Data in Brief 53 (2024) 110221

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Dataset for selection of stable reference genes for accurate quantitative gene expression analysis in silvertip tetra (*Hasemania nana*): Implications for sex differentiation and determination

Hwa Jin Lee[a,c], Ki Tae Kim[b], Min Sun Kim[c], Sang Yoon Lee[a,*]

[a] *CellQua, Inc, Seongnam 13595, South Korea*
[b] *Southeast Sea Fisheries Research Institute, NFRDI, 398-68, Sanyangilju-ro, Sangyang-Up Tongyeong, Gyoengnam 650-943, Korea*
[c] *Department of Biological Sciences, Kongju National University, Gongju 32588, South Korea*

ARTICLE INFO

Article history:
Received 21 September 2023
Revised 18 January 2024
Accepted 14 February 2024
Available online 20 February 2024

Dataset link: PRJNA1017117 (Original data)

Keywords:
Silvertip tetra
RT-qPCR
Reference genes
Gene expression

ABSTRACT

Real-time quantitative PCR (RT-qPCR) is a widely used method for accurate quantitative gene expression analysis. For accurate quantitative verification of RT-qPCR, it is essential to select a reference gene with high expression stability depending on the experimental environment or the different tissues. In this study, we evaluated the stability of nine candidate reference genes, labeled elongation factor 1-alpha (EF1A), ERBB receptor feedback inhibitor 1-like isoform x2 (ERRFI), glyceraldehyde-3-phosphate dehydrogenase (GAPDH), integrin beta2 like (ITGB2), phosphatidylinositol-binding clathrin assembly protein-like isoform x3 (PICALM), 60 s ribosomal protein L5 (RPL5), 60 s ribosomal protein L7 (RPL7), tubulin beta chain (TUBB), and ubiquitin-conjugating enzyme E2A (UBE2A), in the brain (including pituitary gland) gonads and caudal fins of silvertip tetra (*Hasemania nana*) males and females. The stability evaluation of the reference gene was analyzed using a program based on the geNorm, NormFinder, BestKeeper, and RankAggreg algorithms. As a result, RPL5 (brain, caudal fin), EF1A (gonad), and PICALM

* Corresponding author.
E-mail address: sangyoon0423@cellqua.io (S.Y. Lee).

https://doi.org/10.1016/j.dib.2024.110221

(three tissue types) genes were evaluated as the most stable genes in silvertip tetra females. In males, TUBB (brain, caudal fin) and ITGB2 (gonads, three tissue types) genes were the most stable, and in both sexes, TUBB (brain), ITGB2 (caudal fin), RPL5 (gonads), and PICALM (three tissue types) genes are considered appropriate as reference genes for qRT-PCR analysis. However, the GAPDH gene was judged to be inappropriate for use as a reference gene because gene stability in the brain, caudal fin, and gonads was evaluated to be low in all males and females. As an introductory study on silvertip tetra, a new research model fish, the results of this study are expected to provide helpful information regarding sex differentiation and determination in fish.

Specifications Table

Subject	Biological sciences
Specific subject area	Molecular Biology
Type of data	Table, Figure
How data were acquired	Data were obtained from Real-time PCR using a Thermal Cycler Dice ™ real-time PCR system for reference genes.
Data format	Raw, Analyzed
Description of data collection	The nine candidate reference genes were selected and analyzed using a program based on the geNorm, NormFinder, BestKeeper, and RankAggreg algorithms in the brain and pituitary gland, gonads, and caudal fins of silvertip tetra (*Hasemania nana*).
Data source location	Institution: Department of Marine Biotechnology, Gangneung-Wonju National University City: Gangneung-si Country: Republic of Korea Latitude and longitude: 37.770748 / 128.866253
Data accessibility	Repository name: BioProject Data identification number: PRJNA1017117 Direct URL to data: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1017117

1. Value of the Data

- The silvertip tetra (*Hasemania nana*) is a small freshwater fish. The silvertip tetra can distinguish males and females by their body color. Therefore, the silvertip tetra has been identified as a suitable research model for sex differentiation and determination.
- This study provides suitable reference genes for the brain, caudal fin, and gonadal tissues in the silvertip tetra.
- It is essential to evaluate the selection of suitable reference genes according to different experimental conditions and tissue types and verify their stability through various programs such as geNorm, BestKeeper, NormFinder, and RankAggreg.
- This dataset represents the initial reference gene data for the silvertip tetra and is valuable for researching sex differentiation and determination in fish.
- The results of this data are expected to be useful for baseline RT-qPCR-based expression analysis in the new experimental model fish, the silvertip tetra, and other fish.

2. Data Description

The following nine genes [elongation factor 1-alpha (EF1A), ERBB receptor feedback inhibitor 1-like isoform x2 (ERRFI), glyceraldehyde-3-phosphate dehydrogenase (GAPDH), integrin beta2-like (ITGB2), phosphatidylinositol-binding clathrin assembly protein-like isoform x3 (PICALM), 60 s ribosomal protein L5 (RPL5), 60 s ribosomal protein L7 (RPL7), tubulin beta chain (TUBB), and ubiquitin-conjugating enzyme E2A (UBE2A)] were selected as suitable reference gene. The stability of reference genes was assessed by analyzing gene expression in the brain, gonads, and caudal fin tissues of females and males in silvertip tetra. The PCR efficiencies of the nine reference genes were 93–107%, confirming that the appropriate PCR efficiency conditions for RT-qPCR analysis were met (Table 1). The nine reference genes were analyzed using various programs such as geNorm, BestKeeper, NormFinder, and RankAggreg.

2.1. The stability of candidate reference gene by geNorm

The results of the ranking order determined using geNorm program were presented in Fig. 1. The most suitable reference genes in the brain tissue of the silvertip tetra were presented in Fig. 1A. In B and C, we presented the results for the candidate reference genes in the caudal fin and gonad, respectively. When we evaluated the overall expression stability by integrating the Ct values of the brain, gonad, and caudal fin, ITGB2, and PICALM were the most stable genes in females, while GAPDH was the least stable. UBE2A and ITGB2 were the most stable in males, while TUBB was the least stable. Considering both sexes, ITGB2 and PICALM had the highest expression stability, while GAPDH had the lowest (Fig. 1D).

2.2. The stability of candidate reference gene by normfinder

The results of the candidate reference gene using NormFinder program were presented Fig. 2. The ranking orders in the brain, caudal fin, and gonad of the silvertip tetra through NormFinder were presented in Fig. 2A, B, and C, respectively. In Fig. 2D, UBE2A had the highest expression stability, and GAPDH had the lowest. In males, UBE2A and ITGB2 were the most stable genes, while TUBB was the least stable gene. Across all male and female data, UBE2A was the most stable gene and GAPDH was the least stable.

2.3. The stability of candidate reference gene by bestkeeper

The most stable reference gene and unstable gene in the brain, caudal fin, and gonad of the silvertip tetra were represented in Fig. 3A, B, and C. The combined brain, gonad, and caudal fin tissue data for both males and females were presented in Fig. 3D. TUBB genes were found to be most stable in females, while GAPDH was found to be least stable. ITGB2 was found to be most stable in males, while TUBB was found to be least stable. Across all male and female data, the PICALM gene showed the highest expression stability, while the GAPDH gene showed the lowest.

2.4. The stability of candidate reference gene by RankAggreg

Ranking the expression stability based on the geometric mean of the target genes in the brain showed that RPL5 was the most stable gene in females and TUBB in males, and TUBB was the most stable gene in the mixed male-female sample (Table 2). In Fig. 4A, the least stable gene in the combined male and female data was GAPDH. In the caudal fin, the most stable genes were RPL5 in females, TUBB in males, and ITGB2 in all male and female data. In contrast, the least

Table 1
Summary of candidate reference genes and primer sequences for RT-qPCR.

Gene symbol	Gene description	Primer labels	Primer sequences (5′ - 3′)	PCR efficiency (%)
EF1A	Elongation factor 1-alpha	qHNA_EF1A1_FW1	TGGCAAAGTGACCAAGTCTG	96
		qHNA_EF1A1_RV1	GACAATTCCCTCCTTTCCTG	
ERRFl	ERBB receptor feedback inhibitor 1-like isoform x2	qHNA_ERRFl1_FW1	TCAAAGAGTCTGGCTCAGGA	106.27
		qHNA_ERRFl1_RV1	TGTTGAGGTGGCACTGTGTA	
GAPDH	Glyceraldehyde 3 phosphate dehydrogenase	qHNA_GAPDH_FW1	TGATGCTGGTGCAGGAATTG	93.57
		qHNA_GAPDH_RV1	CATGGCACTTCAAACAGTT	
ITGB2	Integrin beta2 like	qHNA_ITGB2_FW1	GCTGTGGCTTTGATTGGACT	102.53
		qHNA_ITGB2_RV1	GTTCCATCGGTGCTTTAGGA	
PICALM	Phosphatidylinositol-binding clathrin assembly protein-like isoform x3	qHNA_PICALM_FW1	CACTATTCCAGCGCACAGTCA	100.08
		qHNA_PICALM_RV1	TTTGGACTGCCAGTTAGAGC	
RPL5	60s ribosomal protein L5	qHNA_RPL5_FW1	ACCCAGTGCACGAAAAGAAG	100.46
		qHNA_RPL5_RV1	TAATGCGGGCAAGAAGAGTC	
RPL7	60s ribosomal protein L7	qHNA_RPL7_FW1	AACCGTGAGGACCAGATCAA	98.53
		qHNA_RPL7_RV1	TTTCTGGCCACTAGGTGGAA	
TUBB	Tubulin beta chain	qHNA_TUBB_FW1	AGTGGATCCCAAACAACGTC	104.28
		qHNA_TUBB_RV1	TGTGAACTCCATCTCGTCCA	
UBE2A	Ubiquitin-conjugating enzyme E2A	qHNA_UBE2A_FW1	GATGAACCGAACCCAAACAG	95
		qHNA_UBE2A_RV1	GAGTTTGAACATTGCCTGGC	

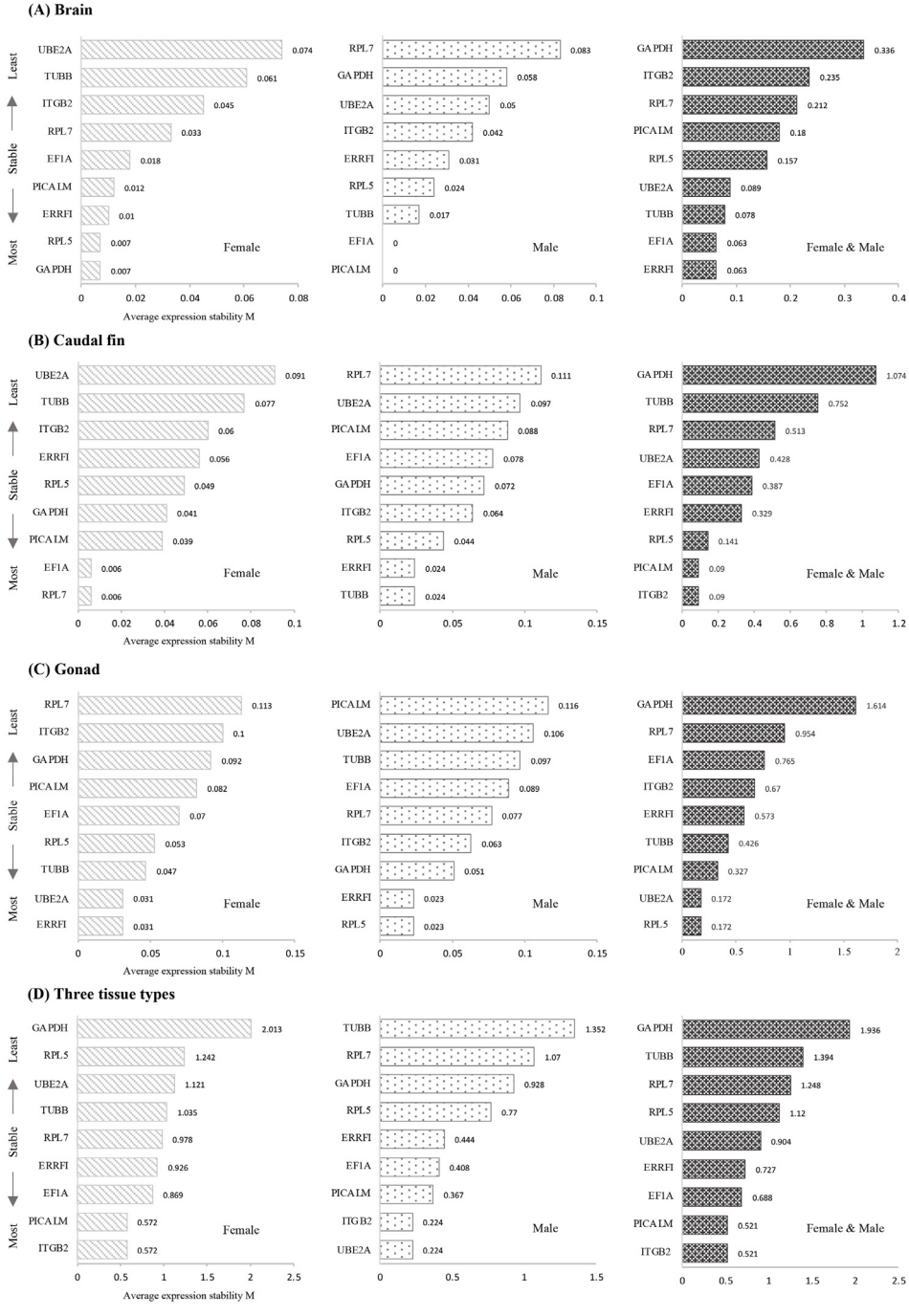


Fig. 1. Mean expression stability values (M) and ranking orders of candidate reference genes using geNorm. A lower value of average expression stability (M) indicates a more stable expression. The samples are (A) brain, (B) caudal fin, (C) gonads and (D) three tissue types.

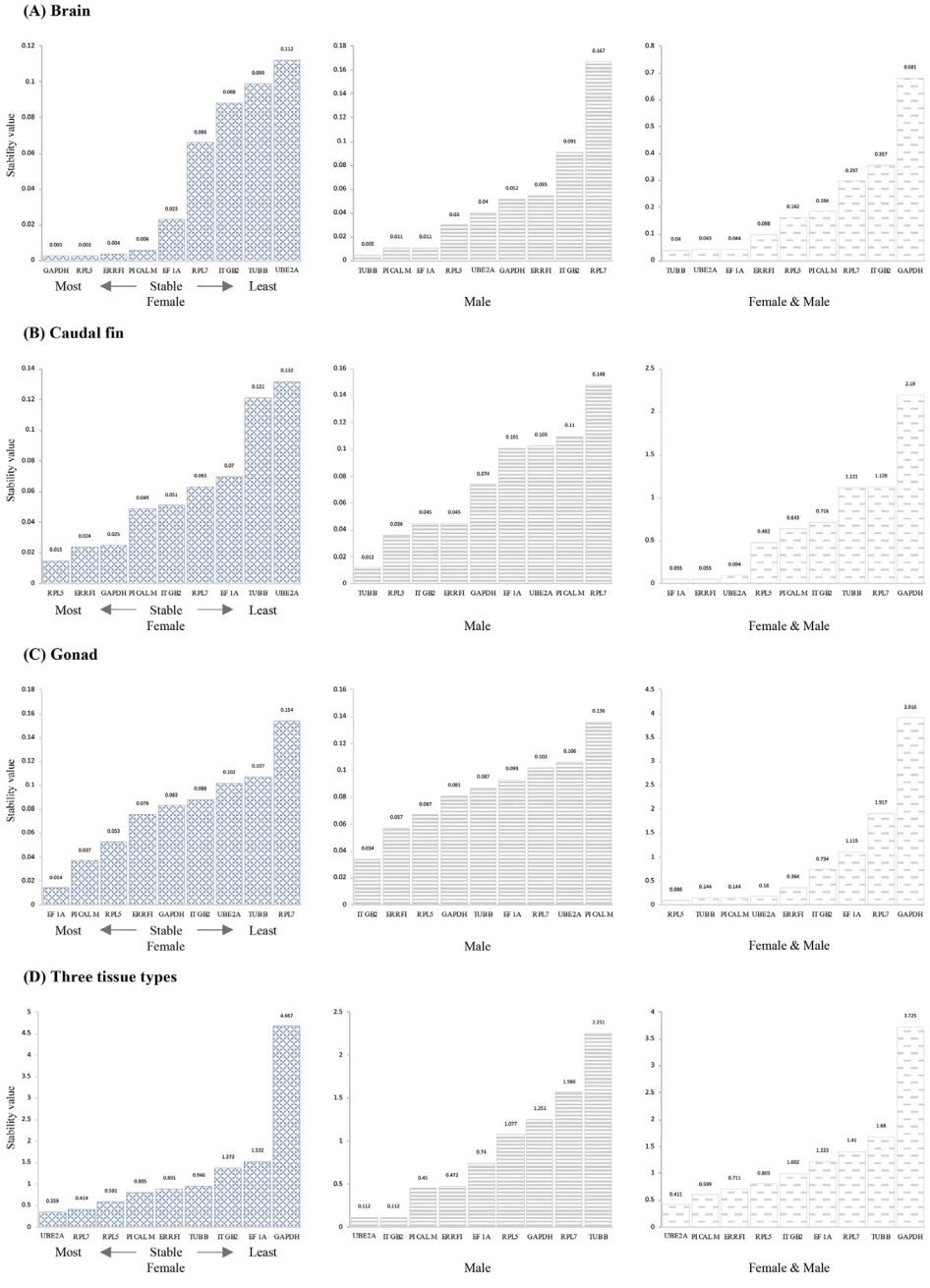


Fig. 2. Expression stability of candidate reference genes calculated by NormFinder. A lower value indicates a more stable expression. The samples are (A) brain, (B) caudal fin, (C gonads, and (D three tissue types.

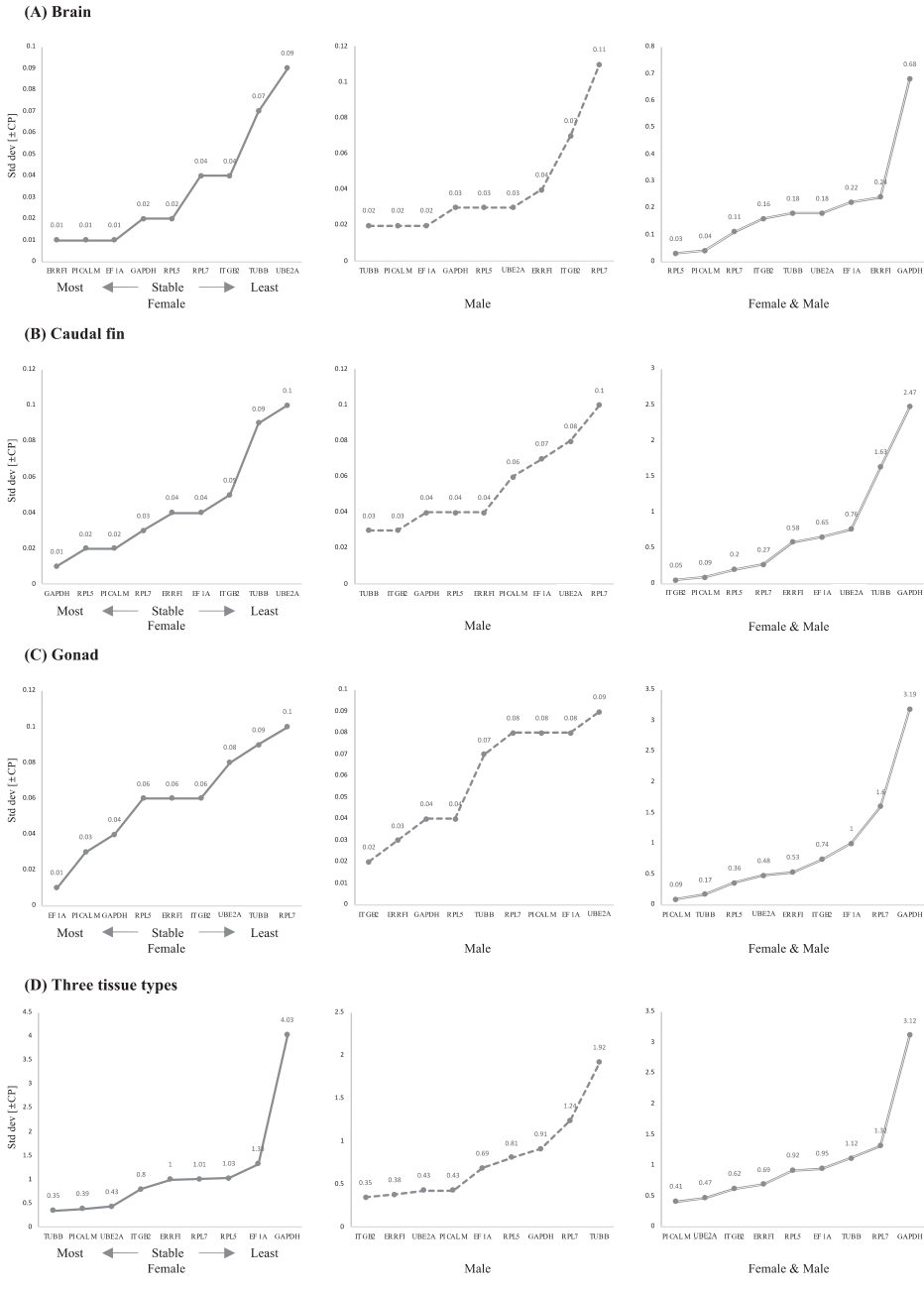
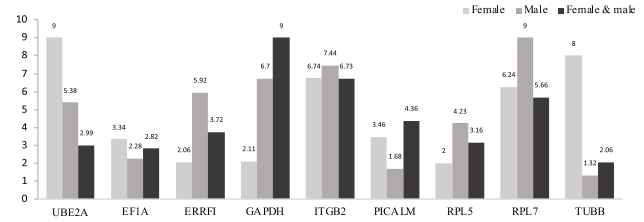
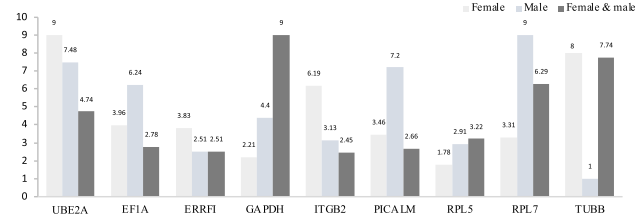


Fig. 3. Expression stability of candidate reference genes calculated by BestKeeper. A lower value indicates more stable expression. The samples refer to (A) brain, (B) caudal fin, (C) gonads, and (D) three tissue types.



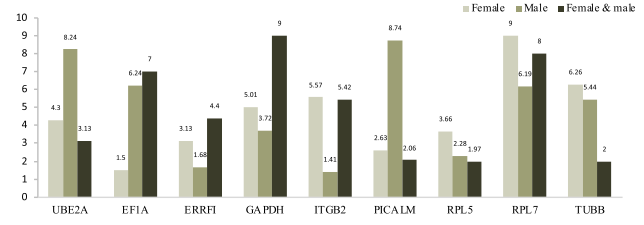
Female	RPL5	ERRFI	GAPDH	EF1A	PICALM	RPL7	ITGB2	TUBB	UBE2A
Male	TUBB	PICALM	EF1A	RPL5	UBE2A	ERRFI	GAPDH	ITGB2	RPL7
Female & male	TUBB	EF1A	UBE2A	RPL5	ERRFI	PICALM	RPL7	ITGB2	GAPDH
	Most	←			Stable		→		Least

(A)



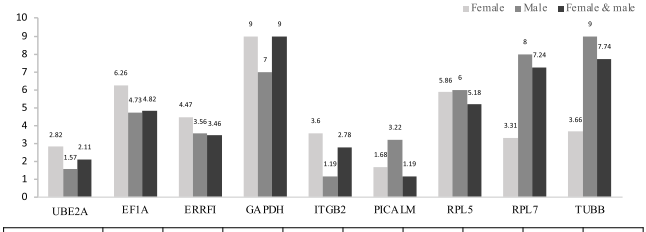
Female	RPL5	GAPDH	RPL7	PICALM	ERRFI	EF1A	ITGB2	TUBB	UBE2A
Male	TUBB	ERRFI	RPL5	ITGB2	GAPDH	EF1A	PICALM	UBE2A	RPL7
Female & male	ITGB2	ERRFI	PICALM	EF1A	RPL5	UBE2A	RPL7	TUBB	GAPDH
	Most	←			Stable		→		Least

(B)



Female	EF1A	PICALM	ERRFI	RPL5	UBE2A	GAPDH	ITGB2	TUBB	RPL7
Male	ITGB2	ERRFI	RPL5	GAPDH	TUBB	RPL7	EF1A	UBE2A	PICALM
Female & male	RPL5	TUBB	PICALM	UBE2A	ERRFI	ITGB2	EF1A	RPL7	GAPDH
	Most	←			Stable		→		Least

(C)



Female	PICALM	UBE2A	RPL7	ITGB2	TUBB	ERRFI	RPL5	EF1A	GAPDH
Male	ITGB2	UBE2A	PICALM	ERRFI	EF1A	RPL5	GAPDH	RPL7	TUBB
Female & male	PICALM	UBE2A	ITGB2	ERRFI	EF1A	RPL5	RPL7	TUBB	GAPDH
	Most	←			Stable		→		Least

(D)

Fig. 4. Expression stability of candidate reference genes calculated by RankAggreg. A lower value indicates more stable expression. The samples refer to (A) brain, (B) caudal fin, (C) gonads, and (D) three tissue types.

Table 2
Ranking of nine candidate reference genes using geNorm, NormFinder, BestKeeper, and RankAggreg programs. The lower ranking values indicate the most stable gene.

Ranking order	geNorm			NormFinder		
	F (♀)	M (♂)	F (♀) & M (♂)	F (♀)	M (♂)	F (♀) & M (♂)
1	ITGB2	UBE2A	ITGB2	UBE2A	UBE2A	UBE2A
2	PICALM	ITGB2	PICALM	RPL7	ITGB2	PICALM
3	EF1A	PICALM	EF1A	RPL5	PICALM	ERRFI
4	ERRFI	EF1A	ERRFI	PICALM	ERRFI	RPL5
5	RPL7	ERRFI	UBE2A	ERRFI	EF1A	ITGB2
6	TUBB	RPL5	RPL5	TUBB	RPL5	EF1A
7	UBE2A	GAPDH	RPL7	ITGB2	GAPDH	RPL7
8	RPL5	RPL7	TUBB	EF1A	RPL7	TUBB
9	GAPDH	TUBB	GAPDH	GAPDH	TUBB	GAPDH
Ranking order	BestKeeper			RankAggreg		
	F (♀)	M (♂)	F (♀) & M (♂)	F (♀)	M (♂)	F (♀) & M (♂)
1	TUBB	ITGB2	PICALM	PICALM	ITGB2	PICALM
2	PICALM	ERRFI	UBE2A	UBE2A	UBE2A	UBE2A
3	UBE2A	UBE2A	ITGB2	RPL7	PICALM	ITGB2
4	ITGB2	PICALM	ERRFI	ITGB2	ERRFI	ERRFI
5	ERRFI	EF1A	RPL5	TUBB	ER1A	EF1A
6	RPL7	RPL5	EF1A	ERRFI	RPL5	RPL5
7	RPL5	GAPDH	TUBB	RPL5	GAPDH	RPL7
8	EF1A	RPL7	RPL7	EF1A	RPL7	TUBB
9	GAPDH	TUBB	GAPDH	GAPDH	TUBB	GAPDH

stable genes were UBE2A in females, RPL7 in males, and GAPDH in mixed samples (Fig. 4B). In the gonads, the most stable genes were EF1A and ITGB2 in females and males, respectively, and RPL5 in the combined male and female data. On the other hand, the least stable genes were RPL7 in females, PICALM in males, and GAPDH in mixed samples (Fig. 4C). When all three tissues were analyzed, the PICALM gene showed the highest expression stability in males, while the GAPDH gene showed the highest expression stability in females (Fig. 4D)

3. Experimental Design, Materials and Methods

3.1. Fish and sample collection

The silvertip tetra (*Hasemania nana*) used in the experiment was purchased from a local aquarium (Busan, Korea). Fish were bred in the laboratory and used as parents (F0), and F1 individuals were produced by the natural spawning method and used in the experiment. The water temperature in the aquarium (25.7 cm × 40 cm × 28.3 cm; W × D × H) was maintained at 24 ± 1.0 °C, and the fish were kept in a circulating filtration system. Fish were anesthetized using clove oil (Thermo Fisher Scientific, USA) at 200 ppm before dissection. Brain (including pituitary gland), caudal fin, and gonad tissues from five male and female fish were harvested by anatomical techniques and stored in an ultra-low temperature freezer (−80°C) until RNA isolation.

3.2. Selection of candidate reference genes and primer design

The sequence information of the nine reference gene candidates was collected from the de-novo assembly of whole-transcriptome RNA-seq of the silvertip tetra performed in the labo-

ratory (SRX21774992, SRX21774993, SRX21774994, SRX21774995, SRX21774996, SRX21774997). The selected genes were EF1A, ERRFI, GAPDH, ITGB2, PICALM, RPL5, RPL7, TUBB, and UBE2A, and primers for RT-qPCR were constructed based on each gene sequence. Primers for each gene were designed using Primer3Plus software (https://primer3plus.com/cgi-bin/dev/primer3plus.cgi). PCR efficiency was performed by synthesizing based on primer length (20 bp), GC content (50%), and amplification products of 150–200 bp length. PCR efficiency values (E) were evaluated using silvertip tetra adult brain, caudal fin, and gonad cDNA templates and QuantStudioTM Real-time PCR software v1.3. The slope of the standard curve determined the PCR efficiency according to the formula D=[10(−1/slope)−1] × 100% (Table 1).

3.3. RNA extraction and cDNA synthesis

Total RNA was extracted from brain, caudal fin and gonad tissues of male and female silvertip tetra using Trizol reagent (Invitrogen, USA). Total RNA was further purified using Ribospin II kit (Geneall, Korea). The experiments were also performed with DNase I treatment to remove contamination of genomic DNA. An Epoch microplate spectrophotometer (BioTeck, USA) was used to confirm that the absorbance values at 260/280 nm were within the range of 1.8–2.0. The total RNA of each sample was then calibrated to the same concentration and stored in an ultra-low temperature freezer (−80 °C). To synthesize cDNA from total RNA, reverse transcription reactions were performed using the Hyperscript RT Master Mix Kit (Geneall), oligo (dT) primers, and random hexamers according to the manufacturer's recommended method.

3.4. Real-time quantitative PCR

RT-qPCR was performed in triplicate according to the manufacturer's recommended method using the StepOnePlus Real-Time PCR system (Applied Biosystems, USA) and PowerUp SYBR Green Master Mix (Applied Biosystems) in a total reaction volume of 20 µl. All reactions were performed using the following settings: predenaturation at 95 °C for 15 min, followed by 45 cycles of 10 s at 95 °C (denaturation), 10 s at 60 °C (annealing), 72 s at 72 °C (extension).

3.5. Analysis of stability of candidate reference genes

The stability of gene expression was assessed using Ct values obtained by RT-qPCR. The programs geNorm, NormFinder, and BestKeeper were used to evaluate gene expression stability (https://www.heartcure.com.au/reffinder/). The geNorm program is a method for calculating the stability value (M value, cutoff <1.5) of reference genes [1] and is known to calculate the gene expression stability value by pairwise mutation analysis of each reference gene with another control gene [2]. As a result, the lower the stability value (M value), the more stable the gene was selected as the gene with the highest gene expression stability. The NormFinder program is used for stability analysis based on the stability of gene expression among a group of selected reference genes [3,4], and the lower the stability value, the more appropriate reference gene was selected [5]. The BestKeeper program was used for gene stability analysis based on the standard deviation of Ct values for each gene and the gene coefficient of variation [6]. The RankAggreg analysis is. A method to analyze the stability ranking of reference genes by synthesizing the results of three programs: geNorm, NormFinder, and BestKeeper [7].

Ethics Statement

The experimental protocol used in this study was approved by the Animal Ethics Committee of Gangneung-Wonju National University (GWNU-2019-26). In addition, all authors who partic-

ipated in the study completed the Animal Welfare & Ethics course certification by accrediting the research ethics and CITI program, research ethics and compliance training program.

Data Availability

PRJNA1017117 (Original data) (BioProject in GenBank).

CRediT Author Statement

Hwa Jin Lee: Investigation, Writing – original draft; **Ki Tae Kim:** Investigation, Writing – original draft; **Min Sun Kim:** Investigation, Writing – original draft; **Sang Yoon Lee:** Supervision, Writing – review & editing.

Acknowledgments

This research was supported by the Bio & Medical Technology Development Program of the National Research Foundation (NRF) funded by the Ministry of Science and ICT (MSIT) of South Korea (NRF-2021M3E5E6021114) and by a grant from National Institute of Fisheries Science (R2024029).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] J. Vandesompele, K. De Preter, F. Pattyn, B. Poppe, N.V. Roy, A.D. Paepe, F. Speleman, Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes, Genome Biol. 3 (2002) 1–12 http://genomebiology.com/2002/3/7/research/0034.1.

[2] H. Wang, X. Zhang, Q. Liu, X. Liu, S. Ding, Selection and evaluation of new reference genes for RT-qPCR analysis in *Epinephelus akaara* based on transcriptome data, PLoS One 12 (2) (2017) 1–18, doi:10.1371/journal.pone.0171646.

[3] C.L. Andersen, J.L. Jensen, T.F. Ørntoft, 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 (15) (2004) 5245–5250, doi:10.1158/0008-5472.CAN-04-0496.

[4] M. Xu, B. Zhang, X. Su, S. Zhang, M. Huang, Reference gene selection for quantitative real-time polymerase chain reaction in populus, Anal. Biochem. 408 (2) (2011) 337–339, doi:10.1016/j.ab.2010.08.044.

[5] Y. Gao, Y. Gao, B. Huang, Z. Meng, Y. Jia, Reference gene validation for quantification of gene expression during ovarian development of turbot (*Scophthalmus maximus*), Sci. Rep. 10 (1) (2020) 1–10, doi:10.1038/s41598-020-57633-3.

[6] R.S. McCulloch, M.S. Ashwell, A.T. O'Nan, P.L. Mente, Identification of stable normalization genes for quantitative real-time PCR in porcine articular cartilage, J. Anim. Sci. Biotechnol. 3 (36) (2012) 1–7 http://www.jasbsci.com/content/3/1/36.

[7] H. Song, W. Mao, Z. Duan, et al., Selection and validation of reference genes for measuring gene expression in *Toona ciliata* under different experimental conditions by quantitative real-time PCR analysis, BMC Plant Biol. 20 (2020) 1–14, doi:10.1186/s12870-020-02670-3.