Selection of a reference gene for studies of adipose tissues of toothed whales

JDM Senevirathna[1,4 *], Ryo Yonezawa[1], Taiki Saka[1], Yoji Igarashi[2], Kazutoshi Yoshitake[1], Shigeharu Kinoshita[1], Noriko Funasaka[3], Shuichi Asakawa[1]

[1]Laboratory of Aquatic Molecular Biology and Biotechnology, Department of Aquatic Bioscience, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Japan

[2]Department of Life Sciences and Chemistry, Graduate School of Bioresources, Mie University, Japan

[3]Department of Marine Bioresources, Graduate School of Bioresources, Mie University, Japan

[4]Department of Animal Science, Faculty of Animal Science and Export Agriculture, Uva Wellassa University, Sri Lanka

***Correspondence**

E-mail: duminda@uwu.ac.lk

Abstract

Fat metabolism in toothed whales is different from other mammals. RT-qPCR is still a reliable technique for studying the relative expressions of various genes involved in metabolism. This study was done for Risso's dolphin, a toothed whale and information produced here will be important for further transcriptomics studies focused on unrevealed marine mammal fat metabolism. In this study, we sought to identify a suitable reference gene with minimum resources. Seven candidate reference genes ZC3H10, FTL, LGALS1, RPL27A, GAPDH, FTH1 and DCN were initially tested for amplification efficiency using RT-qPCR by producing standard curves. Then, three nearly 100% efficient genes FTL, LGALS1 and GAPDH were selected for the gene stability analysis to determine one stable gene across eight different fat tissues, liver, and muscle of Risso's dolphins based on four algorithms, provided in geNorm, NormFinder, BestKeeper and Delta Ct. Finally, a RefFinder comprehensive ranking was done based on stability values and the genes were ranked as: FTL>LGALS1>GAPDH. The FTL and LGHLS were identified as the most stable genes; however, GAPDH was third, a well-known housekeeping gene for mammals. Finally, we suggest using FTL as a reliable reference gene for functional genomics studies on toothed whales in the future.

KEYWORDS; Dolphin Fat, Gene Expression, Gene Stability, qPCR, Reference Gene, Transcriptomics

Introduction

The Risso's dolphin is one of the marine mammals in the order Cetacea, a species of toothed whales which exhibits cosmopolitanism (Gaspari, Airoldi and Hoelzel, 2007; Gaspari and Natoli, 2012). These animals show unique metabolic adaptations to the marine environment. Lipids in toothed whales' function in several parts of the body and are involved in their special adaptations like echolocation. The fatty acid composition of these fats varies based on species, age and type of tissue (Koopman, 2018). A recent study has identified positively selected genes for lipid metabolism in Cetacea and unique features for studying functional modifications of multiple genes (Endo, Kamei and Inoue-Murayama, 2018). Therefore, there is an increasing interest in studying toothed whales at genomic and transcriptomic levels to reveal potential genes and gene ontologies involved in this specialized process of fat metabolism.

Reverse transcription quantitative real-time polymerase chain reaction (RT-qPCR) is the most convenient, reliable, widely used method to analyze gene expression in any type of cell (Bustin and Mueller, 2005). This technique requires an accurate and reliable reference or housekeeping gene (HKG) to normalize specific gene expression data for relative quantification (Almeida-Oliveira *et al.*, 2017). The housekeeping gene is a gene that is expressed in all cells and is involved with basic metabolic and maintenance activities. However, in different tissues in different species, the expression level of reference genes may show significant differences (Radonić *et al.*, 2004). Therefore, identification of a reliable reference gene is important before comparison with genes of interest. Specifically, in toothed whales, the adipose tissue metabolic pathway is still not clear and supportive lipid metabolism genes have not been investigated in transcriptomic level. There are possibilities of having unique de novo biosynthesis pathways of lipids in the toothed whales (Koopman, 2018). Therefore, identification of a stable reference gene is desirable for the future RT-qPCR assays of functional genomics of lipid metabolic pathways.

Some of the commonly used reference genes for mammals are 18S ribosomal RNA (18S rRNA), glyceraldehyde-3-phosphate dehydrogenase (GAPDH), 28S ribosomal RNA (28SrRNA), β actin (ACTB) and succinate dehydrogenase complex subunit A (SDHA). These genes show a high level of expression in various mammalian tissues. However, under different tissue, environmental, and experimental conditions, the level of expression can vary considerably. For example, a disease-associated gene, peroxisome proliferator-activated receptor gamma coactivator 1-alpha (PGC1α), was tested in muscle and adipose tissue with different reference genes and alteration of the expression profile was observed (Perez *et al.*, 2017). In marine mammals, the following genes have been evaluated as HKGs in skin biopsies and blood samples from various species: phosphoglycerate kinase 1 (PGK1), hypoxanthine phosphoribosyl transferase 1 (HPRT1) and ribosomal protein L4 (RPL4) (Chen *et al.*, 2015); ribosomal protein RPLL8 (Buckman *et al.*, 2011); GAPDH (Spinsanti *et al.*, 2006; Mancia, Warr and Chapman, 2008); tyrosin 3-monooxygenase/tryptophan 5-monooxygenase activation protein zeta (YWHAZ) (Beineke *et al.*, 2004, 2007; Muller and Woods, 2013); Ribosomal Protein S9 (RPS9) (Sitt *et al.*, 2008, 2010; Viscarra and Ortiz, 2014); Ribosomal Protein S18 (RPS18) (Martinez-Levasseur *et al.*, 2013). However, the selection of HKGs for marine mammals is very limited and until now, there has been no experiment tested for fat tissues in dolphins to our knowledge.

In this study, the experimental design was arranged with emphasis placed on having a limited number of sample size, a low amount of RNA, a low budget for qPCR and identification of novel HKGs (Fig 1). According to these factors, seven candidate genes were initially selected based on transcriptomic Fragments Per Kilobase of transcript per Million mapped reads (FPKM) values (unpublished data), the coefficient of variance (CV), and gene function. Then, identified genes were evaluated for amplification efficiency by using RT-qPCR, and only three genes were selected for further stability analysis with the hope of identifying the most stable reference gene using eight types of fat tissue, liver tissue, and muscle tissue in Risso's dolphin. The null hypothesis stated there was no significant difference in selected reference gene stability, and this was tested by four statistical algorithms. Finally, this work validated

stable novel genes in the selected tissue types. Therefore, in the future, these findings can be used for gene expression analysis in marine mammal species mainly for fat metabolism studies.

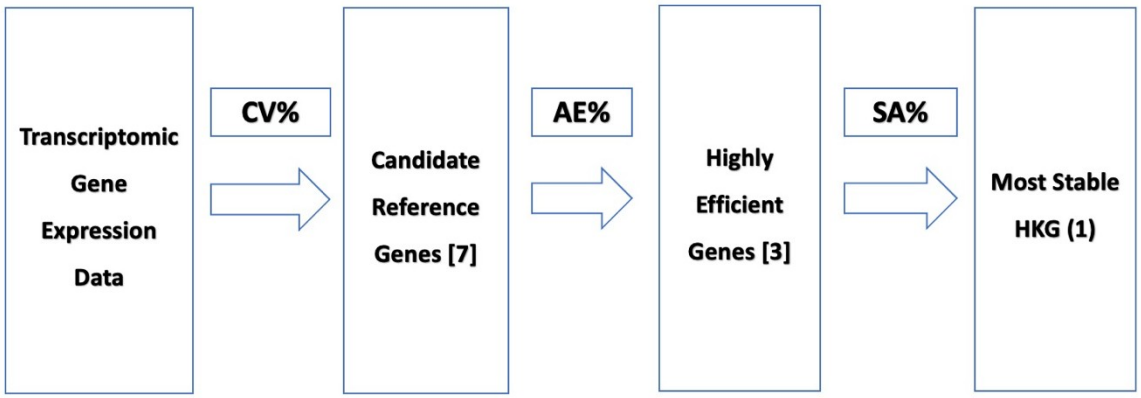


Fig 1. Diagram of experimental design with minimum resources. CV% - Coefficient of Variance calculated by FPKM values, AE% - Amplification Efficiency based on the RT-qPCR, SA% - Stability Analysis by statistical algorithms

Materials and Methods

Sample collection

Ten tissue types were received from three male Risso's dolphins (sample ID: 19TK409, 19TK410 and 19TK411) under the cooperation of the biological surveys of the National Research Institute of Far Seas Fisheries, Japan Fisheries Research and Education Agency. Tissues included four different parts of melon, two types of jaw fat (inner jaw and outer jaw), two types of blubber tissue (inner and outer), liver and muscle. These tissues were preserved in RNAlater solution at the site and then transported to the laboratory and stored at -80°C until RNA extraction.

RNA extraction and cDNA synthesis

The RNA was extracted by the RNAiso Plus (Total RNA extraction reagent) method from TAKARA BIO INC using the manufacturer's instructions (Takara Bio Inc, 2017). In summary, 20 mg of each sample homogenized for 2 minutes at 6000 rpm with a tissue homogenizer (Precellys® 24) with RNAiso plus and beads. Homogenized samples were transferred to a centrifuge tube and incubated for 5 minutes at room temperature (RT). Then, samples were centrifuged at 12,000 x g for 5 minutes at 4°C and supernatants were collected into new tubes. One-fifth the amount of chloroform was added compared to the amount of RNAiso plus to each sample and mixed until the solution became milky. The tubes were kept for 5 minutes at RT and centrifuged at 12,000 x g for 15 minutes at 4°C. The top layer that contained RNA was transferred to new tubes. The RNA pellets were precipitated by adding 1/2 the amount of isopropanol and centrifuging at 12,000 x g for 10 minutes at 4°C. Precipitates were washed using 75% cold ethanol and centrifuged at 7,500 x g for 5 minutes at 4°C. The RNA pellets were dried for several minutes and dissolved with 100 µl RNase-free water.

RNA clean-up was conducted using a NucleoSpin RNA Clean-up XS kit (MACHEREY-NAGEL) following the protocol in (View, 2014). First, DNA in the crude RNA extracts was digested by adding 1/10 mixture of rDNase (Macherey-Nagel 740963) and a reaction buffer and incubating for 10 minutes at 37°C. An equal volume of buffer RCU was then added to each sample and mixed 2 x 5 seconds. NucleoSpin RNA XS columns and centrifuged for 30 seconds at 11,000 x g. Next, 400 µl of buffer RA3 was added to each tube and centrifuged for 30 seconds at 11,000 x g. The flow-through was discarded and 200 ul of buffer RA3 was added to the columns and centrifuged for 2 minutes at 11,000 x g to dry the membrane containing purified RNA. Then, each column was placed in a nuclease-free collection tube (1.5 ml, supplied) and RNA was eluted in 30 µl RNase-free water by centrifuging at 11,000 x g for 30 seconds. RNA concentrations were checked using the Qubit® RNA Assay Kit in Qubit®2.0 Flurometer (Life Technologies, Carlsbad, CA, USA) and RNA quality was verified using the Agilent 2200 Tape Station (Agilent Technologies, Santa Clara, CA, USA) (Technologies, 2015). The RNA samples were kept at -80°C for further use.

First strand cDNA was synthesized for each RNA sample using the PrimerScriptRT master mix kit (Perfect Real Time), using 2 µl of master mix, up to 500 ng purified RNA and RNase-free water in a 10 µl solution on ice and incubating on a thermal cycler at 37°C for 15 minutes, and 85°C for 5 seconds. cDNA concentrations were checked for all samples using the Qubit, and the presence of cDNA was also checked by normal PCR using a selected primer before RT-qPCR. The cDNA samples were stored at -20°C until use.

Selection of candidate reference genes using RNA-seq data and designation of primers

Unpublished transcriptomic gene expression, fragment per kilobase per million reads (FPKM) data generated in a illumine HiSeq RNA-seq analysis of the same tissues from our laboratory were used to select the highly expressed candidate reference genes. The expression stability of these selected genes was calculated and briefly described as follows (Wang *et al.*, 2014). Firstly, using FPKM values of gene expression in 30 samples, mean expression values (MVs) and standard deviations (SDs) were calculated for each gene. Secondly, the coefficient of variation (CV) of each gene was ranked according to the CV value (S1 Table). Seven reference genes were selected including novel reference genes based on a low CV value, gene function in metabolism, averaged over a value of 1000 FPKM and more information can be found in previous publications.

RT-qPCR primers were designed using primer3 (http://bioinfo.ut.ee/primer3-0.4.0/) considering primer length; ~20 bp; product size: 150-250 bp; GC% content: 40%-60% (primers are shown in Table 2). Before RT-qPCR analysis, accuracy and specificity of all primers were checked by normal PCR conditions (using dNTP mix 1.6 µl, 10x Ex Taq buffer 2 µl, Ex Taq 0.1 µl, nuclease free water 13.3 µl,

0.5 µl from each forward and reverse primer, and 2 µl of cDNA as the template under the thermal cycle conditions of 95 °C for 3 minutes, 35 cycles of 98 °C for 10 seconds, 60 °C for 30 seconds, 72 °C for 30 seconds, and 72 °C for 3 minutes final extension) and 2% agarose gel electrophoresis.

Table 1. Information about suitable reference genes

Gene	Name	Function
ZC3H10	Zinc Finger CCCH-Type Containing 10	specifically regulates miRNA biogenesis
FRIL/FTL	Ferritin Light Chain	stores iron in a soluble and nontoxic state
LEG1/LGALS1	Galectin 1	modulates cell-cell and cell-matrix interactions
RPL27A	Ribosomal Protein L27	catalyzes protein synthesis
G3P/GAPDH	Glyceraldehyde-3-Phosphate Dehydrogenase	has both glyceraldehyde-3-phosphate dehydrogenase and nitrosylase activities
FRIH/FTH1	Ferritin Heavy Chain 1	stores iron in a soluble, non-toxic, readily available form
PGS2/DCN	Decorin	plays a role in collagen fibril assembly

Realtime quantitative PCR (RT-qPCR) and amplification efficiency

RT-qPCR was performed in 96-well plates on an ABI 7300 Fast Real-Time PCR system (ABI, Alameda, CA, USA) using the SYBR Premix Ex Taq™ Takara kit (Takara Bio Inc, 2015). One qPCR reaction volume was 20 µl and conditions were as follows: 10 µl SYBR Premix Ex Taq (1x), 0.8 µl of 50 nM of both forward and reverse primers, 0.4 ul of ROX reference dye (50X), 6 µl of nuclease-free water and 2 µl of 100 ng of cDNA. All reactions and no template controls (NTC) were conducted in triplicates. After 40 cycles, a melting curve analysis was conducted ranging from 55°C to 95°C and cycle values (Ct

187 data) were obtained from the relevant software of the RT-qPCR system by automatically determining the
188 threshold values.
189
190 Standard curves were acquired for each primer pair by amplification in serial dilutions such as
191 1:1, 1:10, 1:100, 1:1000, and 1:10000 for all samples (S2 Table and S1 Fig). The correlation coefficient
192 (R^2) and the PCR amplification efficiency (E) for each gene were calculated from the slope of a standard
193 curve,
194 $E = (10^{(-1/slope)} - 1) \text{ X } 100\%$
195 (https://www.chem.agilent.com/store/biocalculators/calcSlopeEfficiency.jsp?_requestid=1040745)
196 (Bustin *et al.*, 2009). Based on the closer to 100% 0f E and closer to 1 of R^2 values, three reference genes
197 were selected for further analysis of gene stability to test for all biological samples in three replicates.
198

199 ## Gene stability analysis

200
201 A normality test for Ct data was conducted in the SPSS 17 statistical package (SPSS Inc, 2008)
202 using the Kolmogorov-Smirnov test (KS-test) and the normality of different genes was evaluated by p
203 value. Different statistical algorithms such as geNorm (Vandesompele *et al.*, 2002), NormFinder
204 (Andersen, Jensen and Ørntoft, 2004), BestKeeper (Pfaffl *et al.*, 2004), comparative Delta Ct method
205 (Silver *et al.*, 2006), and RefFinder (Xie *et al.*, 2012) were used to determine the stability of candidate
206 reference genes by an online version of RefFinder developed by Dr. Zhang's Lab
207 (http://www.ciidirsinaloa.com.mx/RefFinder-master/?type=reference#; (Xie *et al.*, 2012)). The raw Ct
208 data from each sample were used as an input file. The geNorm performed a pair-wise comparison among
209 genes and calculated the expression stability value (M) for each gene, in which a low M value is more
210 stable. NormFinder also calculated an M value of gene stability based on the inter- and intra-group
211 variance, and gives the lowest M value to the most stable genes. The BestKeeper algorithm was used to
212 determine standard deviation and genes were rated based on variance, which also gives low values of SD

and CV for stable genes. Delta Ct values were obtained from pairwise differences between Ct values of reference genes to determine the stability of genes by repeatability among all samples. RefFinder considered the results of all above algorithms and gave a completed final ranking of reference genes. After selecting the most stable novel genes, they were analyzed for gene stability in 10 different tissues using three pieces of data from three replicates and comprehensive graphs were prepared for the discussion.

Results

In this study, a total of 30 samples, 10 tissues from three individual Risso's dolphins were used to identify potential reference genes for normalization of qPCR data across fat tissues and to validate transcriptomic data. Gene IDs, gene names, and functions of the candidate reference genes are included in Table 1. The genes, primer information, amplicon length, and amplification information are found in Table 2.

Table 2. Primer information and amplification efficiencies for suitable reference genes

Gene ID	Primer Sequence (5'-3') forward/reverse	Amplicon Size (bp)	Tm °C	Amplicon Efficiency (%) SD	R^2	Slope
ZC3H10	GATGAGGCAGAGGTCCAAGT	234	58.4	82.259	0.975	-3.836
	TGGCCATCGATGTTCCAGAT		56.3			
FRIL/FTL	GGCTTCTATTTCGACCGCGA	228	58.4	95.184	0.995	-3.443
	GGCCTCGTTCAGGTTCTTCT		58.4			
LEG1/LGALS1	ATGGCTTGTTGTGACGCATT	182	54.3	94.394	0.997	-3.464
	GGTCCTGGGCAAGTTTCTTG		58.4			
RPL27A	GGGTGAAAGGTAAGCGGAAG	167	58.4	122.383	0.954	-2.881

	CTGAGGTGCCGTCATCAATG		58.4			
G3P/GAPDH	GTCGGAGTGAACGGATTTGG	218	58.4	94.882	0.999	-3.451
	TGGAAGATGGTGATGGCCTT		56.3			
FRIH/FTH1	GCCTCCTACGTCTACCTGTC	197	60.4	79.929	0.885	-3.92
	TTCTCCCAGTCATCACGGTC		58.4			
PGS2/DCN	AGAAGCTCTCCTACATCCGC	213	58.4	1936.58	0.914	-0.764
	AATGAGGAGTGTTGGCCAGA		56.3			

Identification of candidate reference genes based on transcriptome data

Based on the RNA-seq data from our laboratory (not yet published), seven highly expressed (FPKM>1000) genes with CV ranging from 46.14% to 175.75% were identified in this study (Table 1, S1 Table). Some genes were newly identified as candidate reference genes such as ferritin light chain (FTL), zinc finger CCCH-type containing 10 (ZC3H10), galectin 1 (LGALS1), ferritin heavy chain 1 (FTH1) and decorin (DCN).

Primer specificity, amplification efficiency, and expression profiles

The specificity of each primer pair was confirmed by 2% agarose gel electrophoresis with a single band (S2 Fig). All genes presented a single melting peak in the qPCR, indicating specific amplification (S3 Fig). The cDNA-free NTC (negative control) samples did not show any melting curve product. These results indicate an absence of errors in the RT-qPCR amplification. qPCR performance was good according to the results of coefficient (R2) and efficiency (E) (Table 1 and S1 Figure). According to the five-point standard curve, three genes showed good slopes and efficiencies: FTL (-

3.443, 95.184%), LGALS1 (-3.464, 94.394%) and GAPDH (-3.451, 94.882%), respectively. Therefore, these genes were selected for gene stability analysis and finally to select a potential reference gene for dolphin fat tissues.

All the Ct data of the different genes were in the normal distribution according to the KS-test ($p > 0.05$) (Fig 2). The expression profile of the three selected genes was tested by the mean quantification cycle (Ct) values for each sample, after testing for normality using the KS-test. The Ct values for three selected genes were 15.05 for FTL, 17.00 for GAPDH and 22.64 for LGALS1 (Fig 2). This shows the expression of these candidate genes was highly transcribed (Chen *et al.*, 2015) and suited with reference gene expression level (15<Ct>30). The lowest Ct value is for FTL gene, while the highest expression and highest Ct value are for LGALS1.

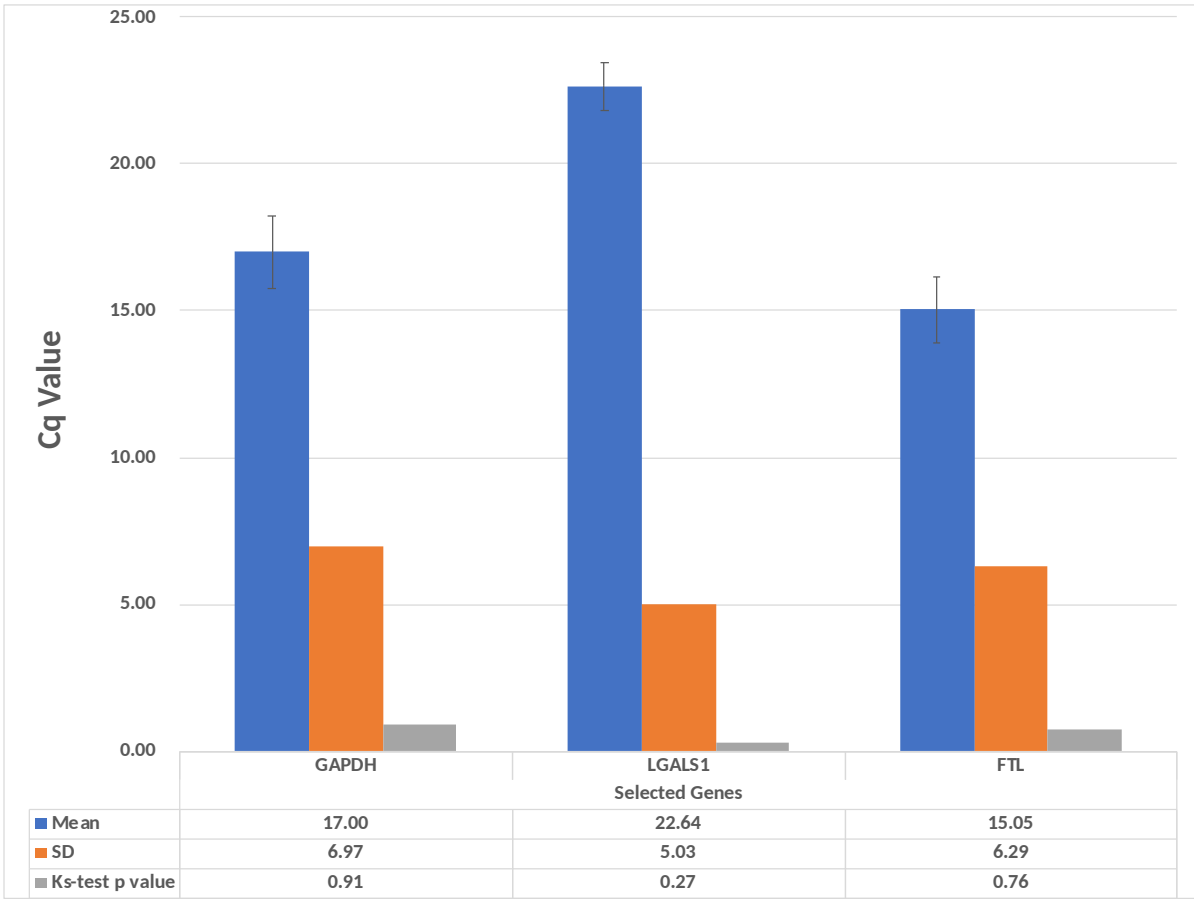


Fig 2. Distribution of mean Ct values of 3 selected reference genes in all samples. Bar chart showing mean Ct values with standard error of mean, standard deviation (SD) and KS-test p values.

Validation of reference genes and identification of the most the stable house-keeping genes

The geNorm statistical algorithm evaluated the stability of the selected reference genes using M value, the lowest values given to the most stable genes, FTL and LGALS1. GAPDH gene had the highest M value suggesting it was the least stable (Table 3 and S4 Fig). NormFinder and Delta CT showed the best stability (1.495) for the FTL gene, followed by LGALS1 and GAPDH. Interestingly, the BestKeeper gives the most stable value for the LGALS1 gene and then FTL, followed by GAPDH. The BestKeeper algorithm also calculated the Pearson correlation coefficient (r) among gene stability values, and it was significantly different ($p < 0.05$) (Table 4). Therefore, in the RefFinder comprehensive ranking, the FTL gene has been identified as the most stable reference or housekeeping gene for transcriptome analysis, especially for different types of fat tissue in toothed whales (Table 5).

Table 3. Results of stability among three selected reference genes. Four statistical algorithms (Delta CT, BestKeeper, NormFinder, and geNorm) have been used estimation of stability values concerning 30 samples.

Selected Reference Genes	**RefFinder Comprehensive Ranking**		**Delta CT**		**BestKeeper**		**NormFinder**		**geNorm**	
	Stability Value	Rank	Average of SD	Rank	SD	Rank	Stability Value	Rank	M value	Rank

FTL	1.189	1	3.334	1	4.385	2	1.495	1	2.991	1
LGALS1	1.414	2	3.889	2	3.278	1	3.029	2	2.991	1
GAPDH	3.000	3	4.232	3	5.425	3	3.707	3	3.818	2

Table 4. Pearson correlation coefficient (r) by BestKeeper statistical algorithm

	GAPDH	LGALS1	FTL
LGALS1	0.698	-	-
FTL	0.834	0.857	-
p-value	0.001	0.001	-

Table 5. Summary of RefFinder Gene Ranking from all algorithms

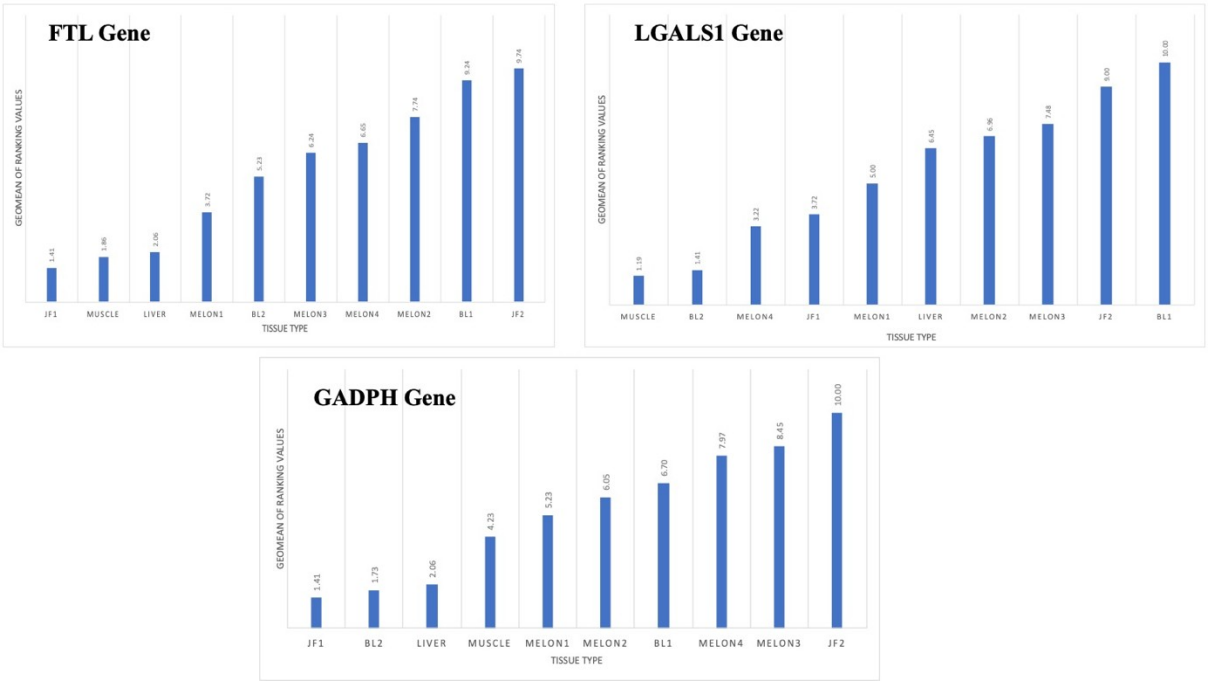
Ranking Order (Better--Good--Average)			
Method	1	2	3
Delta CT	FTL	LGALS1	GAPDH
BestKeeper	LGALS1	FTL	GAPDH
NormFinder	FTL	LGALS1	GAPDH
geNorm	LGALS1| FTL		GAPDH
Recommended comprehensive ranking	FTL	LGALS1	GAPDH

Stability of FTL and LGALS1 genes in different tissues

Based on the comprehensive gene stability values, the FTL gene was most stable in the inner jaw fat sample (JF1), having the lowest stability value, 1.414, and had the lowest stability in the outer jaw fat

sample (JF2) and inner blubber (BL1). However, interestingly, the LGALS1 gene was more stable in muscle samples than other fat tissues and was least stable in BL1 and JF2 (Fig 3).

Fig 3. Comprehensive gene stability in ten different tissues of Risso's dolphins. JF1 – inner jaw fat; JF2 – outer jaw fat, BL1 – inner blubber fat; BL2 – outer blubber fat. Four parts of melon fats have been identified in this study and named 1, 2, 3, and 4.



Discussion

Several factors are affected by the level of RNA expression at the experimental level. In this research, all the samples were collected at the same time and three biological replicates were used. Then, samples were preserved in RNAlater solution. Therefore, we believe that the total RNA of different types of tissue samples was not degraded and not significantly affected in a way that would change the Ct value of the RT-qPCR technique. This assumption was also confirmed by the normality test, which showed $p>0.05$, meaning all samples were in the normal distribution.

316

317 Selection of candidate reference genes for this kind of study may depend on several factors, like
318 transcriptomic expression data, gene function, type of tissue, availability of samples, quantity and quality
319 of RNA extracted from cells, and budget of the experiment. In this study, we specifically focused on
320 finding a suitable reference gene in toothed whales for a variety of adipose tissues such as melon, jaw fat,
321 and blubber. Liver and muscle tissues were also used as controls since they usually express many types of
322 genes. In comparative genomics analysis, it is evident that many genes express in dolphin for aquatic
323 adaptations, blubber, and fat storage (Mancia, 2018). Their fat tissues have different fatty acid, and fatty
324 alcohol combinations; thus, the metabolism of these fats is still unclear. Considering all these aspects, we
325 identified candidate genes with the lowest CV percentage. Reference genes for qPCR can be identified by
326 high-throughput transcriptomic data by next generation sequencing (Gao *et al.*, 2018; Pombo *et al.*, 2019;
327 Bao *et al.*, 2020). All our candidate reference genes showed a high level of expression by more than 1000
328 FPKM values and according to different sources, their functions included regulatory, maintenance,
329 metabolic processes, and catalysis in cells. Finally, this experiment was designed to identify seven
330 candidate reference genes, including five uncommon reference genes and two already known HKGs.
331 Then, the experiment was intended to select three genes for stability analysis with amplification
332 efficiency being considered. The FTL gene was identified as a novel reference gene using 30 various
333 tissue samples by RT-qPCR. We believe this experimental design is adequate for identifying a stable
334 HKG or at least a potential reference gene with minimum experimental work and a low budget,
335 particularly for precious samples like marine mammals.

336

337 RT-qPCR has become widely used as an accurate and sensitive method to determine gene
338 expression under various conditions, for many types of cells in different organisms. Reference genes are
339 important for determining the level of expression of target genes like up-regulation or down-regulation by
340 normalization (36, 37). Several reference genes can be diagnosed by statistical algorithms such as
341 NormFinder, BestKeeper, geNorm, and Delta Ct (Xie *et al.*, 2012). Therefore, the use of multiple HKGs

342 in the normalization of genes of interest is common practice nowadays. The FTL and LGALS1 genes we
343 identified are suitable for future functional genomics analysis of the metabolism of toothed whales. For
344 the selection of stable genes, it is also recommended that more than one algorithm is used, which allows
345 for analysis of data in different ways. Housekeeping genes also maintain normal cellular functions,
346 therefore, it also confirmed in our experiment by the stably expression of FTL, LGALS1, and GADPH
347 genes in muscle, and liver tissues.
348
349 In our study, FTL has been identified as the most stable gene for this tissue of Risso's dolphins,
350 which may be a new finding and can be used for future studies. Also, FTL can now be used as an HKG in
351 marine mammal molecular biology research. FTL is a protein, ferritin light chain, mainly involved in the
352 storage of iron in a soluble and nontoxic state in different cells (*FTL Protein Human Recombinant* |
353 *Ferritin Antigen* | *ProSpec*, no date). Supportively, the ERgene; Python library for reference gene
354 selection has identified the FTL gene as one of stable endogenous reference gene for various studies
355 (Zeng *et al.*, 2020). In another study, TATA-binding protein (TBP) and ATPF1 were experimentally
356 identified as stable reference genes in the adipose tissue of mice (Almeida-Oliveira *et al.*, 2017).
357 However, in our experimental samples, these genes were not highly expressed. LGALS1 was ranked
358 second and one of the functions of this gene is modulating immune responses (Nishi *et al.*, 2008).
359 According to the recent updates, LGALS1 is highly expressed in fat tissues (NCBI, 2021), therefore we
360 also believe it as a stable gene in fats of toothed whales. Moreover, FTL and LGALS1 have not been used
361 as HKGs for many experiments to our knowledge and are thus novel reference genes for marine
362 mammals. GAPDH was classified third, however, it is a very common reference gene for many tissues of
363 numerous animal species. RPS9 and YWHAZ genes were also identified as suitable reference genes from
364 other dolphin blood samples, but never for Risso's dolphins.

Conclusions

For the first time in toothed whale research, this study has systematically selected and evaluated stable reference genes for RT-qPCR in several types of tissue including melon, jaw fat, blubber, liver, and muscle, based on RNA data. Although the study had many limitations, seven highly expressed candidate reference genes were checked for efficiency. Two novel genes, FTL and LGALS1, were identified as stable reference genes for the first time in toothed whales. The FTL gene was the optimal reference gene for all types of tissues tested. It seems plausible that this provides an accurate normalization factor for expression data in genes of interest in different toothed whales' fat tissues.

Acknowledgments

We would like to thank the Taiji Fishery Association, researchers of National Research Institute of Far Seas Fisheries, and researchers of Japan Fisheries Research and Education Agency for helping this research.

Data Accessibility Statement

Selected transcriptomic data will be deposited in DDBJ (DNA Data Bank of Japan) or any other GenBank. All other data will be publicly available with this publication.

References

Almeida-Oliveira, F. *et al.* (2017) 'Reference genes for quantitative PCR in the adipose tissue of mice with metabolic disease', *Biomedicine and Pharmacotherapy*. Elsevier Masson SAS, 88, pp. 948–955. doi: 10.1016/j.biopha.2017.01.091.

Andersen, C. L., Jensen, J. L. and Ørntoft, T. F. (2004) 'Andersen, Jensen, Ørntoft - 2004.pdf', pp. 5245–5250.

Bao, Z. *et al.* (2020) 'Identification and selection of reference genes for quantitative transcript analysis in Corydalis yanhusuo', *Genes*, 11(2). doi: 10.3390/genes11020130.

Beineke, A. *et al.* (2004) 'Development of a lymphocyte-transformation-assay for peripheral blood lymphocytes of the harbor porpoise and detection of cytokines using the reverse-transcription polymerase chain reaction', *Veterinary Immunology and Immunopathology*. Elsevier, 98(1–2), pp. 59–68. doi: 10.1016/j.vetimm.2003.10.002.

Beineke, A. *et al.* (2007) 'Increased blood interleukin-10 mRNA levels in diseased free-ranging harbor porpoises (Phocoena phocoena)', *Veterinary Immunology and Immunopathology*, 115(1–2), pp. 100–106. doi: 10.1016/j.vetimm.2006.09.006.

Buckman, A. H. *et al.* (2011) 'PCB-associated changes in mRNA expression in killer whales (Orcinus orca) from the NE pacific ocean', *Environmental Science and Technology*, 45(23), pp. 10194–10202. doi: 10.1021/es201541j.

Bustin, S. A. *et al.* (2009) 'The MIQE guidelines: Minimum information for publication of quantitative real-time PCR experiments', *Clinical Chemistry*, 55(4), pp. 611–622. doi: 10.1373/clinchem.2008.112797.

Bustin, S. A. and Mueller, R. (2005) 'Real-time reverse transcription PCR (qRT-PCR) and its potential use in clinical diagnosis', *Clinical Science*, 109(4), pp. 365–379. doi: 10.1042/cs20050086.

Chen, I. H. *et al.* (2015) 'Selection of suitable reference genes for normalization of quantitative RT-PCR in peripheral blood samples of bottlenose dolphins (Tursiops truncatus)', *Scientific Reports*. Nature Publishing Group, 5. doi: 10.1038/srep15425.

Endo, Y., Kamei, K.-I. and Inoue-Murayama, M. (2018) 'Genetic Signatures of Lipid Metabolism Evolution in Cetacea'. doi: 10.1101/250597.

Freitas, F. C. P. *et al.* (2019) 'Evaluation of reference genes for gene expression analysis by real-time quantitative PCR (qPCR) in three stingless bee species (Hymenoptera: Apidae: Meliponini)', *Scientific Reports*, 9(1), pp. 1–13. doi: 10.1038/s41598-019-53544-0.

FTL Protein Human Recombinant | Ferritin Antigen | ProSpec (no date). Available at:

https://www.prospecbio.com/ferritin_human (Accessed: 29 December 2020).

Gao, D. *et al.* (2018) 'Transcriptome-wide identification of optimal reference genes for expression analysis of Pyropia yezoensis responses to abiotic stress', *BMC Genomics*. BMC Genomics, 19(1), pp. 1–14. doi: 10.1186/s12864-018-4643-8.

Gaspari, S., Airoldi, S. and Hoelzel, A. R. (2007) 'Risso's dolphins (Grampus griseus) in UK waters are differentiated from a population in the Mediterranean Sea and genetically less diverse', *Conservation Genetics*, 8(3), pp. 727–732. doi: 10.1007/s10592-006-9205-y.

Gaspari, S. and Natoli, A. (2012) 'Grampus griseus (Mediterranean Subpopulation)', *The IUCN Red List of Threatened Species 2012*, 8235. Available at: http://www.iucnredlist.org/details/9461/0.

Koopman, H. N. (2018) 'Function and evolution of specialized endogenous lipids in toothed whales', *The Journal of Experimental Biology*, 221(Suppl 1), p. jeb161471. doi: 10.1242/jeb.161471.

Kozera, B. and Rapacz, M. (2013) 'Reference genes in real-time PCR', *Journal of Applied Genetics*, 54(4), pp. 391–406. doi: 10.1007/s13353-013-0173-x.

Mancia, A. (2018) 'On the revolution of cetacean evolution', *Marine Genomics*. Elsevier, 41(August), pp. 1–5. doi: 10.1016/j.margen.2018.08.004.

Mancia, A., Warr, G. W. and Chapman, R. W. (2008) 'A transcriptomic analysis of the stress induced by capture-release health assessment studies in wild dolphins (Tursiops truncatus)', *Molecular Ecology*, 17(11), pp. 2581–2589. doi: 10.1111/j.1365-294X.2008.03784.x.

Martinez-Levasseur, L. M. *et al.* (2013) 'Control and target gene selection for studies on UV-induced genotoxicity in whales', *BMC Research Notes*, 6(1). doi: 10.1186/1756-0500-6-264.

Muller, H. K. and Woods, G. M. (2013) 'Ultraviolet Radiation Effects on the Proteome of Skin Cells', in *Adv Exp Med Biol*, pp. 111–119. doi: 10.1007/978-94-007-5896-4_8.

NCBI (2021) *LGALS1 galectin 1 [Homo sapiens (human)] - Gene - NCBI*. Available at: https://www.ncbi.nlm.nih.gov/gene?Db=gene&Cmd=DetailsSearch&Term=3956 (Accessed: 17 February 2021).

Nishi, N. *et al.* (2008) 'Functional and structural bases of a cysteine-less mutant as a long-lasting

substitute for galectin-1', *Glycobiology*, 18(12), pp. 1065–1073. doi: 10.1093/glycob/cwn089.

Perez, L. J. *et al.* (2017) 'Validation of optimal reference genes for quantitative real time PCR in muscle and adipose tissue for obesity and diabetes research', *Scientific Reports*. Nature Publishing Group, 7(1). doi: 10.1038/s41598-017-03730-9.

Pfaffl, M. W. *et al.* (2004) 'Determination of most stable housekeeping genes, differentially regulated target genes and sample integrity : BestKeeper', *Biotechnology Letters*, 26, pp. 509–515. doi: JCMM395 [pii]\r10.1111/j.1582-4934.2008.00395.x.

Pombo, M. A. *et al.* (2019) 'Transcriptome-based identification and validation of reference genes for plant-bacteria interaction studies using Nicotiana benthamiana', *Scientific Reports*. Springer US, (July 2018), pp. 1–10. doi: 10.1038/s41598-018-38247-2.

Radonić, A. *et al.* (2004) 'Guideline to reference gene selection for quantitative real-time PCR', *Biochemical and Biophysical Research Communications*. Academic Press Inc., 313(4), pp. 856–862. doi: 10.1016/j.bbrc.2003.11.177.

Silver, N. *et al.* (2006) 'Selection of housekeeping genes for gene expression studies in human reticulocytes using real-time PCR', *BMC Molecular Biology*, 7. doi: 10.1186/1471-2199-7-33.

Sitt, T. *et al.* (2008) 'Quantitation of leukocyte gene expression in cetaceans', *Developmental and Comparative Immunology*. doi: 10.1016/j.dci.2008.05.001.

Sitt, T. *et al.* (2010) 'Cellular immune responses in cetaceans immunized with a porcine erysipelas vaccine', *Veterinary Immunology and Immunopathology*. Elsevier B.V., 137(3–4), pp. 181–189. doi: 10.1016/j.vetimm.2010.05.003.

Spinsanti, G. *et al.* (2006) 'Selection of reference genes for quantitative RT-PCR studies in striped dolphin (Stenella coeruleoalba) skin biopsies', *BMC Molecular Biology*, 7. doi: 10.1186/1471-2199-7-32.

SPSS Inc (2008) 'SPSS Statistics for Windows, Version 17.0.', *Chicago: SPSS Inc.*

Takara Bio Inc (2015) 'RR420A SYBR ® Premix Ex Taq ™ (Tli RNaseH Plus) Product Manual'.

Takara Bio Inc (2017) 'RNAiso Plus (Total RNA extraction reagent) Product Manual'.

Technologies, A. (2015) 'Agilent 2200 TapeStation System', p. 90.

Vandesompele, J. *et al.* (2002) *Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes*. Available at: http://genomebiology.com/2002/3/7/research/0034.1Correspondence:.rankSpeleman.

View, M. (2014) 'RNA clean-up User manual', (May).

Viscarra, J. A. and Ortiz, R. M. (2014) 'deprivation', 62(7), pp. 889–897. doi: 10.1016/j.metabol.2012.12.014.Cellular.

Wang, Z. *et al.* (2014) 'Selection of reference genes for quantitative reverse-transcription polymerase chain reaction normalization in Brassica napus under various stress conditions.', *Molecular genetics and genomics : MGG*. Mol Genet Genomics, 289(5), pp. 1023–1035. doi: 10.1007/s00438-014-0853-1.

Xie, F. *et al.* (2012) 'miRDeepFinder: A miRNA analysis tool for deep sequencing of plant small RNAs', *Plant Molecular Biology*, 80(1), pp. 75–84. doi: 10.1007/s11103-012-9885-2.

Zeng, Z. *et al.* (2020) 'ERgene: Python library for screening endogenous reference genes', *Scientific Reports*. Nature Publishing Group UK, 10(1), pp. 1–7. doi: 10.1038/s41598-020-75586-5.

Supporting information

S1 Fig. Efficiency of the selected genes based on standard curves

S2 Fig. Image of 2% agarose gel electrophoresis of selected candidate 7 reference genes with 100 bp ladder

S3 Fig. qPCR dissociation curves for all candidate reference genes

S4 Fig. Histograms of gene stability analysis from five different methods

S1 Table. Gene expression values of selected genes for the validation as reference genes of RD for fat tissues.

S2 Table. Ct values of standard curves in different gene.

491