C. *et al.* Selection of reference genes for RT-qPCR analysis in a predatory biological control agent, *Coleomegilla maculata* (Coleoptera: Coccinellidae). *Sci. Rep.* **5**, 18201 (2015).
33. Yang, Y. *et al.* Identification and evaluation of suitable reference genes for normalization of MicroRNA expression in *Helicoverpa armigera* (Lepidoptera: Noctuidae) using quantitative real-time PCR. *J. Insect Sci.* **17**, 1–10 (2017).
34. Zhang, L. *et al.* Selection of reference genes for expression analysis of plant-derived microRNAs in *Plutella xylostella* using qRT-PCR and ddPCR. *PLoS ONE* **14**, e0220475 (2019).
35. Du, H. *et al.* Response variability of juvenile hormone and juvenile hormone analogs in workers of the Formosan subterranean termite (*Coptotermes formosanus* Shiraki). *J. Asia-Pac. Entomol.* **23**, 1010–1013 (2020).
36. Chen, C. *et al.* Real-time quantification of microRNAs by stem-loop RT-PCR. *Nucleic Acids Res.* **33**, e179 (2005).
37. Vandesompele, J. *et al.* Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* **3**, research0034.1 (2002).
38. Andersen, C. L., Jensen, J. L. & Ørntoft, T. F. 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 (2004).
39. Pfaffl, M. W., Tichopad, A., Prgomet, C. & Neuvians, T. P. 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 (2004).
40. Silver, N., Best, S., Jiang, J. & Thein, S. L. Selection of housekeeping genes for gene expression studies in human reticulocytes using real-time PCR. *BMC Mol. Biol.* **7**, 33 (2006).
41. Xie, F., Xiao, P., Chen, D., Xu, L. & Zhang, B. miRDeepFinder: A miRNA analysis tool for deep sequencing of plant small RNAs. *Plant Mol. Biol.* **80**, 75–84 (2012).
42. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the  $2^{-\Delta\Delta CT}$  method. *Methods* **25**, 402–408 (2001).
43. Hellemans, J., Mortier, G., De Paep, A., Speleman, F. & Vandesompele, J. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol.* **8**, R19 (2007).
44. Cornette, R., Matsumoto, T. & Miura, T. Histological analysis of fat body development and molting events during soldier differentiation in the damp-wood termite, *Hodotermopsis sjostedti* (Isoptera, Termopsidae). *Zool. Sci.* **24**, 1066–1074 (2007).
45. Roush, S. & Slack, F. J. The *let-7* family of microRNAs. *Trends Cell Biol.* **18**, 505–516 (2008).
46. Kramer, M. F. Stem-loop RT-qPCR for miRNAs. *Curr. Protoc. Mol. Biol.* Chapter 15: Unit 15.10 (2011).

## Acknowledgements

This research was supported by the GDAS Project of Science and Technology Development (Grant No. 2019GDA-SYL-0104018), the National Natural Science Foundation of China (Grant No. 31702068), Project of State Key Laboratory of Integrated Management of Pest Insects and Rodents (Grant No. IPM2211), and the GIZ (Grant No. GIABR-gjrc201601).

## Author contributions

Conceptualization, H.D.; Investigation, R.H.; Methodology, H.D. and D.C.; Data curation, H.D. and R.H.; Writing—review & editing, H.D., Z.L., D.C. and H.Z.; Writing—original draft, H.D.; Funding acquisition, Z.L., H.D. and H.Z.; Formal analysis, H.D.; Resources, Z.L. and C.H.; Supervision, Z.L. All authors have read and approved the final manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-35926-7>.

**Correspondence** and requests for materials should be addressed to Z.L.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023