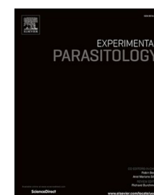
Experimental Parasitology 248 (2023) 108511

Contents lists available at ScienceDirect

Experimental Parasitology

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



Mining *Lepeophtheirus salmonis* RNA-Seq data for qPCR reference genes and their application in *Caligus elongatus*

Andreas Borchel [1,*], Christiane Eichner [1], Aina-Cathrine Øvergård

University of Bergen, SLRC-Sea Lice Research Centre, Department of Biological Sciences, Pb. 7803, 5020, Bergen, Norway

ARTICLE INFO

Keywords:
Gene expression
quantitative PCR
Copepod
RNA-Seq
Reference gene

ABSTRACT

Lepeophtheirus salmonis and *Caligus elongatus* are two parasitic copepod species posing a significant threat to salmonid aquaculture. Consequently, several gene expression studies are executed each year to gain new knowledge and treatment strategies. Though, to enable accurate gene expression measurements by quantitative real time PCR, stable reference genes are needed. Previous studies have mainly focused on a few genes selected based on their function as housekeeping genes, as these are often stably expressed in various cells and tissues. In the present study, however, RNA-sequencing data from 127 *L. salmonis* samples from different life stages and diverse environmental conditions were used to identify new candidate reference genes displaying low variation. From this, six genes were selected, and the stability validated by qPCR on samples from different life stages. Since neither a genome nor comprehensive RNA sequencing data are available for *C. elongatus*, homologous genes to those identified for *L. salmonis* were identified within a *C. elongatus* transcriptome assembly and validated by qPCR in different life stages. Overall, the genes eukaryotic translation initiation factor 1A (EIF1A) and serine/threonine-protein phosphatase 1 (PP1) displayed the highest stability in *L. salmonis*, while the combination of PP1 and ribosomal protein S13 (RPS13) was found to have the highest stability in *C. elongatus*. These genes are well-suited reference genes for qPCR applications which allow for accurate normalization of target genes.

1. Introduction

Lepeophtheirus salmonis and *Caligus elongatus* are fish parasites belonging to the order of Siphonostomatoida within the class of copepods. Especially the former parasite poses a considerable economic and ecological problem in salmonid aquaculture (Costello, 2009). Even though their last common ancestor lived more than 80 million years ago (Kumar et al., 2017), both parasites share a rather similar way of life and have a similar habitus. Both undergo a life cycle consisting of eight life stages separated by molts; the shedding of old and production of new cuticle. The first life stages, nauplius I and II, are lecithotrophic and planktonic; while the copepodid stage is the infective stage which must find and attach to a host. From there on, when the lice are in their parasitic stage, the life cycles differ to some point. While L. *salmonis* has two chalimus stages followed by two pre-adult stages before the final molt to adult (Hamre et al., 2013), *C. elongatus* goes through four chalimus stages directly followed by the adult stage (Piasecki and Mackinnon, 1995).

Analyzing the ontogenetic expression patterns of genes can give a deep insight into gene function in sea lice. For example, genes upregulated as soon as the copepodid settles on the fish might be involved in the interaction of parasite and host, food uptake or digestion, while genes that are strongly expressed exclusively in either the adult female or male can be involved in reproduction. Quantitative real-time PCR (qPCR) is a commonly used method to measure gene expression in cells, tissues, or whole organisms as sea lice. To obtain accurate results, reference genes are used to normalize the expression values of the genes of interest (Jacob et al., 2013). However, to meaningfully counterbalance potential differences in RNA or cDNA input in the qPCR reactions and other technical factors affecting a precise gene expression measurement, these reference genes must not be affected by the investigated biological factors (Kozera and Rapacz, 2013).

In the early years of qPCR, typically only one reference gene was used for normalization, which was often a housekeeping gene like β-actin, glyceraldehyde-3-phosphate dehydrogenase (GAPDH) or 18S rRNA. However, since there is always some natural variation in the transcript level of a given gene, it is recommended to use two or more genes involved in different cellular processes as reference genes (Bustin

* Corresponding author.
E-mail address: andreas.borchel@uib.no (A. Borchel).

[1] Both authors contributed equally to this work.

https://doi.org/10.1016/j.exppara.2023.108511
Received 30 January 2023; Received in revised form 10 March 2023; Accepted 11 March 2023
Available online 13 March 2023
0014-4894/© 2023 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

et al., 2009). For most qPCR studies published on *L. salmonis*, the eukaryotic elongation factor 1 alpha (EF1α) is used as the only reference gene, as its developmental stability has been found acceptable (Frost and Nilsen, 2003). However, when that study was performed, the genetic information available for *L. salmonis* was limited. Therefore, only four candidate genes were tested as potential reference genes based on the use of these in other species. Today, in the age of next generation sequencing, much more data is available to make a more informed decision on the expression stability of basically all genes of a given organism. For L. *salmonis* the complete genome is available, consisting of roughly 13000 genes (Skern-Mauritzen et al., 2021), and the expression of these genes has been measured using RNA-sequencing (RNA-Seq) in more than 400 samples of different salmon louse stages, tissues and environmental conditions (Short Read Archive, https://www.ncbi.nlm.nih.gov/sra, January 2023). Thus, there is a great potential of identifying new and better reference gene candidates in such data sets. Unfortunately, this magnitude of genetic information is not available for *C. elongatus*, as neither the genome nor transcriptomes have been published, and GenBank contains only 31 entries for this species as of January 2023. Therefore, we aimed to use the data obtained in *L. salmonis* to bring forward the research on *C. elongatus*.

In this study, the most suitable reference genes for *L. salmonis* were first identified based on RNA sequencing data, followed by the design and validation of qPCR assays for the selected genes. Further, orthologues were identified in *C. elongatus* transcriptomic data, and their sequences were confirmed to allow the design of accurate qPCR assays. The stability of the genes was further evaluated to be used as reference genes analyzing gene expression in different *C. elongatus* life stages.

2. Methods

2.1. Animal husbandry and sampling

For L. *salmonis*, animals from the lab strain LsGulen were cultivated as previously described (Hamre et al., 2009). For *C. elongatus*, a lab strain called CeSenja was used which had been brought into the lab from a salmon aquaculture facility in Senja in Northern Norway in February 2019, originated on the offspring of six egg strings. The lice of this strain were kept on Atlantic salmon, analogous to the husbandry described for *L. salmonis*. The free-living stages of both lice strains were kept in a flow-through system, before infesting Atlantic salmon (*Salmo salar*). The salmon were kept in accordance with Norwegian animal experiment legislation, and were hand fed on a commercial diet and reared in sea water with a salinity of 34.5 ppt. The water temperature was around 9 °C (*C. elongatus*) and 10 °C (*L. salmonis*).

Samples of the early life stages of both species (nauplius, copepodid, chalimus) contained several animals pooled together to yield enough RNA for later analysis, while preadult and adult stages were sampled individually. While L. *salmonis* samples were taken roughly during the predicted middle of each stage, parasitic stages of *C. elongatus* were sampled at the following daydegrees post infection (ddpi): parasitic copepodid at 18 ddpi, chalimus I at 60 ddpi, chalimus II at 94 ddpi, chalimus III at 138 ddpi, chalimus IV (male and female) at 182 ddpi, young adult female at 199 ddpi and adult male and female 379 ddpi.

All sampled animals were put in RNAlater (Life Technologies), kept overnight at 4 °C and stored at −20 °C until further processing.

2.2. Candidate gene selection

To identify potential reference genes in *C. elongatus* and new reference genes for *L. salmonis*, we used a selection of RNA-Seq derived data from *L. salmonis* generated in our lab in the last 10 years (Borchel et al., 2021; Eichner et al., 2018; Heggland et al., 2020; Øvergård et al., 2022, unpublished data). Overall, data from 127 samples were included in the analysis (Supplementary Table 1), consisting of paired-end Illumina--generated short-reads from all *L. salmonis* life stages and lice kept at different salinities. The data analysis was the same for all samples. In short, the reads were first mapped against the salmon louse genome (LSalAtl2s) using STAR aligner (Dobin et al., 2013) in the Galaxy web platform using the public server usegalaxy.no (Afgan et al., 2018). The counts of mapped reads for each gene were obtained using FeatureCounts (Liao et al., 2014). Inter-sample normalization was then performed using DeSeq2 (Love et al., 2014), on a combined count table followed by a normalization of the obtained values by the gene length. For each gene, the mean, standard deviation, and coefficient of variation (CV, defined as standard deviation/mean) were calculated. The genes were then ranked according to their CV, assuming that genes with a low CV are well-suited candidate reference genes.

Five new candidate reference genes were chosen, based on CV and normalized count numbers (NCNs). The two highest ranking genes with the lowest CV had rather low NCNs. Therefore, the next three genes were selected based on a low CV and higher NCNs. These were the gene with the lowest CV among genes with an NCN ≥1000 and two genes with the lowest CV among genes with an NCN ≥10000. Additionally, we included three previously established reference genes, EF1α1 (Frost and Nilsen, 2003), 40S ribosomal protein S13 (RPS13) (Frost and Nilsen, 2003), and ADP/ATP carrier protein 3 (ADT3) (Eichner et al., 2015). This gave a total of eight genes selected for further analysis in *L. salmonis*.

To find reference genes for *C. elongatus*, we searched for orthologues to the chosen *L. salmonis* genes in a de-novo-transcriptome assembled from adult female RNA-Seq data (in prep.). Primers were designed based on sequences predicted from the transcriptome, PCR products generated and sequenced using GoTaq G2 DNA Polymerase (Promega) and Big Dye 3.1 (ThermoFisher Scientific), respectively. Additionally, β-tubulin was included as it has been used as a reference gene in *Caligus rogercresseyi* (Gallardo-Escárate et al., 2014). This gave a total of nine genes selected for further analysis in *C. elongatus*.

2.3. RNA isolation, cDNA synthesis & qPCR

For L. *salmonis*, total RNA was isolated using Trizol reagent (Sigma Aldrich) according to the manufacturer's protocol. The RNA (1 μg) was DNase-treated with DNase I (Invitrogen), according to the manufacturer's protocol. For *C. elongatus*, total RNA was isolated by combining Trizol reagent with RNeasy (Qiagen) RNA isolation kits, as previously described (Harasimczuk et al., 2018), including an on-column DNase treatment.

The RNA purity and quantity were determined using a Nanodrop spectrophotometer, where the 260/280 and 260/230 ratios were between 1.9 and 2.2, with few lower-concentrated samples outside this range for the 260/230 ratio.

Equal amounts of DNase-treated RNA (182 ng for *L. salmonis*, 100 ng for *C. elongatus*) were then reverse transcribed using the AffinityScript cDNA synthesis kit (Agilent) according to the supplier's recommendation and diluted 1:10 before storage at −20 °C until use.

The qPCR was run on QuantStudio 3 Real-Time PCR machines (Applied Biosystems), using 1x PowerUP SYBR Green Master Mixes (ThermoFisher Scientific), 500 nM forward and reverse primers (Table 1) and 2 μl diluted cDNA in 10 μl reactions (3.6 ng RNA-equivalents for *L. salmonis*, 2 ng for *C. elongatus*). The thermocycling parameters were as follows: initiation, 50 °C, 2 min; holding, 95 °C, 2 min; 40 cycles of 95 °C, 15 s & 60 °C, 1 min, followed by a concluding melting curve. All measurements were done in technical duplicates. Standard curves with a minimum of five points (1:5) were made for each assay to calculate PCR efficiencies, given by the equation $E\% = (10^{1/slope} - 1) \times 100$ (Ginzinger, 2002). Efficiencies of the new assays were between 94% and 109% (Supplementary Table 2). Every plate run, contained non template controls to check for contamination.

For L. *salmonis*, five of the used primer pairs had one primer placed on an exon-junction, two other primer pairs spanned an intron (RPS13 and ADT3) of 66 bp and 53 bp size, respectively. One gene (EIF1A) was annotated with only one exon, therefore the primers were placed within

Table 1
Primers used in qPCR.

Gen	FW	RV
Lepeophtheirus salmonis		
EF1A	GGTCGACAGACGTACTGGTAAATCC	TGCGGCCTTGGTGGTGGTTC
EIF1A	CCCTCTTCGCCAAAAGTGAC	CAGACGAGGCTCGCAACTTA
ADT3	CTGGAGAGGGAATTTGGCTAACGTG	GACCCTGGACACCGTCAGACTTCA
RPS13	GCCGGTGTTTAACAATCATCAA	GGGCTTCGAGTCCTTGTATGC
B52	CAGATAAGGATGCTATGGGTTGG	ATGTGCGCGACAGATGAGAT
PP1	GATGAATGCAAACGGCGATA	CGTCTGTAGGTCAGGGCTCA
NELFD	ACGGATAATTGGAGCCATCG	GAAGCTACCGGGAACCTTATG
MTMR2	GCATTTCAATATCAAGTCAATGAAGG	GAAGCCCCTGACGACGATAC
Caligus elongatus		
EF1A	GACGTACCGGCAAGTCCAC	GGAGGGAAGTCGGAGAAGG
EIF1A	GGAAAGGAGGCAAGAATCGC	TCCCCAGCATCTTTGTCACT
ADT3	AACTCACGCTTTCCGTCCTT	AGAGCGTCTTCAAGACTGCC
RPS13	GATCCTCCGCATCCTCAAGG	ACGGCCTTCTTGATGAGGTG
B52	TGAAGATTCAAATCACACACCCG	ATCTGGCGGCAGAATGATCC
PP1	GGGAAGCCACCGTACTCAAA	CCAGCCCATCTTGTTGGAGT
NELFD	CTGTAACCGCCGAATAACATGG	CTCTTACGGTGGCGAATATGGA
MTMR2	GACTTTGTCTTCTGCTTGAGCC	TTCCTCACCATCATTCTGGACC
Tub	GGGAACGCTCCTCATTTCCA	GACGGTGTCAGAGACCTTGG
Trypsin	ACGATCTCTTCTCCGACAGC	ACGTCAAAGTTGGTTCCGTG
Myosin	AGAAGCGACTCCAGGTTGAA	CCAAGCGCTTCGTATCTTCC

this exon. For *C. elongatus*, the deliberate positioning of the primers on exon junctions was impossible as genomic sequence information was missing. However, our analysis of a PCR reaction on genomic DNA revealed, that the ADT3 primer was spanning an intron of 53 bp.

For the intron-spanning primers, melting curves were analyzed for double peaks to check for genomic DNA contamination, which was not observed. Additionally, -RT controls performed in *C. elongatus* did not show signs of contamination with genomic DNA.

2.4. Analysis

Overall, we followed the suggestions for identification of reference genes in longitudinal studies (Sundaram et al., 2019) including analysis of CV, raw expression profiles and use of NormFinder (Andersen et al., 2004), which takes differences between groups into account. Several additional algorithms were used to analyze the stability of the potential reference genes, collected in the tool RefFinder (Xie et al., 2012). This tool combines the methods Genorm (Vandesompele et al., 2002), Best-Keeper (Pfaffl et al., 2004), NormFinder (Andersen et al., 2004), and a delta ct based method (Silver et al., 2006).

3. Results

3.1. Reference genes for L. salmonis

Potential reference genes for *L. salmonis* were selected based on CV and normalized counts from RNA-Seq data as described in the methods part (Table 2). The novel genes had lower CVs (8–20%) than the established reference genes (25–37%).

To validate the suitability of the eight candidate reference genes, qPCR was performed on samples from different *L. salmonis* life stages. The obtained quantification cycle (c_q)-values were in general relatively stable for all genes (Fig. 1A), where EF1A, RPS13 and ADT3 showed the highest expression with the lowest measured c_q-values, while MTMR2 and NELFD displayed the lowest expression. As also seen in the transcriptomic data, the CV of the linearized c_q values (2^{-Cq}) was rather high for the three established reference genes (28.5–29.8%) and lower for the new candidate genes (22.9–27.2%), except for NELFD which had the highest CV (38.1%). The lowest CV was observed for EIF1A (22.6%) and MTMR2 (24.1%). Raw expression profiles showed a rather stable expression for all genes throughout the life stages (Fig. 1 B). A grouped Normfinder analysis (Fig. 1C) found highest stability of EIF1a and PP1, which were also best to be combined. The established reference genes

Table 2
Selected candidate reference genes from *L. salmonis*.

Symbol	Gene id	Gene name	Average normalized counts	Variation coefficient
ADT3	EMLSAG00000008594	ADP/ATP carrier protein 3	23833	37%
RPS13	EMLSAG00000007940	40S ribosomal protein S13	43351	33%
EF1α	EMLSAG00000004106	eukaryotic translation elongation factor 1 alpha 1	88735	25%
EIF1A	EMLSAG00000000100	eukaryotic translation initiation factor 1A	10887	20%
B52	EMLSAG00000007146	Serine-arginine protein 55 B52	12745	19%
PP1	EMLSAG00000010494	serine/threonine-protein phosphatase 1	1833	11%
NELFD	EMLSAG00000005404	Negative elongation factor D	491	8%
MTMR2	EMLSAG00000012255	Myotubularin-related protein 2	778	8%

EF1A and especially ADT3 were determined to be least stable. Also the algorithms employed by RefFinder suggested that ADT3 was least stable, while PP1 and EIF1A ranked among the most stable genes (Supplementary Table 3).

3.2. Reference genes for C. elongatus

After identifying *C. elongatus* homologs for all the candidate reference genes within the transcriptome assembly, PCR fragments were sequenced to verify their identity. The sequences are deposited in Genbank (PP1: OQ560354, ADT3: OQ560355, EF1A: OQ560356, RPS13: OQ560357, MTMR2: OQ560358, NELFD: OQ560359, Tubulin: OQ560360, EIF1A: OQ560361).

qPCR was further performed on different *C. elongatus* life stages (Fig. 2A). They showed variation coefficients in a similar range as observed in *L. salmonis*. The lowest variation was found for PP1 (24.2%) < Tub < TMR2 < EIF1A < RPS13 < B52 < ADT3 < NELFD < EF1A (43.3%).

Life stage had a statistically significant effect on all genes (ANOVA), however the differences varied in their strength (Fig. 2). While ADT3 and EF1A showed rather strong differences between stages, expression levels of the novel reference genes were generally more stable.

When running a Normfinder analysis, considering the different life stages as groups, the most stable reference gene was again determined to be EIF1A, whereas the best combination of two genes was PP1 and RPS13 (Fig. 2C). Also the RefFinder analyses found EIF1A to be highly stable, while the stability ranking for PP1 and RPS13 varied between different algorithms (Supplementary Table 3).

To apply the candidate reference genes, the expression of myosin (OQ560363) and a trypsin gene (OQ560362) were measured in the different *C. elongatus* life stages and normalized against all candidate reference genes (Fig. 3). Overall, the expression patterns obtained by normalization with the different reference genes were similar, but specific differences were present. The expression of myosin, a protein generally highly abundant in muscle cells, started on a rather high level after hatching as nauplius 1, before dropping in the attached stages and later increasing again. However, normalizing to different reference

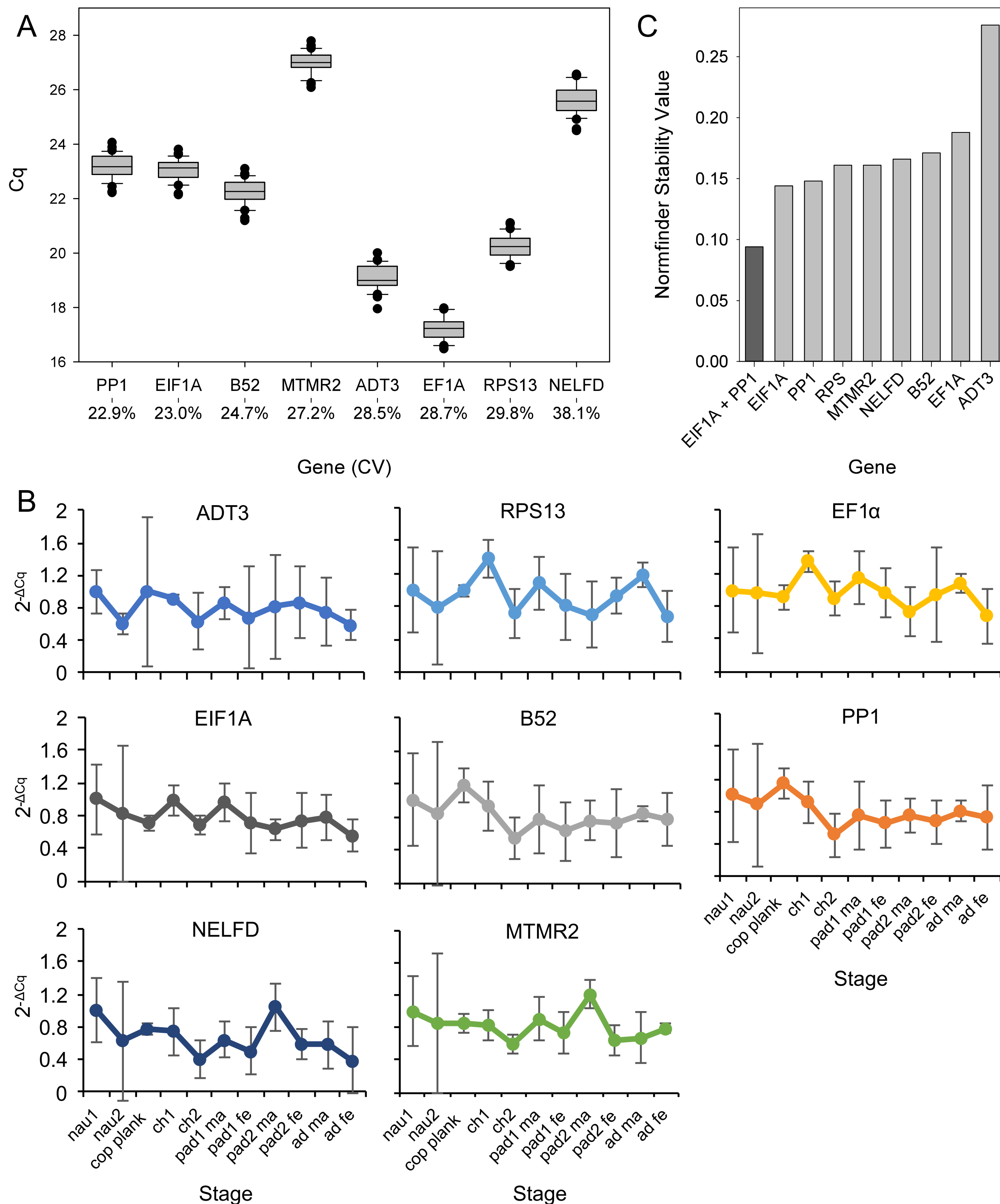


Fig. 1. Reference gene stability in whole body *L. salmonis* samples from different life stages (N = 3, total of 33 samples). A) C_q values of the different candidate reference genes in different life stages. Boxes give 25th and 75th percentile and median, whiskers indicate 90th and 10th percentiles. Dots represent values outside these percentiles. Genes are ordered by their linearized coefficient of variation (CV, value underneath gene name). B) Raw expression levels ($2^{-\Delta Cq}$). Dots represent means ± SD. The nauplius 1 stage is used as calibrator. C) Stability values as determined by Normfinder analysis. The most stable combination of genes is marked dark grey. Lower stability values indicate more stable genes.; nau = nauplius, cop = copepodid, ch = chalimus, pad = preadult, ad = adult, ma = male, fe = female; for gene names see Table 2.

genes, resulted in some cases in a changed order of which stage had the highest expression (for example compare EF1A and NELFD). The expression of trypsin, a digestive enzyme, increased steadily with age. However, the numerical values of relative expression varied between reference genes also for this target gene.

4. Discussion

Stable reference genes are essential for accurate gene expression quantification by qPCR. Often, potential reference genes are selected based on their previous evaluation and/or use in other species, and only a handful of genes are often analyzed during the selection process (e.g. Frost and Nilsen, 2003; Jorgensen et al., 2006; Mahanty et al., 2017;

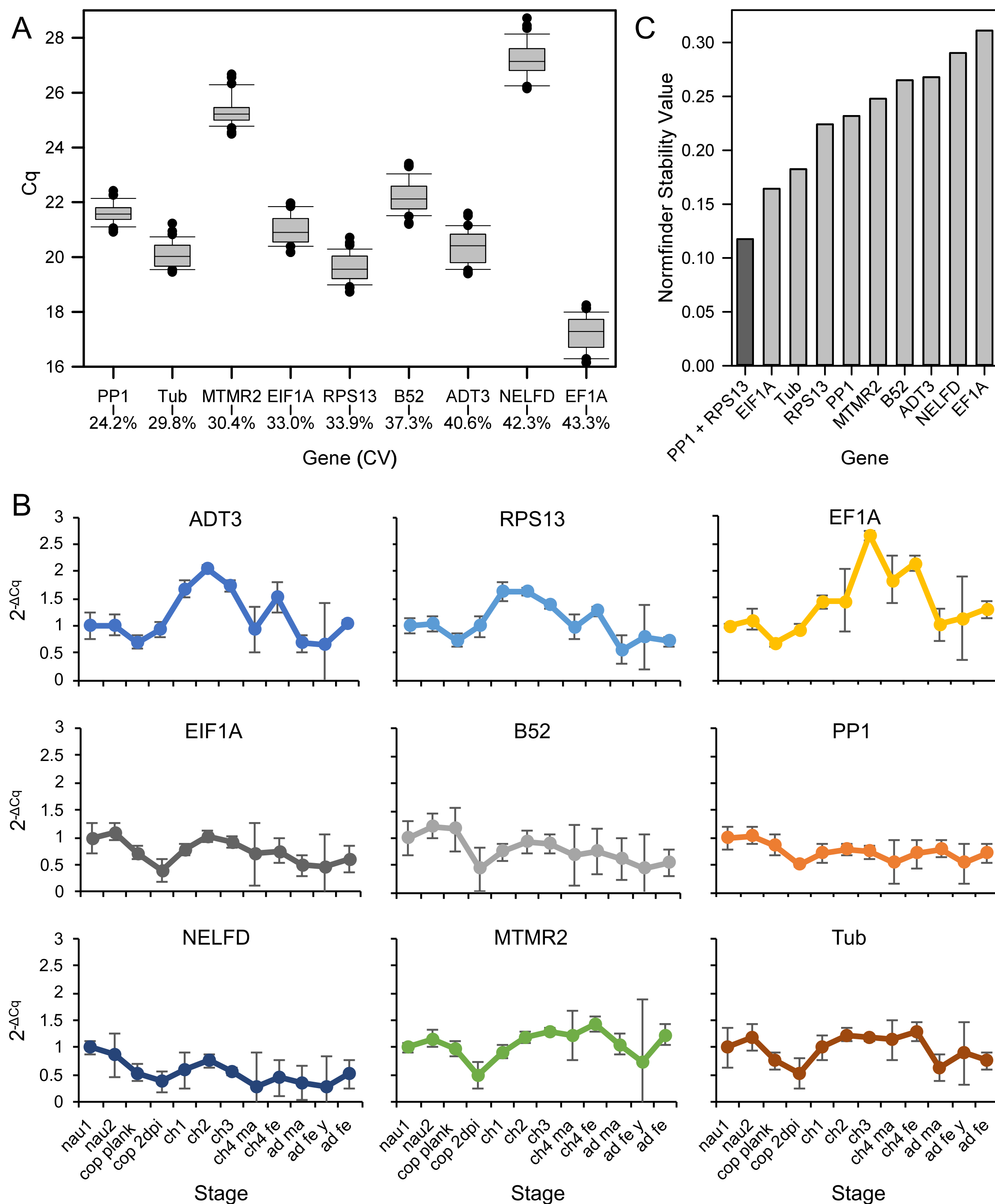


Fig. 2. Reference gene stability in *C. elongatus* samples from different life stages (N = 3, total of 36 samples). A) C_q values of the different candidate reference genes in different life stages. Boxes give 25th and 75th percentile and median, whiskers indicate 90th and 10th percentiles. Dots represent values outside these percentiles. Genes are ordered by their linearized coefficient of variation (CV, value underneath gene name). B) Raw expression levels ($2^{-\Delta Cq}$). Dots represent means ± SD. The nauplius 1 stage is used as calibrator. C) Stability values as determined by Normfinder analysis. The most stable combination of genes is marked dark grey. Lower stability values indicate more stable genes.; nau = nauplius, cop = copepodid, ch = chalimus, ad = adult, ma = male, fe = female, y = young; for gene names see Table 2.

Olsvik et al., 2005; Wang et al., 2018). Therefore, more suitable reference genes are likely to be found by an extended analysis, as performed in the present study looking at extensive RNA-seq datasets created from a variety of *L. salmonis* samples. By this approach, we could identify new reference genes showing a more stable expression pattern than previously used reference genes in *L. salmonis*. A similar RNA-Seq-based approach has also been proven useful– though with fewer samples – to determine reference genes in e.g. apple (Zhou et al., 2017), *Arabidopsis* (dos Santos et al., 2020) and pregnant women (Chim et al., 2017).

Another study compared the stability of reference genes identified by an RNA-Seq pipeline and conventional reference genes (Sampathkumar et al., 2022). There it was found that RNA-Seq-based reference genes not necessarily perform better than conventional reference genes. The conventional reference genes yielded even better stability values. In our

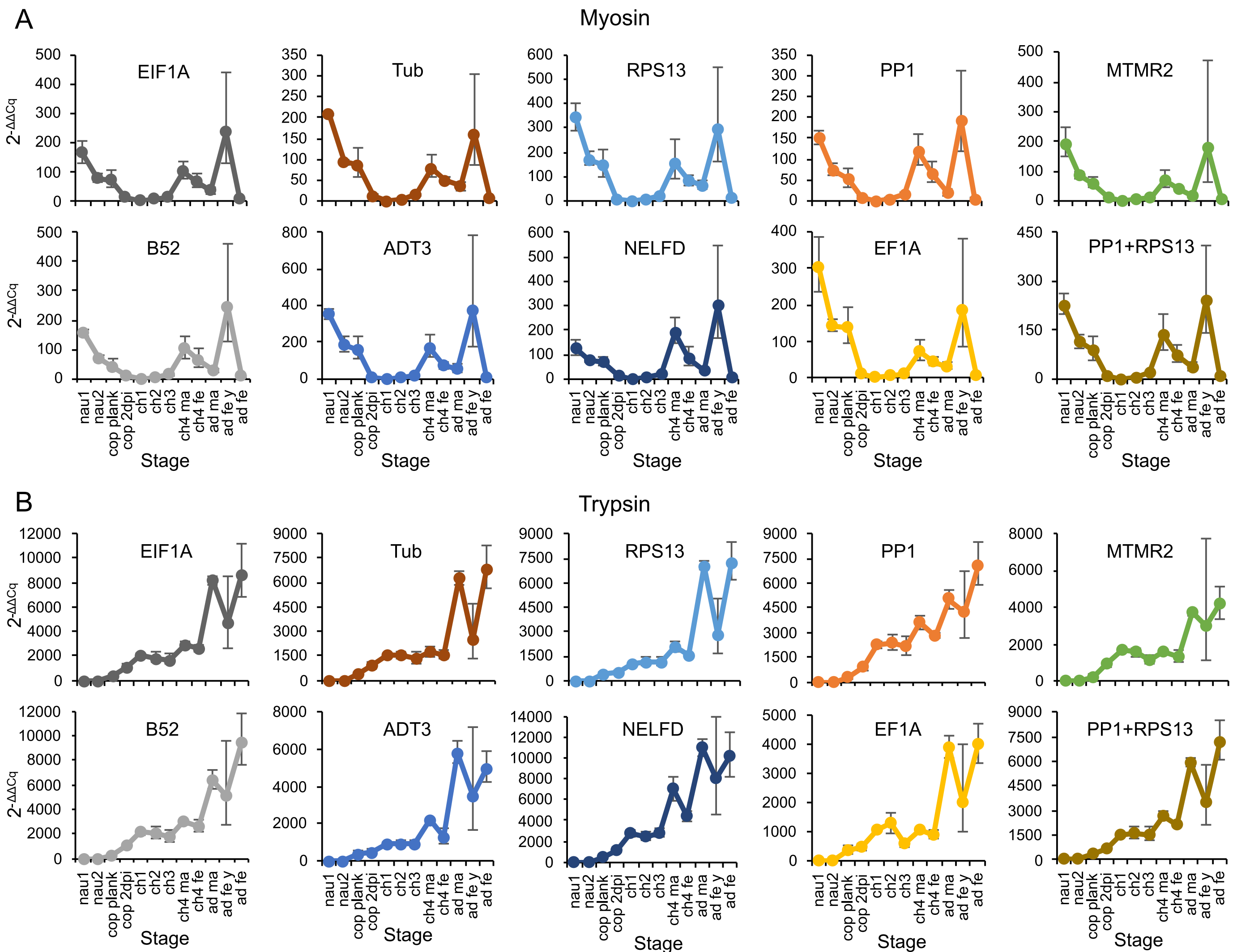


Fig. 3. Relative expression (average $2^{-\Delta\Delta Cq}$ ± SD, N = 3) of A) myosin and B) trypsin in samples of the different stages of *C. elongatus*. The reference genes investigated in this study were used as endogenous control. The reference gene used is shown in the header of each diagram. In every gene's last panel, the expression level calculated by using a combination of two reference genes recommended by the grouped Normfinder algorithm is shown. The stage with the lowest gene expression (chalimus 1 and nauplius 1 respectively) was chosen as a calibrator and set to 1.

experiments the results were mixed. According to Normfinder, both in *C. elongatus* and *L. salmonis*, a novel reference gene was most stable, however for both species some of the conventional genes were also among the top ranks. In the case of *C. elongatus* the combination of a novel (PP1) and a conventional (RPS13) reference gene was even determined to be most stable. For this project, the use of RNA-Seq data was reasonable, as these data had already been generated and could be used without further costs.

Employing an RNA-seq approach, we identified five new *L. salmonis* reference genes which are generally adequate for normalization of qPCR results. This has been especially validated for measurements throughout the life cycle of the salmon louse. These genes are most likely stable also under various environmental conditions, as the selection process was based on RNA Seq data from lice in different salinity and with and without a host. However, the suitability of a given reference genes should be evaluated for every new condition in question. E.g., in a study where *L. salmonis* were subjected to oxidative stress due to the knock down (KD) of a heme peroxidase, EF1α and ADT3 varied greatly between normal and KD copepodids (Øvergård et al., 2017). Therefore, it is helpful to have a selection of new potential reference genes at hand when qPCR is applied to new experimental setups. For analysis of expression during the *L. salmonis* life cycle the combination of EIF1A and PP1 was found to be most stable.

When applying the putative reference genes to investigate the expression of the two *C. elongatus* target genes, myosin and trypsin, the expression patterns observed mirrored the biological functions of the measured genes. Myosin is needed for muscle contraction (Geeves and Holmes, 1999); therefore, it is reasonable that it is less expressed in the sessile stages compared to the free-living and mobile stages where the lice is more dependent on a higher muscle power to move and attach to the host. Trypsin is among other physiological processes involved in digestion (Kvamme et al., 2005); therefore, it might be less needed in lecithotrophic free-living stages but its expression increases when attached to a host that provides nutrients to be digested.

The analyzed target genes were strongly regulated between stages, with c_q value differences of 8 and 13 for myosin and trypsin, respectively. Therefore, even small differences in the expression of the used reference genes had a big impact on the calculated fold changes. E.g., this can be seen as an expression of trypsin in adult female lice about 10000 times as high as in nauplius 1 when measured with NELFD ($\Delta\Delta\ c_q = -13.3$), while it is only about 4000 times higher when measured with EF1A ($\Delta\Delta\ c_q = -12.0$). Here the fold change varied strongly due to the

low expression levels in the calibrator stage being sensitive to small changes in the delta c_q values. Using an average of PP1 and RPS13 resulted in a fold change of about 7200 between nauplius 1 and adult female, corresponding to around the average of the fold change of all reference genes used, and also gave an intermediate expression pattern. Generally, it is recommended to use more than one reference gene for qPCR normalization purposes, and the choice of reference gene combinations should be well reasoned (Bustin et al., 2009). For example, the combination of coregulated genes might be detrimental, as these could be inadequate to counterbalance differences in cDNA concentration. For L. *salmonis*, EIF1A1 and PP1 were determined as the best combination of stable genes. While EIF1A is part of translation (Passmore et al., 2007), PP1 dephosphorylates proteins (Bollen et al., 2010), arguing that a combination of these genes is reasonable. For *C. elongatus* the combination of PP1 and RPS13 was recommended by the Normfinder algorithm, which also seems reasonable as RPS13 is a ribosomal protein while PP1 is a phosphatase, without known interaction of both proteins.

For *C. elongatus*, we could not revert to a vast diversity and magnitude of genetic information as for *L. salmonis*. Therefore, the same candidate reference genes as chosen for *L. salmonis* were evaluated in *C. elongatus* in addition to β-tubulin. Overall, this seemed to be an appropriate strategy, as the genes were also found to be rather stable in the different *C. elongatus* life stages with the most stable genes identified as EIF1A, β-tubulin and PP1. This indicates that the stability of especially EIF1A as well as PP1 might be conserved in both sea louse species, suggesting that they might also be good candidates for reference genes in other sea lice or potentially copepods in general. EIF1A was even found to be the most stable gene by several algorithms. At the same time, our findings call for carefulness when analyzing gene expression data during sea louse development and most likely animal development in general. Even though the novel reference genes are most likely among the most stable genes, the raw c_q values showed some statistically significant variation between life stages for each gene. Similar challenges were observed during the identification of reference genes for the early murine development of spinal cord and cerebellum (Sundaram et al., 2019). In that article a general workflow for the identification of reference genes in longitudinal studies was suggested, consisting of an analysis of the CV and raw expression profiles with a final NormFinder analysis considering the different groups. Overall, we followed this workflow, which suggests removing genes with CV over 50% from the NormFinder analysis. As all candidate genes from this study had a lower CV, this step was unnecessary, which also highlights the good selection of candidate genes based on RNA-Seq data. Based on this, the combination of PP1 and RPS13 is recommended for studies on *C. elongatus* ontogeny.

5. Conclusion

In the present work, we have identified novel reference genes for the two sea louse species *L. salmonis* and *C. elongatus*. The genes found to be stably expressed between life stages in *L. salmonis*, were also stable in *C. elongatus*. This work facilitates the implementation of new reference genes in *L. salmonis*, where they might complement the already established reference genes. Additionally, we have contributed the first nuclear-encoded mRNA sequences from *C. elongatus* to GenBank and identified reference genes that can be used in this species. As its economic importance is on the rise, the availability of reference genes for qPCR assays is of high importance.

Author statement

Andreas Borchel: Conceptualization, Methodology, Formal analysis, Investigation, Resources, Writing - Original Draft, Writing - Review & Editing, Visualization.

Christiane Eichner: Conceptualization, Methodology, Formal analysis, Investigation, Resources, Writing - Original Draft, Writing - Review & Editing, Visualization.

Aina-Cathrine Øvergård: Conceptualization, Resources, Writing - Review & Editing, Funding acquisition.

Declaration of competing interest

None.

Data availability

Data will be made available on request.

Acknowledgements

We want to thank Heidi Kongshaug for excellent support in the lab, and Lars Are Hamre and Per Gunnar Espedal for excellent animal care. Moreover, we thank Lars Are Hamre for bringing the CeSenja strain into the lab and learning how best to cultivate *C. elongatus*, so that this study was possible.

This work was funded by the project “ModuLus” (901564) funded by the Norwegian Seafood Research Fund (FHF).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.exppara.2023.108511.

References

Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Čech, M., Chilton, J., Clements, D., Coraor, N., Grüning, B.A., Guerler, A., Hillman-Jackson, J., Hiltemann, S., Jalili, V., Rasche, H., Soranzo, N., Goecks, J., Taylor, J., Nekrutenko, A., Blankenberg, D., 2018. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Res. 46, W537–W544. https://doi.org/10.1093/nar/gky379.

Andersen, C.L., Jensen, J.L., Ørntoft, T.F., 2004. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. Cancer Res. 64, 5245–5250. https://doi.org/10.1158/0008-5472.CAN-04-0496.

Bollen, M., Peti, W., Ragusa, M.J., Beullens, M., 2010. The extended PP1 toolkit: designed to create specificity. Trends Biochem. Sci. 35, 450–458. https://doi.org/10.1016/j.tibs.2010.03.002.

Borchel, A., Heggland, E.I., Nilsen, F., 2021. The transcriptomic response of adult salmon lice (Lepeophtheirus salmonis) to reduced salinity. Comp. Biochem. Physiol. - Part Genomics Proteomics 37, 100778. https://doi.org/10.1016/j.cbd.2020.100778.

Bustin, S.A., Benes, V., Garson, J.A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M.W., Shipley, G.L., Vandesompele, J., Wittwer, C.T., 2009. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clin. Chem. 55, 611–622. https://doi.org/10.1373/clinchem.2008.112797.

Chim, S.S.C., Wong, K.K.W., Chung, C.Y.L., Lam, S.K.W., Kwok, J.S.L., Lai, C.-Y., Cheng, Y.K.Y., Hui, A.S.Y., Meng, M., Chan, O.-K., Tsui, S.K.W., Lee, K.-Y., Chan, T.-F., Leung, T.-Y., 2017. Systematic selection of reference genes for the normalization of circulating RNA transcripts in pregnant women based on RNA-seq data. Int. J. Mol. Sci. 18, 1709. https://doi.org/10.3390/ijms18081709.

Costello, M.J., 2009. The global economic cost of sea lice to the salmonid farming industry. J. Fish. Dis. 32, 115–118. https://doi.org/10.1111/j.1365-2761.2008.01011.x.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., Gingeras, T.R., 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21. https://doi.org/10.1093/bioinformatics/bts635.

dos Santos, K.C.G., Desgagné-Penix, I., Germain, H., 2020. Custom selected reference genes outperform pre-defined reference genes in transcriptomic analysis. BMC Genom. 21, 35. https://doi.org/10.1186/s12864-019-6426-2.

Eichner, C., Dondrup, M., Nilsen, F., 2018. RNA sequencing reveals distinct gene expression patterns during the development of parasitic larval stages of the salmon louse (*Lepeophtheirus salmonis*). J. Fish. Dis. 41, 1005–1029. https://doi.org/10.1111/jfd.12770.

Eichner, C., Øvergård, A.-C., Nilsen, F., Dalvin, S., 2015. Molecular characterization and knock-down of salmon louse (Lepeophtheirus salmonis) prostaglandin E synthase. Exp. Parasitol. 159, 79–93. https://doi.org/10.1016/j.exppara.2015.09.001.

Frost, P., Nilsen, F., 2003. Validation of reference genes for transcription profiling in the salmon louse, *Lepeophtheirus salmonis*, by quantitative real-time PCR. Vet. Parasitol. 118, 169–174. https://doi.org/10.1016/j.vetpar.2003.09.020.

Gallardo-Escárate, C., Valenzuela-Muñoz, V., Nuñez-Acuña, G., Chávez-Mardones, J., Maldonado-Aguayo, W., 2014. Transcriptome analysis of the couch potato (CPO)

protein reveals an expression pattern associated with early development in the salmon louse Caligus rogercresseyi. Gene 536, 1–8. https://doi.org/10.1016/j.gene.2013.11.100.

Geeves, M.A., Holmes, K.C., 1999. Structural mechanism of muscle contraction. Annu. Rev. Biochem. 68, 687–728. https://doi.org/10.1146/annurev.biochem.68.1.687.

Ginzinger, D.G., 2002. Gene quantification using real-time quantitative PCR: an emerging technology hits the mainstream. Exp. Hematol. 30, 503–512. https://doi.org/10.1016/S0301-472X(02)00806-8.

Hamre, L.A., Eichner, C., Caipang, C.M.A., Dalvin, S.T., Bron, J.E., Nilsen, F., Boxshall, G., Skern-Mauritzen, R., 2013. The salmon louse *Lepeophtheirus salmonis* (copepoda: caligidae) life cycle has only two chalimus stages. PLoS One 8, e73539. https://doi.org/10.1371/journal.pone.0073539.

Hamre, L.A., Glover, K.A., Nilsen, F., 2009. Establishment and characterisation of salmon louse (*Lepeophtheirus salmonis* (Krøyer 1837)) laboratory strains. Parasitol. Int. 58, 451–460. https://doi.org/10.1016/j.parint.2009.08.009.

Harasimczuk, E., Øvergård, A.-C., Grotmol, S., Nilsen, F., Dalvin, S., 2018. Characterization of three salmon louse (Lepeophtheirus salmonis) genes with fibronectin II domains expressed by tegumental type 1 glands. Mol. Biochem. Parasitol. 219, 1–9. https://doi.org/10.1016/j.molbiopara.2017.12.002.

Heggland, E.I., Dondrup, M., Nilsen, F., Eichner, C., 2020. Host gill attachment causes blood-feeding by the salmon louse (Lepeophtheirus salmonis) chalimus larvae and alters parasite development and transcriptome. Parasites Vectors 13, 225. https://doi.org/10.1186/s13071-020-04096-0.

Jacob, F., Guertler, R., Naim, S., Nixdorf, S., Fedier, A., Hacker, N.F., Heinzelmann-Schwarz, V., 2013. Careful selection of reference genes is required for reliable performance of RT-qPCR in human normal and cancer cell lines. PLoS One 8, e59180. https://doi.org/10.1371/journal.pone.0059180.

Jorgensen, S.M., Kleveland, E.J., Grimholt, U., Gjoen, T., 2006. Validation of reference genes for real-time Polymerase chain reaction studies in Atlantic salmon. Mar. Biotechnol. 8, 398–408. https://doi.org/10.1007/s10126-005-5164-4.

Kozera, B., Rapacz, M., 2013. Reference genes in real-time PCR. J. Appl. Genet. 54, 391–406. https://doi.org/10.1007/s13353-013-0173-x.

Kumar, S., Stecher, G., Suleski, M., Hedges, S.B., 2017. TimeTree: a resource for timelines, timetrees, and divergence times. Mol. Biol. Evol. 34, 1812–1819. https://doi.org/10.1093/MOLBEV/MSX116.

Kvamme, B.O., Kongshaug, H., Nilsen, F., 2005. Organisation of trypsin genes in the salmon louse (Lepeophtheirus salmonis, Crustacea, copepoda) genome. Gene 352, 63–74. https://doi.org/10.1016/j.gene.2005.03.011.

Liao, Y., Smyth, G.K., Shi, W., 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30, 923–930. https://doi.org/10.1093/bioinformatics/btt656.

Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550. https://doi.org/10.1186/s13059-014-0550-8.

Mahanty, A., Purohit, G.K., Mohanty, S., Nayak, N.R., Mohanty, B.P., 2017. Suitable reference gene for quantitative real-time PCR analysis of gene expression in gonadal tissues of minnow Puntius sophore under high-temperature stress. BMC Genom. 18, 617. https://doi.org/10.1186/s12864-017-3974-1.

Olsvik, P.A., Lie, K.K., Jordal, A.-E.O., Nilsen, T.O., Hordvik, I., 2005. Evaluation of potential reference genes in real-time RT-PCR studies of Atlantic salmon. BMC Mol. Biol. 6, 21. https://doi.org/10.1186/1471-2199-6-21.

Øvergård, A.-C., Eichner, C., Nilsen, F., Dalvin, S., 2017. Molecular characterization and functional analysis of a salmon louse (Lepeophtheirus salmonis, Krøyer 1838) heme peroxidase with a potential role in extracellular matrixes. Comp. Biochem. Physiol. A. Mol. Integr. Physiol. 206, 1–10. https://doi.org/10.1016/j.cbpa.2017.01.004.

Øvergård, A.-C., Midtbø, H.M.D., Hamre, L.A., Dondrup, M., Bjerga, G.E.K., Larsen, Ø., Chettri, J.K., Buchmann, K., Nilsen, F., Grotmol, S., 2022. Small, charged proteins in salmon louse (Lepeophtheirus salmonis) secretions modulate Atlantic salmon (Salmo salar) immune responses and coagulation. Sci. Rep. 12, 7995. https://doi.org/10.1038/s41598-022-11773-w.

Passmore, L.A., Schmeing, T.M., Maag, D., Applefield, D.J., Acker, M.G., Algire, M.A., Lorsch, J.R., Ramakrishnan, V., 2007. The eukaryotic translation initiation factors eIF1 and eIF1A induce an open conformation of the 40S ribosome. Mol. Cell 26, 41–50. https://doi.org/10.1016/j.molcel.2007.03.018.

Pfaffl, M.W., Tichopad, A., Prgomet, C., Neuvians, T.P., 2004. 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.

Piasecki, W., Mackinnon, B.M., 1995. Life cycle of a sea louse, *Caligus elongatus* von Nordmann, 1832 (copepoda, Siphonostomatoida, caligidae). Can. J. Zool. 73, 74–82. https://doi.org/10.1139/z95-009.

Sampathkumar, N.K., Sundaram, V.K., Danthi, P.S., Barakat, R., Solomon, S., Mondal, M., Carre, I., Jalkh, T.E., Padilla-Ferrer, A., Grenier, J., Massaad, C., Mitchell, J.C., 2022. RNA-Seq is not required to determine stable reference genes for qPCR normalization. PLoS Comput. Biol. 18, e1009868 https://doi.org/10.1371/journal.pcbi.1009868.

Silver, N., Best, S., Jiang, J., Thein, S.L., 2006. 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.

Skern-Mauritzen, R., Malde, K., Eichner, C., Dondrup, M., Furmanek, T., Besnier, F., Komisarczuk, A.Z., Nuhn, M., Dalvin, S., Edvardsen, R.B., Klages, S., Huettel, B., Stueber, K., Grotmol, S., Karlsbakk, E., Kersey, P., Leong, J.S., Glover, K.A., Reinhardt, R., Lien, S., Jonassen, I., Koop, B.F., Nilsen, F., 2021. The salmon louse genome: copepod features and parasitic adaptations. Genomics 113, 3666–3680. https://doi.org/10.1016/J.YGENO.2021.08.002.

Sundaram, V.K., Sampathkumar, N.K., Massaad, C., Grenier, J., 2019. Optimal use of statistical methods to validate reference gene stability in longitudinal studies. PLoS One 14, e0219440. https://doi.org/10.1371/journal.pone.0219440.

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., Speleman, F., 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol. 3 https://doi.org/10.1186/gb-2002-3-7-research0034 research0034.1.

Wang, H., Wen, H., Li, Y., Zhang, K., Liu, Y., 2018. Evaluation of potential reference genes for quantitative RT-PCR analysis in spotted sea bass (Lateolabrax maculatus) under normal and salinity stress conditions. PeerJ 6, e5631. https://doi.org/10.7717/peerj.5631.

Xie, F., Xiao, P., Chen, D., Xu, L., Zhang, B., 2012. miRDeepFinder: a miRNA analysis tool for deep sequencing of plant small RNAs. Plant Mol. Biol. 80, 75–84. https://doi.org/10.1007/s11103-012-9885-2.

Zhou, Z., Cong, P., Tian, Y., Zhu, Y., 2017. Using RNA-seq data to select reference genes for normalizing gene expression in apple roots. PLoS One 12, e0185288. https://doi.org/10.1371/journal.pone.0185288.