scientific reports

Check for updates

OPEN

Validation of reliable reference genes for comparison of gene expression across species in the *Anopheles* Hyrcanus Group

Do Eun Lee[1,2], Si Hyeock Lee[3,4] & Ju Hyeon Kim[1,5]✉

The *Anopheles* Hyrcanus Group comprises 25 recognized mosquito species that exhibit significant variations in critical vector characteristics despite having similar morphology and ecological niches. Cross-species transcriptional profiling with appropriate reference genes is essential to evaluate the physiological differences among these closely related species. This study aimed to identify reliable pan-species reference genes among six *Anopheles* Hyrcanus Group species found in South Korea. We assessed the expression stability of eight candidate genes across five developmental stages in each of the six species. Utilizing qPCR and analysis programs, our analysis revealed that the transcription of *ribosomal protein L8* (*RPL8*) and *ribosomal protein L13a* was the most stable at the larval stage. In contrast, *ribosomal protein L32* and *ribosomal protein S17* (*RPS17*) exhibited stability across all tested adult stages. In comparisons across different developmental stages within individual species, *RPS17* emerged as a reliable reference gene for *Anopheles belenrae*, *Anopheles pullus*, *Anopheles sinensis*, and *Anopheles sineroides*, whereas *ribosomal protein S7* and *RPL8* were suitable for *Anopheles kleini* and *Anopheles lesteri*. These selected reference genes may facilitate comparative gene expression profiling within the *Anopheles* Hyrcanus Group, thereby providing insights into the molecular mechanisms underlying their differences, such as distinct vector competence and baseline insecticide susceptibility.

Keywords *Anopheles* Hyrcanus Group, Reference gene, qPCR, Ribosomal protein, Cross-species comparison, Developmental stages

The *Anopheles* Hyrcanus Group, comprising at least 25 species, is widely distributed throughout East Asia[1]. Some mosquitoes in this group are capable of transmitting *Plasmodium vivax*, and the primary vectors vary depending on the country and region because of the species distribution and their vectorial capacity[2–5]. In South Korea, six species in the *Anopheles* Hyrcanus Group (*Anopheles belenrae*, *Anopheles kleini*, *Anopheles lesteri*, *Anopheles pullus*, *Anopheles sinensis*, and *Anopheles sineroides*) have been identified in malaria high-risk areas, particularly in the northern Gyeonggi province[6]. As the molecular differentiation of *An. belenrae* and *An. kleini* was performed in 2005, which found that they were distinct species from *An. sinensis*[7], in-depth research has been conducted on the *Anopheles* Hyrcanus Group. Epidemiology studies in the northern Gyeonggi province revealed that *An. pullus* is prevalent during early summer, *An. kleini* in midsummer, and *An. sinensis* dominates from mid to late-summer[6]. Comparative studies on the vector competence of *An. sinensis*, *An. kleini*, *An. pullus,* and *An. belenrae* determined that *An. kleini* exhibits the strongest vector competence, whereas *An. sinensis* demonstrated the lowest susceptibility to *Plasmodium vivax*[4,8]. Despite their distinct physiological and ecological characteristics, comparative studies across these species have been challenging because of their overlapping habitats and difficulties in species identification.

To establish a comprehensive understanding of the differences in their epidemiologically important physiology, it is crucial to compare the expression profiles of genes associated with vector-pathogen interactions, immune responses, and xenobiotic metabolism across the different species within the *Anopheles* Hyrcanus Group. Furthermore, comparative gene expression studies among the different species offer the opportunity

[1]Department of Tropical Medicine and Parasitology, Seoul National University College of Medicine, Seoul 03080, Republic of Korea. [2]Core Research Facilities, Seoul National University College of Medicine, Seoul 03080, Republic of Korea. [3]Department of Agricultural Biotechnology, Seoul National University, Seoul 08826, Republic of Korea. [4]Research Institute of Agriculture and Life Sciences, Seoul National University, Seoul 08826, Republic of Korea. [5]Institute of Endemic Diseases, Seoul National University College of Medicine, Seoul 03080, Republic of Korea. ✉email: biomyst5@snu.ac.kr

to identify conserved genetic pathways and the regulatory elements that play fundamental roles in mosquito biology. However, existing gene expression studies have primarily considered the comparative analysis of genes related to insecticide resistance within the most predominant species, *An. sinensis*[9].

Quantitative polymerase chain reaction (qPCR) is a widely used method to compare transcription levels of specific genes and relies on relative quantification based on a reference gene. To ensure an accurate comparison of transcription levels of the target gene across samples, it is essential to select a reference gene with a consistent expression level for normalization. The reference gene is typically selected from among housekeeping genes that exhibit high expression levels. However, variations in the stability of gene transcription can occur even within the same orthologous genes across different species and can be influenced by factors such as tissue type, developmental stage, and environmental conditions[10–12]. Therefore, it is crucial to select an appropriate reference gene with confirmed transcription stability that is least influenced by varying conditions.

Most studies on insect reference genes focus on identifying appropriate internal reference genes within a particular species. Commonly used reference genes, such as ribosomal proteins, tubulin, actin, glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*), and elongation factor 1α (*EF1α*), are selected for their consistently high and stable expression levels[13]. In *Anopheles gambiae*, the *ribosomal protein S7* (*RPS7*) gene has been commonly used as a reference gene to normalize data when comparing the expression levels of cytochrome P450s (CYP450s) and odorant receptors[14,15]. For *An. sinensis*, the *ribosomal protein S18* (*RPS18*) gene was used for the transcriptome analysis of pyrethroid-resistant strains[9]. In addition, both the *RPS7* and *ribosomal protein L49* (*RPL49*) genes were used to compare the expression levels of CYP450 between resistant strains[16].

Some studies have identified reference genes specifically for the purpose of comparing gene expression levels between species or genera. For example, one study employed microarray techniques to select the most suitable reference gene for six *Eucalyptus* species[17]. Moreover, common reference genes have been identified among birds from different genera[18], as well as intergenus species of livestock[19]. Selecting genes with consistent expression levels across multiple species poses a challenge because of genetic variance and the difficulty of designing primers with acceptable efficiency for all target species. Nevertheless, employing such an approach enables research that encompasses broader and more diverse gene expression studies.

In this study, we aimed to identify suitable reference genes for (1) comparisons across developmental stages within a single species and (2) interspecies comparisons that encompass the developmental stages of six mosquito species in the *Anopheles* Hyrcanus Group. The representative developmental stages used were 4th instar larva (4L), pupa (pup), 24-h-old adult female (24 h-A), 72-h-old adult female (72 h-A), and mixed-aged, oviposited female (ovi-A). qPCR and four analysis programs (geNorm[20], Bestkeeper[21], NormFinder[22], and RefFinder[23]) were utilized to determine gene stability. Leveraging these reference genes will facilitate the investigation into physiological differences among the *Anopheles* Hyrcanus Group species.

Materials and methods

Mosquito collection and sample preparation

Adult mosquitoes were collected from three locations in Gyeonggi province, South Korea (Paju: 37.871385, 126.772513; Gimpo: 37.708802, 126.643029; and Hwaseong: 37.098884, 126.783453) using black light traps (BT global, South Korea) from June to September over a three-year period (2021–2023). Species identification was performed using a multiplex PCR method with the *internal transcribed spacer 2* gene sequences as the molecular marker[24] with slight modifications. Oviposited eggs from individual blood-fed mosquitoes were collected, and TetraMin flakes (Tetra, VA, USA) were provided daily to the hatched larvae. A total of 8–10 individuals were collected from each developmental stage (4L, pup, 24 h-A, 72 h-A, and ovi-A) for each species and stored at − 80℃ until RNA extraction. Due to the low population density, 2–3 individuals were used for the oviposited stage of *An. kleini*, *An. lesteri*, and *An. sineroides*.

To obtain the mRNA sequences of each gene from the six mosquito species used in this study, 10 female mosquitoes were homogenized with a 2.0 mm stainless-steel bead using a bullet blender (Next Advance, NY, USA), and total RNA was extracted using TRI reagent (MRC, OH, USA) according to the manufacturer's protocol. The samples were treated with DNase I (Takara Bio Inc., Shiga, Japan) to remove any DNA contamination. cDNA was synthesized from 1.5 μg of DNase I-treated RNA using SuperScript IV reverse transcriptase (Invitrogen, Carlsbad, USA).

Verification of candidate reference gene sequences

Eleven candidate genes were initially selected based on previous transcriptional studies on mosquitoes and commonly used housekeeping genes in other insect species[13,16,25]. These included: *actin*, *alpha-tubulin* (*α-tub*), *EF1α*, *GAPDH*, *ribosomal protein L13a* (*RPL13a*), *ribosomal protein L32* (*RPL32*), *RPL49*, *ribosomal protein L8* (*RPL8*), *ribosomal protein S17* (*RPS17*), *RPS18*, and *RPS7*. Gene fragments were amplified using primers designed from the conserved regions of each gene obtained from at least five species of *Anopheles* mosquitoes with completed genome sequencing (Supplementary tables 1, 2). The reaction mixture (25 μL total) contained 10 ng of template cDNA, 2 μL of 2.5 mM dNTP, 2.5 μL of 10 × buffer, 0.4 μM of each primer, 0.12 μL of EX Taq polymerase (Takara Bio Inc.), and double distilled water (ddH_2O) up to 25 μL. A 3 min preincubation at 95℃ was followed by 34 cycles at 95℃ for 20 s, 55℃ for 30 s, and 72℃ for 1 min, with a final extension at 72°℃ for 5 min. The PCR products were purified using a Monarch Clean-up kit (New England Biolabs, Ipswich, MA).

Amplified fragments less than 600 bp in size (*RPL49*, *RPL32*, *RPS17*, *RPS18*, and *RPS7*) were directly sequenced, whereas larger amplicons were cloned for precise sequencing. Purified amplicons were ligated into the pGEM-T easy vector (Promega, Madison, MU, USA) using T4 DNA ligase (Promega) and transformed into DH5α competent cells. After overnight culture at 37℃ on an LB agar plate, colonies were isolated and subsequently cultured in LB broth. Plasmids were then extracted using the AccuPrep Plasmid Mini Extraction

Kit (Bioneer, Daejeon, Korea) and sequenced by Macrogen (Seoul, Korea) using an ABI3730xl genetic analyzer. The obtained sequences were submitted to GenBank (Supplementary table 3).

qPCR primer design and efficiency testing

To design a common primer set for qPCR in the conserved regions, candidate gene sequences from six species were aligned using CLC Main Workbench 8.1.2 software (QIAGEN, Aarhus, Denmark). Primers with a length of 20–21 bp, a GC content of 45–60%, and product sizes ranging from 90 to 180 bp were selected. Primer dimerization was checked using a Multiple Primer Analyzer (ThermoFisher Scientific). qPCR was performed with varying amounts of cDNA (0.024, 0.12, 0.6, 3, and 15 ng) to assess efficiency. The reaction mix consisted of 4 µL of cDNA, 1 µL of the 5 pmol primer set, and 5 µL of TB Green premix EX Taq (Takara Bio Inc.). Amplification followed the following thermal protocol: 95℃ for 60 s (initial step); 35 cycles of 95℃ for 15 s, 57℃ for 20 s, and 72℃ for 25 s; and a final melting curve analysis from 60–95℃ at a rate of 1℃/sec. Fluorescence was measured in each cycle and during the melting curve analysis. The C_t values and efficiency of each primer set in the six different species were automatically calculated using LightCycler 96 software (Roche). Following the efficiency test, the same qPCR reactions were performed for all sample groups and genes using 10 ng of cDNA with two technical and three biological replicates.

Stability evaluations and validation of reference genes

Four normalization methods—delta C_t, geNorm, BestKeeper, and NormFinder—were used to evaluate the stability of the eight candidate reference genes, with RefFinder employed to determine the overall ranking. The delta C_t method calculates the average differences in threshold cycle (C_t) values between pairs of genes based on C_q values to identify the gene with the highest expression stability[26]. Using one gene as a reference, the delta Ct values for the remaining genes were calculated, and the standard deviations (SDs) were derived. This process was repeated for each of the eight candidate genes. The geNorm program evaluates stability based on stability value (M), which represents the average pairwise variation (V) between genes, with a cutoff value of 0.15. It also identifies the optimal number of reference genes for normalization through pairwise variation analysis (V_n/V_{n+1})[20]. BestKeeper algorithm identifies stable reference genes based on minimal C_q variations. It calculates metrics such as geometric mean, arithmetic mean, standard deviation (SD), and coefficient of variation (CV), and performs pairwise correlation analyses to determine the best reference genes[21]. NormFinder ranks candidate genes according to their expression stability while considering sample grouping or experimental design[22]. The most stable genes with the lowest stability value (SV) were identified based on log-transformed C_q values.

For validation of the selected reference genes for each species and developmental stage, *CHS2* and *Vg1* were tested as the target genes using the two most stable reference genes and the least stable reference gene for each test. qPCR was performed using 10 ng of template cDNA, and the reaction mixture and conditions were the same as described for the efficiency test.

Results

Primer set design and efficiency test

Primer sets for 10 of these candidate genes (excluding *RPS18*) were designed from conserved regions across the six *Anopheles* Hyrcanus Group species (Table 1). Although the *RPS18* gene was used in a previous study with *An. sinensis*, it was excluded from our study because no suitable primers that met our criteria (a length of 20–21 bp, a GC content of 45%–60%, and a product size ranging from 90 to 180 bp) could be found within the conserved regions. After efficiency testing, the *RPL49* and *α-tub* primers, which exhibited lower average amplification efficiencies (89.1 and 92.1%, respectively) than the other reference genes, were excluded from further analysis. The remaining primers demonstrated robust performance, with average efficiencies ranging from 95.8 to 103.1%. Notably, their species-specific efficiencies ranged from 90.7 to 106%, with an $R^2 \geq 0.99$, thus satisfying the criteria for reference genes.

qPCR for C_q values and comparative delta C_t

qPCR was performed to compare the transcript abundance and stability of the candidate genes. C_q values for the eight remaining candidate reference genes across the five developmental stages and six species were determined using LightCycler 96 software (Roche, Mannheim, Germany). The average C_q values for these eight genes ranged from 14.77 (*EF1a*) to 16.95 (*GAPDH*), thus indicating their robust expression levels (Fig. 1). Among the eight candidate reference genes, *EF1α* exhibited the highest abundance across the six species, with average C_q values ranging from 14.48 in *An. lesteri* to 15.07 in *An. kleini*.

The gene with the lowest SD was determined to be the most stable in the delta C_t analysis, which varied across species: *RPL8* in *An. belenrae* (SD = 0.63) and *An. kleini* (SD = 0.74); *RPS7* in *An. lesteri* (SD = 0.52); *RPL32* in *An. pullus* (SD = 0.79); and *RPS17* in *An. sinensis* (SD = 0.76) and *An. sineroides* (SD = 0.63) (Fig. 2, Table 2). In the 4L and ovi-A stages, *RPL13a* was found to be the most stable gene (SD = 0.42 and SD = 0.25, respectively), while *RPL32* was the most stable gene in the other developmental stages (SD = 0.30–0.75) (Table 3).

geNorm

In each species, the use of two genes consistently exhibited the least pairwise variation when compared to using three or four genes. However, an exception was observed in *An. lesteri* and *An. sineroides*, where lower pairwise variation values were estimated for three genes (Fig. 3). When testing the different stages across six species, three genes were found to be superior to two genes, as indicated by their lower pairwise variations. However, it is worth noting that two genes were also acceptable, as their $V_{2/3}$ values were < 0.1. When assessing the M values of the eight candidate genes, the combination of *RPL32, RPL8, RPL13a,* and *RPS17* consistently exhibited lower values, indicating greater stability across all of the sample groups (Tables 2, 3, Fig. 4). Furthermore, the

Gene	F/R	Sequence (5′-3′)	Size (bp)	Efficiency (%)					
				An. belenrae	*An. kleini*	*An. lesteri*	*An. pullus*	*An. sinensis*	*An. sineroides*
actin	F	CTGAAGTACCCGATCGAGCA	90	99.6	101.5	94.3	96.8	97.4	98.1
	R	CACTCGCAGCTCGTTGTAGA							
**a-tub*	F	GCCAACCAGATGGTCAAGTG	116	93.9	83.6	84.2	94.8	95.5	82.4
	R	GTCTTGATGGTGGCGATGG							
EF1a	F	AGGCCGGTATCTCCAAGAAC	105	104.5	103.1	102.6	102.3	97.8	105.3
	R	CGGTCGAGTCCATCTTGTTC							
GAPDH	F	TGGTGAACGGCCAGAAGATC	161	102.4	105.1	101.9	101.6	98.5	95.4
	R	GATGATGACCTTCTTGGCGC							
RPL13a	F	CGGCCATTTCTTCCGCAACA	111	103.7	90.7	93.4	94.9	97.1	95.2
	R	CACAGCATACGGCTGGGAG							
RPL32	F	GCGGACGAAGAAGTTCATCC	113	101.8	99.0	94.8	96.5	96.4	95.4
	R	GCGGACGAAGAAGTTCATCC							
**RPL49*	F	ACCGGAGTGGAAGTACGTC	159	93.4	93.9	89.5	87.4	88.6	99.8
	R	AGATACACGGGCAGCATGTG							
RPL8	F	ATCGACAAGCCCATCCTGAAG	151	106.6	99.8	104.0	103.9	98.2	105.9
	R	CAGTCGAAGCCTTACCGATG							
RPS17	F	AGAGGTGGCCATCATTCCGA	104	95.7	100.0	92.4	97.5	96.2	98.8
	R	TAGAAATACCACGCACTTGCG							
RPS7	F	GTGCGCGAGCTTGAGAAGAA	115	103.0	100.2	92.4	92.7	94.9	97.3
	R	GTGGTCGCTTCTGCTTGT TG							

Table 1. Primers used for qPCR.

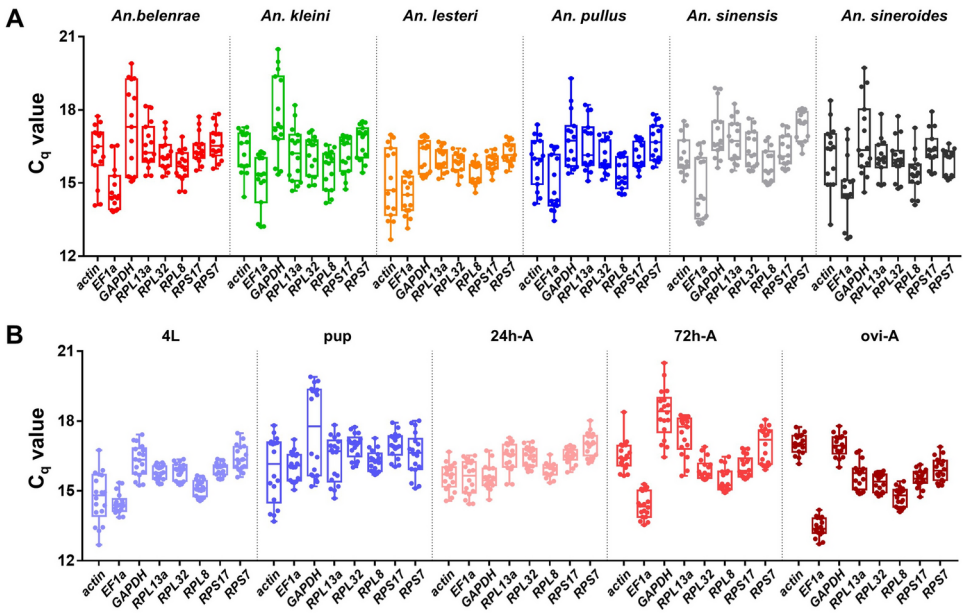

Fig. 1. C_q values of the eight candidate genes in the six *Anopheles* Hyrcanus Group species (**A**) and five developmental stages (**B**).

most stable genes with the lowest M values varied across species: *RPL32* and *RPS17* in *An. belenrae* (M = 0.1), *An. pullus* (M = 0.16), and *An. sineroides* (M = 0.16); *RPL32* and *RPL8* in *An. kleini* (M = 0.15) and *An. sinensis* (M = 0.1); and *RPL13a* and *RPS7* in *An. lesteri* (M = 0.08).

BestKeeper and NormFinder

BestKeeper analysis, based on the SD and CV values using pairwise correlation, identified that *RPS17* was the most stable gene for *An. belenrae*, *An. lesteri*, and *An. pullus*, while *RPS7* was found to be the most stable gene for *An. kleini*, *An. sinenesis*, and *An. sineroides* (Table 2). In addition, *RPS17* emerged as the best reference gene for the 4L and ovi-A stages across the six species. *RPL8* was determined to be the best reference gene for the 24 h-A

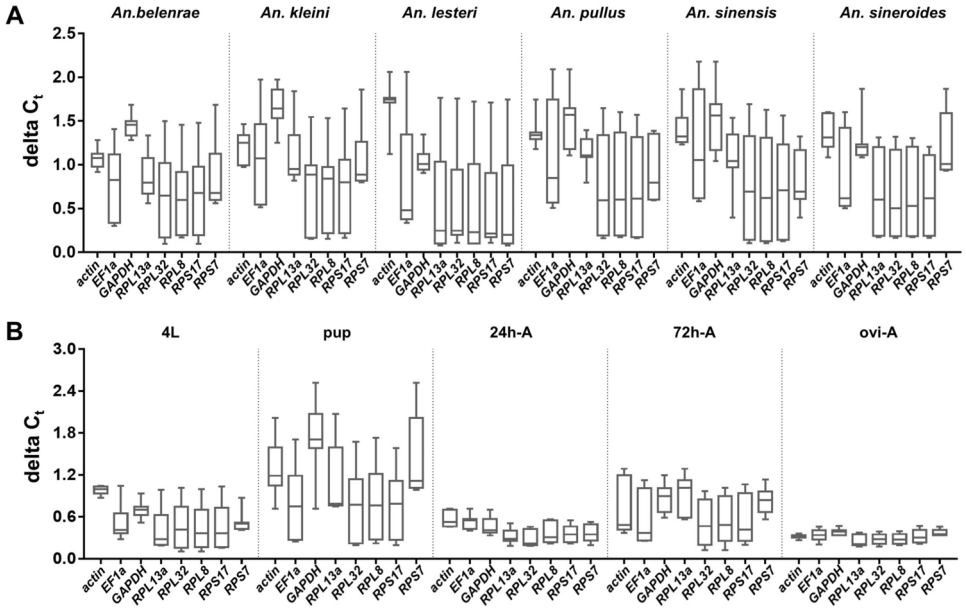

Fig. 2. Delta C_t values of the eight candidate genes in the six *Anopheles* Hyrcanus Group species (**A**) and five developmental stages (**B**).

and pup stages, while *RPL32* exhibited the highest stability as the reference gene for the 72 h-A in the *Anopheles* Hyrcanus Group (Table 3).

By NormFinder algorithm, the most stable genes for each species with the lowest SV, were determined as follows: *RPL8* in *An. belenrae* and *An. kleini* (SV = 0.31 and 0.26, respectively); *RPS17* in *An. lesteri*, *An. pullus*, and *An. sineroides* (SV = 0.26, 0.38, and 0.17, respectively); and *RPS7* in *An. sinensis* (SV = 0.33) (Table 2). When testing across the different developmental stages, results similar to those of the delta C_t method were obtained, where *RPL13a* emerged as the most stable gene in the 4L and ovi-A stages (SV = 0.23 and 0.11, respectively), while *RPL32* was the most stable gene in the 24 h-A and 72 h-A stages (SV = 0.07 and 0.24, respectively) (Table 3).

RefFinder

Using the results obtained from the four normalized methods described above, the overall ranking of gene stability was determined using RefFinder[23]. For *An. belenrae*, *An. pullus*, *An. sinensis*, and *An. sineroides*, *RPS17* was identified as the most stable reference gene, while *RPL8* and *RPS7* were identified as the most stable genes for *An. kleini* and *An. lesteri,* respectively. When using two reference genes, the following combinations were recommended: *RPS17* + *RPL8* for *An. belenrae* and *An. sinensis; RPL8* + *RPL32* for *An. kleini*; *RPS17* + *RPS7* for *An. lesteri;* and *RPS17* + *RPL32* for *An. pullus* and *An. sineroides*. When comparing the expression levels between species, *RPL32* was identified as the most stable gene for the pup and adult stages, while *RPL13a* was found to be the most stable gene for the 4L stage. For normalization with two reference genes, *RPL13a* + *RPL8* was recommended for the 4L stage, *RPL32* + *RPL8* for the pup and 72 h-A stages, and *RPL32* + *RPS17* or *RPL32* + *RPL13a* for the 24 h-A and ovi-A stages.

Validation of reference genes

Chitin synthase 2 (*CHS2)* and *vitellogenin 1* (*Vg1)* genes are known to exhibit distinct expression patterns throughout the mosquito life cycle[27,28]. To validate the selected reference genes, qPCR was performed, and the transcription levels of *CHS2* and *Vg1* were normalized using either the two most stable genes or the least stable gene and compared. When using *RPS7* and *RPS17*, which were identified as the most stable genes for *An. lesteri,* as references, the expression level of *CHS2* was determined to be highest in the 24 h-A stage (Fig. 5A). However, when normalized using the least stable gene, *actin*, a different pattern of *CHS2* emerged during the adult stages, with a 4.9-fold higher expression level in the ovi-A stage compared to the 24 h-A stage. The expression pattern of *Vg1* across stages varied depending on whether it was normalized to the most or least stable gene (Fig. 5B).

When comparing the expression levels of *CHS2* in the 24 h-A stage across the six species, *An. pullus* and *An. sinensis* exhibited the lowest expression levels when the stable reference genes (*RPL32* and *RPL13a*) were used for normalization. In contrast, *An. belenrae* showed the lowest expression level when an unstable gene (*EF1α*) was used as the reference (Fig. 5C). The expression level of *Vg1* also varied among the six species, but the order of expression level for each species was more consistent when using the *RPL32* and *RPL13a* genes (*An. pullus* < *An. kleini* < *An. sinensis* < *An. belenrae* < *An. sineroides* < *An. lesteri*) compared to using *EF1α* (*An. belenrae* < *An. pullus* < *An. sineroides* < *An. sinensis* < *An. kleini* < *An. lesteri*) (Fig. 5D). These results emphasize the importance of appropriate reference gene selection and confirm the reliability of the reference genes identified in this study.

		delta C_t		geNorm		BestKeeper		NormFinder		RefFinder
		Average SD	Rank	M	Rank	SD	Rank	SV	Rank	Overall ranking
An. belenrae	*actin*	1.08	7	0.68	7	0.94	7	0.61	7	7
	Ef1a	0.76	4	0.25	4	0.7	5	0.52	4	4
	GAPDH	1.45	8	0.87	8	1.61	8	1.17	8	8
	RPL13a	0.87	5	0.53	6	0.75	6	0.54	5	6
	RPL32	0.64	2	0.10	1	0.51	2	0.36	3	3
	RPL8	0.63	1	0.15	3	0.51	3	0.31	1	2
	RPS17	0.65	3	0.10	1	0.46	1	0.35	2	1
	RPS7	0.87	6	0.43	5	0.56	4	0.59	6	5
An. kleini	*actin*	1.19	7	0.84	7	0.74	5	0.81	6	7
	Ef1a	1.05	4	0.37	4	0.85	6	0.82	7	5
	GAPDH	1.66	8	1.05	8	1.40	8	1.20	8	8
	RPL13a	1.13	6	0.70	6	0.92	7	0.67	4	6
	RPL32	0.75	2	0.15	1	0.68	4	0.33	3	2
	RPL8	0.74	1	0.15	1	0.68	3	0.26	1	1
	RPS17	0.76	3	0.18	3	0.66	2	0.33	2	3
	RPS7	1.08	5	0.58	5	0.60	1	0.72	5	4
An. lesteri	*actin*	1.70	8	0.78	8	1.27	8	1.40	8	8
	Ef1a	0.78	6	0.25	6	0.68	7	0.70	6	6
	GAPDH	1.05	7	0.47	7	0.63	6	0.82	7	7
	RPL13a	0.54	3	0.08	1	0.42	5	0.32	3	4
	RPL32	0.57	5	0.17	5	0.32	2	0.32	5	5
	RPL8	0.53	2	0.09	3	0.40	4	0.32	3	3
	RPS17	0.54	4	0.15	4	0.27	1	0.26	1	2
	RPS7	0.52	1	0.08	1	0.37	3	0.30	2	1
An. pullus	*actin*	1.37	7	0.89	7	0.82	5	0.98	7	7
	Ef1a	1.11	5	0.37	4	1	8	0.84	6	5
	GAPDH	1.51	8	1.05	8	0.91	6	1.24	8	8
	RPL13a	1.13	6	0.69	6	0.92	7	0.65	5	6
	RPL32	0.79	1	0.16	1	0.57	3	0.42	2	2
	RPL8	0.8	3	0.17	3	0.54	2	0.45	4	3
	RPS17	0.79	2	0.16	1	0.47	1	0.38	1	1
	RPS7	0.89	4	0.49	5	0.67	4	0.44	3	4
An. sinensis	*actin*	1.41	7	0.89	7	0.62	4	1.10	7	7
	Ef1a	1.19	6	0.38	4	1.17	8	0.90	6	6
	GAPDH	1.51	8	1.04	8	0.87	7	1.17	8	8
	RPL13a	1.05	5	0.67	6	0.74	6	0.55	5	5
	RPL32	0.8	3	0.10	1	0.65	5	0.45	4	4
	RPL8	0.77	2	0.10	1	0.61	3	0.39	3	2
	RPS17	0.76	1	0.13	3	0.54	2	0.35	2	1
	RPS7	0.84	4	0.53	5	0.53	1	0.33	1	3
An. sineroides	*actin*	1.35	8	0.92	8	1.04	7	1.00	8	8
	Ef1a	0.93	5	0.34	5	0.95	6	0.65	5	6
	GAPDH	1.27	7	0.78	7	1.29	8	0.98	7	7
	RPL13a	0.66	4	0.18	3	0.55	3	0.36	4	3
	RPL32	0.65	2	0.16	1	0.61	5	0.29	2	2
	RPL8	0.65	3	0.18	4	0.59	4	0.32	3	4
	RPS17	0.63	1	0.16	1	0.54	2	0.17	1	1
	RPS7	1.25	6	0.58	6	0.50	1	0.84	6	5

Table 2. Ranking of gene stability across the six species was determined using various algorithms.

Discussion

When comparing the gene expression levels among the different species, it is essential to consider that even within the same genus, substantial variation may exist within the nucleotide sequences of the same orthologous target or reference genes (Supplementary table 3). Therefore, precise primer design is essential to avoid errors resulting from sequence polymorphisms at the priming sites between species[29]. Differences in the sequence

		delta C_t		geNorm		BestKeeper		NormFinder		RefFinder
		Average SD	Rank	M	Rank	SD	Rank	SV	Rank	Overall ranking
4L	*actin*	0.98	8	0.56	8	0.86	8	0.69	8	8
	Ef1a	0.51	5	0.25	5	0.33	5	0.33	6	5
	GAPDH	0.70	7	0.42	7	0.56	7	0.54	7	7
	RPL13a	0.42	1	0.17	4	0.27	2	0.23	1	1
	RPL32	0.45	4	0.10	1	0.30	4	0.27	4	4
	RPL8	0.42	2	0.10	1	0.29	3	0.25	2	2
	RPS17	0.45	3	0.14	3	0.23	1	0.31	5	3
	RPS7	0.53	6	0.32	6	0.44	6	0.26	3	6
pup	*actin*	1.27	6	0.86	7	1.25	7	1.06	6	6
	Ef1a	0.80	4	0.25	4	0.48	4	0.40	4	4
	GAPDH	1.71	8	1.08	8	1.81	8	1.50	8	8
	RPL13a	1.12	5	0.46	5	0.66	5	0.74	5	5
	RPL32	0.75	1	0.19	1	0.42	2	0.25	2	1
	RPL8	0.78	3	0.23	3	0.35	1	0.39	3	2
	RPS17	0.75	2	0.19	1	0.43	3	0.17	1	3
	RPS7	1.41	7	0.66	6	0.72	6	1.13	7	7
24 h-A	*actin*	0.56	8	0.41	8	0.45	7	0.43	8	7
	Ef1a	0.53	7	0.36	7	0.65	8	0.38	7	8
	GAPDH	0.46	6	0.30	6	0.39	5	0.34	6	6
	RPL13a	0.32	2	0.18	1	0.35	3	0.16	2	2
	RPL32	0.30	1	0.18	1	0.35	4	0.07	1	1
	RPL8	0.35	3	0.22	3	0.25	1	0.20	3	3
	RPS17	0.36	5	0.23	4	0.25	2	0.22	4	4
	RPS7	0.36	4	0.25	5	0.39	6	0.23	5	5
72 h-A	*actin*	0.76	5	0.31	5	0.48	5	0.60	6	5
	Ef1a	0.61	4	0.23	4	0.45	4	0.41	4	4
	GAPDH	0.88	7	0.63	7	0.84	7	0.66	7	6
	RPL13a	0.94	8	0.71	8	0.90	8	0.72	8	8
	RPL32	0.52	1	0.12	1	0.37	1	0.24	1	1
	RPL8	0.55	2	0.12	1	0.43	3	0.31	2	2
	RPS17	0.57	3	0.19	3	0.40	2	0.35	3	3
	RPS7	0.83	6	0.51	6	0.69	6	0.60	5	7
ovi-A	*actin*	0.32	5	0.26	6	0.32	3	0.20	5	5
	Ef1a	0.34	6	0.24	5	0.32	2	0.22	6	6
	GAPDH	0.39	8	0.32	8	0.39	7	0.28	8	8
	RPL13a	0.25	1	0.17	1	0.33	4	0.11	1	2
	RPL32	0.28	2	0.17	1	0.36	6	0.15	3	1
	RPL8	0.28	3	0.19	3	0.35	5	0.14	2	4
	RPS17	0.31	4	0.21	4	0.29	1	0.20	4	3
	RPS7	0.37	7	0.29	7	0.42	8	0.23	7	7

Table 3. Ranking of gene stability across the five developmental stages determined using various algorithms.

can lead to variations in the efficiency of the primer set, despite designing the qPCR primers from conserved regions. In this study, certain primer sets, such as those for *α-tub* and *RPL49*, exhibited considerably lower average amplification efficiencies, possibly because of imperfect sequence matching at the priming sites, which resulted in their exclusion from further testing. The performance of primer sets for the remaining reference genes appeared to be acceptable, as indicated by their higher average amplification efficiencies.

In some *Anopheles* species, such as *An. gambiae*, *RPS7* has traditionally been used as a reference gene for normalization in transcriptional studies[30]. However, it has been observed that the expression level of *RPS7* fluctuates across the developmental stages of *An. gambiae*[31]. As gene stability varies among species and experimental conditions[10], the indiscriminate use of *RPS7* as a reference candidate can introduce bias when interpreting the expression results. For example, in this study, *RPS7* was the most stable gene in *An. lesteri*, but its stability notably decreased in *An. belenrae* and *An. sineroides*, thus indicating significant cross-species variation in *RPS7* expression stability. Moreover, pronounced differences in expression levels among the six species were observed, making it unsuitable for cross-species comparisons.

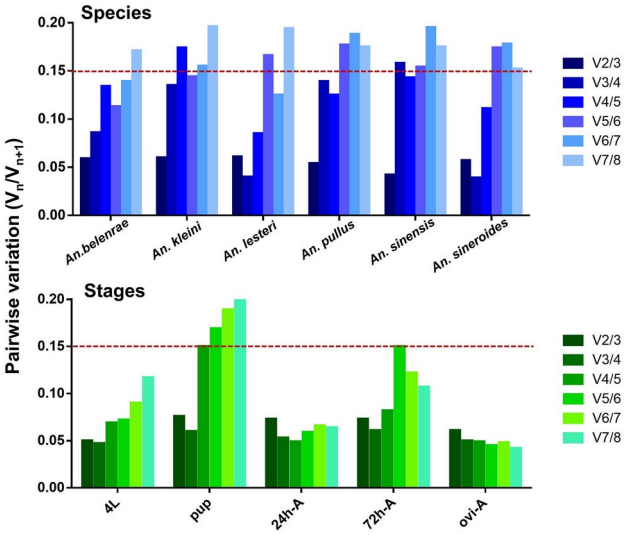

Fig. 3. Pairwise variation values of the six *Anopheles* Hyrcanus Group species and five developmental stages with a 0.15 cutoff value.

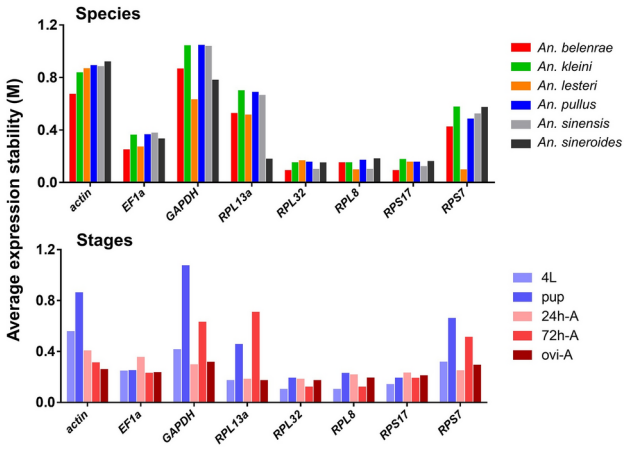

Fig. 4. geNorm-based average expression stability (M) values of the six *Anopheles* Hyrcanus Group species and five developmental stages.

The most stable genes across the six species and five developmental species were those that encoded ribosomal proteins, which are essential components of the large and small subunits. Ribosomal protein genes are widely recognized for their high sequence conservation across organisms, which makes them excellent candidates as reference genes in many insect species. They have been validated as optimum reference genes across diverse experimental conditions, including developmental stages. For example, in *Cimex hemipterus*, *RPL8* was the most stable gene across many different experimental conditions (developmental stages, sex, tissue, gas stimulation, and temperature)[32]. Similarly, *RPL32* in *Frankliniella occidentalis* showed remarkable stability in different developmental stages and temperature conditions[33]. In *Helopeltis theivora, RPL13a* was identified as the most stable gene across the different developmental stages and sexes[34], while the expression of *RPS17* remained stable throughout the entire life stage of *Aedes aegypti*[25]. In contrast, despite being selected as the most stable genes under various experimental conditions for many other insect species[13], genes such as *actin*, *EF1α*, and *GAPDH* exhibited unstable expression patterns among the *Anopheles* Hyrcanus Group species and developmental stages.

In conclusion, this study successfully identified reliable reference genes for the normalization of target gene expression across the five developmental stages in six species in the *Anopheles* Hyrcanus Group distributed in South Korea. The use of two reference genes (*RPL13a* + *RPL8*) is recommended when comparing the expression profiles of genes related to important larval physiology, including detoxification metabolism against various larvicidal agents and immune reactions to aquatic microbes. For the cross-species comparison of diverse physiological aspects of female mosquitoes, including vector competence, pathogen immune responses, and insecticide susceptibility, the simultaneous use of these genes (either *RPL32* + *RPS17* or *RPL32* + *RPL13a*) as reference standards will be recommended for future studies. While these reference genes are valuable for

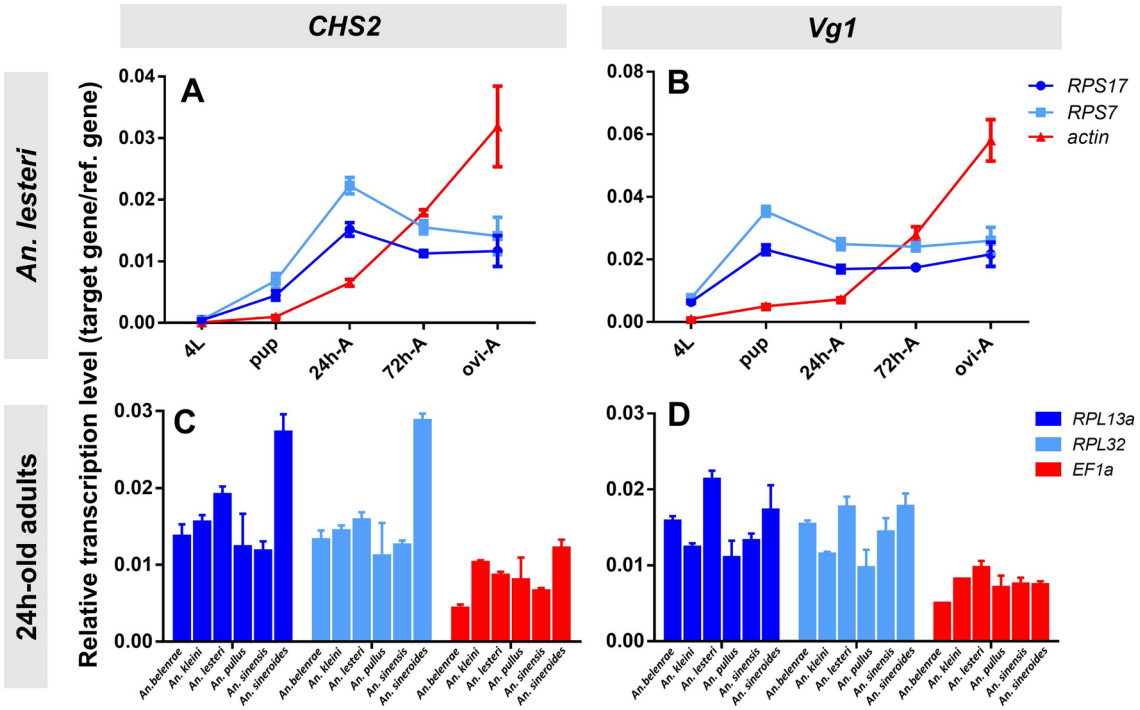

Fig. 5. Comparison of relative transcription levels of the two target genes, *CHS2* and *Vg1*, across the five developmental stages of *Anopheles lesteri* (**A**, **B**) and 24 h-old adult females (24 h-A) of the six *Anopheles* Hyrcanus Group species (**C**, **D**). Values were normalized using the best (blue) or worst (red) reference genes selected in this study.

comparing various gene expression profiles across different species or different developmental stages, further validation is required to assess their suitability for other experimental conditions, such as tissue-specific expression profiling, investigations involving varying temperatures, and the evaluation of infection dynamics. This study only focused on six mosquito species found within the same region; therefore, further investigation is required to determine whether these results apply to other species within the *Anopheles* Hyrcanus Group, which could broaden the utility of our findings for additional research. Nevertheless, these validated reference genes are essential for advancing research on mosquito biology and vector control and enabling a deeper understanding of mosquito behavior and physiology.

Data availability

All data generated during this study are included in this published article and its Supplementary Information files. The associated sequences are available on GenBank (http://www.ncbi.nlm.nuh.gov/genbank; OQ581089—OQ581147, OQ122132—OQ122143, OQ984005 -OQ984034).

Received: 11 October 2024; Accepted: 23 December 2024
Published online: 16 March 2025

References

1. Harbach, R. E. in *Anopheles mosquitoes-New insights into malaria vectors* (IntechOpen, 2013).
2. Zhang, C. et al. Survey of asymptomatic malaria and mosquito vectors in Muang Khua district of Phongsaly province, China-Laos border. *Int. J. Infect Dis.* **96**, 141–147. https://doi.org/10.1016/j.ijid.2020.03.066 (2020).
3. Poncon, N. et al. Biology and dynamics of potential malaria vectors in Southern France. *Malar. J.* **6**, 18. https://doi.org/10.1186/1475-2875-6-18 (2007).
4. Lee, W. J. et al. *Anopheles kleini*, *Anopheles pullus*, and *Anopheles sinensis*: potential vectors of *Plasmodium vivax* in the Republic of Korea. *J. Med. Entomol.* **44**, 1086–1090. https://doi.org/10.1603/0022-2585(2007)44[1086:akapaa]2.0.co;2 (2007).
5. Zhang, S. et al. *Anopheles* vectors in mainland China while approaching malaria elimination. *Trends Parasitol.* **33**, 889–900. https://doi.org/10.1016/j.pt.2017.06.010 (2017).
6. Kim, H. C. et al. Distribution and larval habitats of *Anopheles* species in northern Gyeonggi Province, Republic of Korea. *J. Vector Ecol.* **36**, 124–134. https://doi.org/10.1111/j.1948-7134.2011.00149.x (2011).
7. Rueda, L. M. Two new species of *Anopheles* (*Anopheles*) Hyrcanus Group (Diptera: Culicidae) from the Republic of South Korea. *Zootaxa* **941**, 1–26 (2005).
8. Ubalee, R. et al. Vector competence and the susceptibility of *Anopheles pullus* and *Anopheles belenrae* to *Plasmodium vivax*-infected blood from Thai patients. *J. Med. Entomol.* **59**, 1047–1052. https://doi.org/10.1093/jme/tjab226 (2022).
9. Zhu, G. et al. Transcriptome profiling of pyrethroid resistant and susceptible mosquitoes in the malaria vector, *Anopheles sinensis*. *BMC Genomics* **15**, 448. https://doi.org/10.1186/1471-2164-15-448 (2014).
10. Teng, X., Zhang, Z., He, G., Yang, L. & Li, F. Validation of reference genes for quantitative expression analysis by real-time rt-PCR in four Lepidopteran insects. *J. Insect Sci.* **12**, 60. https://doi.org/10.1673/031.012.6001 (2012).

11. Shakeel, M., Rodriguez, A., Tahir, U. B. & Jin, F. Gene expression studies of reference genes for quantitative real-time PCR: an overview in insects. *Biotechnol. Lett.* **40**, 227–236. https://doi.org/10.1007/s10529-017-2465-4 (2018).
12. Nardini, L. et al. Malaria vectors in the Democratic Republic of the Congo: the mechanisms that confer insecticide resistance in *Anopheles gambiae* and *Anopheles funestus*. *Malar. J.* **16**, 448. https://doi.org/10.1186/s12936-017-2099-y (2017).
13. Lu, 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. https://doi.org/10.3389/fphys.2018.01560 (2018).
14. Muller, P. et al. Field-caught permethrin-resistant *Anopheles gambiae* overexpress CYP6P3, a P450 that metabolises pyrethroids. *PLoS Genet.* **4**, e1000286. https://doi.org/10.1371/journal.pgen.1000286 (2008).
15. Liu, C. et al. Distinct olfactory signaling mechanisms in the malaria vector mosquito *Anopheles gambiae*. *PLoS Biol.* https://doi.org/10.1371/journal.pbio.1000467 (2010).
16. Yan, Z. W. et al. Genome-wide and expression-profiling analyses suggest the main cytochrome P450 genes related to pyrethroid resistance in the malaria vector, *Anopheles sinensis* (Diptera Culicidae). *Pest Manag. Sci.* **74**, 1810–1820. https://doi.org/10.1002/ps.4879 (2018).
17. de Oliveira, L. A. et al. Reference genes for the normalization of gene expression in eucalyptus species. *Plant Cell Physiol.* **53**, 405–422. https://doi.org/10.1093/pcp/pcr187 (2012).
18. Olias, P., Adam, I., Meyer, A., Scharff, C. & Gruber, A. D. Reference genes for quantitative gene expression studies in multiple avian species. *Plos ONE* https://doi.org/10.1371/journal.pone.0099678 (2014).
19. Moura, M. T. et al. Inter-genus gene expression analysis in livestock fibroblasts using reference gene validation based upon a multi-species primer set. *Plos ONE* https://doi.org/10.1371/journal.pone.0221170 (2019).
20. Vandesompele, J. et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* https://doi.org/10.1186/gb-2002-3-7-research0034 (2002).
21. 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. https://doi.org/10.1023/b:bile.0000019559.84305.47 (2004).
22. Andersen, C. L., Jensen, J. L. & Orntoft, 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. https://doi.org/10.1158/0008-5472.CAN-04-0496 (2004).
23. 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.* https://doi.org/10.1007/s11103-012-9885-2 (2012).
24. Li, C. et al. A newly recognized species in the *Anopheles* Hyrcanus Group and molecular identification of related species from the Republic of South Korea (Diptera: Culicidae). *Zootaxa* **939**, 1–8 (2005).
25. Dzaki, N., Ramli, K. N., Azlan, A., Ishak, I. H. & Azzam, G. Evaluation of reference genes at different developmental stages for quantitative real-time PCR in *Aedes aegypti*. *Sci. Rep.* **7**, 43618. https://doi.org/10.1038/srep43618 (2017).
26. 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. https://doi.org/10.1186/1471-2199-7-33 (2006).
27. Zhang, X., Zhang, J., Park, Y. & Zhu, K. Y. Identification and characterization of two chitin synthase genes in African malaria mosquito, *Anopheles gambiae*. *Insect Biochem. Mol. Biol.* **42**, 674–682. https://doi.org/10.1016/j.ibmb.2012.05.005 (2012).
28. Provost-Javier, K. N., Chen, S. & Rasgon, J. L. Vitellogenin gene expression in autogenous *Culex tarsalis*. *Insect Mol. Biol.* **19**, 423–429. https://doi.org/10.1111/j.1365-2583.2010.00999.x (2010).
29. Boyle, B., Dallaire, N. & MacKay, J. Evaluation of the impact of single nucleotide polymorphisms and primer mismatches on quantitative PCR. *Bmc Biotechnol.* **9**, 75. https://doi.org/10.1186/1472-6750-9-75 (2009).
30. Vizioli, J. et al. Gambicin: a novel immune responsive antimicrobial peptide from the malaria vector *Anopheles gambiae*. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 12630–12635. https://doi.org/10.1073/pnas.221466798 (2001).
31. Togawa, T., Dunn, W. A., Emmons, A. C., Nagao, J. & Willis, J. H. Developmental expression patterns of cuticular protein genes with the R&R consensus from *Anopheles gambiae*. *Insect Biochem. Mol. Biol.* **38**, 508–519. https://doi.org/10.1016/j.ibmb.2007.12.008 (2008).
32. Kong, D. et al. Identification and validation of reference genes for expression analysis using qRT-PCR in *Cimex hemipterus* (Hemiptera: Cimicidae). *Insects* https://doi.org/10.3390/insects13090784 (2022).
33. Zheng, Y. T., Li, H. B., Lu, M. X. & Du, Y. Z. Evaluation and validation of reference genes for qRT-PCR normalization in *Frankliniella occidentalis* (Thysanoptera: Thripidae). *PLoS ONE* **9**, e111369. https://doi.org/10.1371/journal.pone.0111369 (2014).
34. Wang, Z. et al. Evaluation and validation of reference genes for quantitative real-time PCR in *Helopeltis theivora* Waterhouse (Hemiptera: Miridae). *Sci. Rep.* **9**, 13291. https://doi.org/10.1038/s41598-019-49479-1 (2019).

Acknowledgements

We would like to thank Xuerong Jin for the technical assistance with some experiments.

Author contributions

DEL conducted the experiments, analyzed the data, and wrote the manuscript. SHL and JHK designed and coordinated the project and reviewed the manuscript. All authors have approved the final version of the manuscript.

Funding

This work was supported by the Research Resettlement Fund for the new faculty of Seoul National University.

Declarations

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/s41598-024-84361-9.

Correspondence and requests for materials should be addressed to J.H.K.